

ACP : choses à retenir





L'ACP permet de décrire un jeu de données, de le résumer, d'en réduire la dimensionnalité.

L'ACP répond à différentes questions :

Etude des individus (i.e. des athlètes) : deux athlètes sont proches s'ils ont des résultats similaires.

On s'intéresse à la variabilité entre individus. Y a-t-il des similarités entre les individus pour toutes les variables ? Peut-on établir des profils d'athlètes ? Peut-on opposer un groupe d'individus à un autre ?

Etude des variables (i.e. des performances) : on étudie les liaisons linéaires entre les variables. Les objectifs sont de résumer la matrice des corrélations et de chercher **des variables synthétiques**:
peut-on résumer les performances des athlètes par un petit nombre de variables ?



Les sorties FactoMineR

Nous utilisons le package FactoMineR pour réaliser l'ACP. FactoMineR se charge de faire les calculs et produit en retour des sorties.

```
premiere_acp <- PCA(X = fifa_100, scale.unit = TRUE, quanti.sup =  
idx_var_quanti_illustratives, quali.sup = idx_var_quali_illustratives, graph = FALSE)
```



Le nombre de composantes à retenir dans l'analyse

```
> summary(premiere_acp)
```

Call:

```
PCA(X = fifa_100, scale.unit = TRUE, quanti.sup = idx_var_quanti_illustratives,  
     quali.sup = idx_var_quali_illustratives, graph = FALSE)
```

Eigenvalues

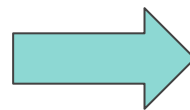
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	Dim.10
Variance	9.547	4.167	1.873	1.047	0.767	0.369	0.285	0.217	0.167	0.156
% of var.	50.248	21.933	9.860	5.509	4.037	1.941	1.501	1.141	0.882	0.819
cumulative % of var.	50.248	72.181	82.041	87.549	91.586	93.527	95.029	96.170	97.051	97.871

D'après la règle de Kaiser-Guttman, on retient les composantes pour lesquelles la valeur propre est supérieure à 1. Dans le cas présent on retient 4 dimensions qui résument 87.5% de l'information contenue dans le jeu de données.

Qu'est-ce qu'une composante ?

Une composante, ou dimension ou variable synthétique est une combinaison linéaire des variables initiales.

short_name	height_cm	weight_kg	pace	shooting	passing	dribbling	defending	physic	attacking_crc	attacking_fin	attacking_he	
L, Messi	170	72		85	92	91	95	34	65	85	95	70
R, Lewandowski	185	81		78	92	79	86	44	82	71	95	90
Cristiano Ronaldo	187	83		87	94	80	88	34	75	87	95	90
Neymar Jr	175	68		91	83	86	94	37	63	85	83	63
K, De Bruyne	181	70		76	86	93	88	64	78	94	82	55
J, Oblak	188	87								13	11	15
K, Mbappé	182	73		97	88	80	92	36	77	78	93	72
M, Neuer	193	93								15	13	25
M, ter Stegen	187	85								18	14	11
H, Kane	188	89		70	91	83	83	47	83	80	94	86
N, Kanté	168	70		78	66	75	82	87	83	68	65	54
K, Benzema	185	81		76	86	81	87	39	77	75	90	89
T, Courtois	199	96								14	14	13
H, Son	183	78		88	87	82	86	43	69	83	88	68
Casemiro	185	84		65	73	76	73	86	90	58	64	79
V, van Dijk	193	92		78	60	71	72	91	84	53	52	87
S, Mané	175	69		91	83	80	89	44	77	78	86	84
M, Salah	175	71		90	87	81	90	45	75	79	91	59
Ederson	188	86								20	14	14
J, Kimmich	177	75		70	73	86	84	83	79	91	68	72
Alisson	191	91								17	13	29
G, Donnarumma	196	90								12	12	12
Sergio Ramos	184	82		70	70	76	74	88	84	66	65	92
L, Suárez	182	83		72	90	82	84	47	83	80	93	84
T, Kroos	183	76		53	81	91	81	71	68	88	77	58
R, Lukaku	191	94		84	87	74	78	39	83	73	92	89
K, Navas	185	80								11	15	11



short_name	Dim1	Dim2	Dim3	Dim4
L, Messi	11,9577324	23,0900908	19,7237335	21,2765171
R, Lewandowski	24,5784033	24,7274475	16,7734192	18,1523701
Cristiano Ronaldo	2,02750981	20,5672244	19,6835708	8,21811876
Neymar Jr	5,68693151	20,7302889	19,0227958	18,0186379
K, De Bruyne	11,7923846	6,8965484	16,1189785	23,5264866
J, Oblak	15,8923083	1,81302832	23,097224	2,1814655
K, Mbappé	16,4599667	3,42803222	9,99121997	7,86402571
M, Neuer	0,75306392	8,3561836	6,61187736	10,900236
M, ter Stegen	2,92968551	18,5117924	16,6864347	16,5757187
H, Kane	2,51419498	24,5391894	13,325366	22,7028922
N, Kanté	13,374679	0,36791324	1,46740428	11,5641022
K, Benzema	21,2974193	2,49217886	14,1726507	11,405643
T, Courtois	11,3032901	24,8942063	21,0588943	22,8101329
H, Son	16,1603037	3,61974624	11,8186012	13,7385729
Casemiro	3,43228534	20,8578322	13,6282035	18,0403019
V, van Dijk	6,51115263	13,1606742	1,11442728	14,0276073
S, Mané	20,3972035	9,40255603	23,9028039	20,0475646
M, Salah	4,10503207	1,72562874	7,00232751	11,2708995
Ederson	19,1217492	13,7058971	14,1510663	24,8889195
J, Kimmich	8,80473875	17,8547656	24,0596595	5,70997578
Alisson	5,76369542	16,0279505	6,75717504	17,134172
G, Donnarumma	7,84091775	24,9830677	10,6682023	13,9699443
Sergio Ramos	8,4660334	8,18929296	4,75425506	14,6610595
L, Suárez	23,355274	17,9138698	1,30501092	8,20962718
T, Kroos	0,61781001	1,00443332	14,8178701	7,29775294
R, Lukaku	12,2447686	3,44468436	17,0289973	13,3057125
K, Navas	12,2676278	14,6338774	17,2906204	4,56248656



Par exemple :

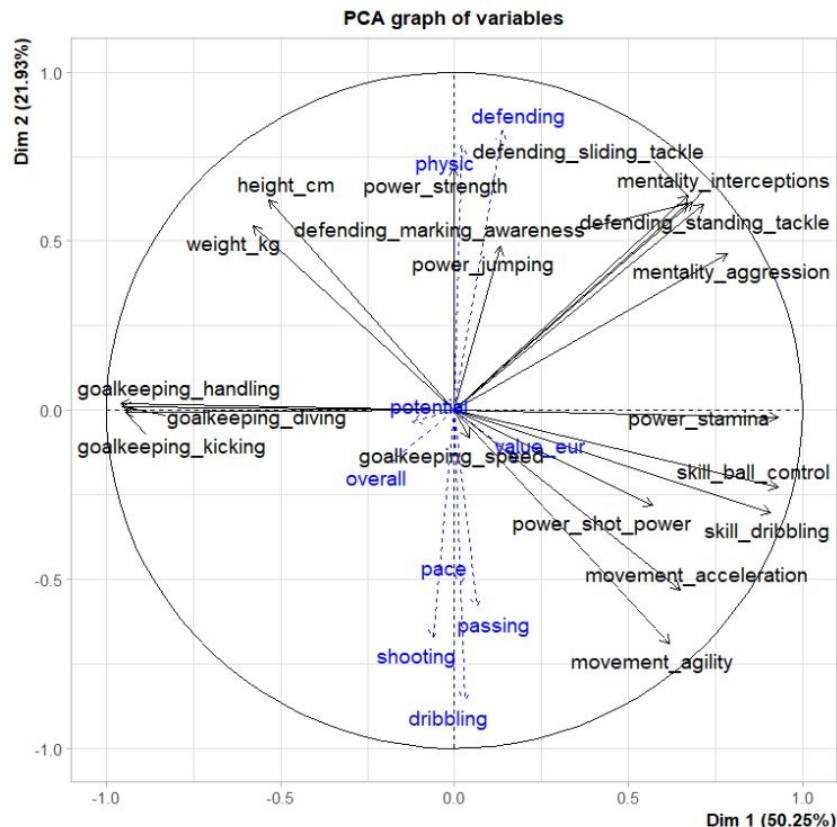
Dim 1 = $0,4 * \text{height_cm} + 0,15 * \text{weight_cm} + 0.35 * \text{shooting} \dots$

À l'intérieur du cercle, il y a des flèches qui partent du centre. Elles sont plus ou moins grandes, et peuvent aller jusqu'à toucher le cercle, sans jamais le dépasser.

Il y a-t-il des variables qui sont corrélées entre elles ?

Comment chaque variable initiale est associée à une

La projection de la flèche (représentant la variable v) sur l'axe Dim1 correspond au coefficient de corrélation entre v et Dim1 . Une flèche nous permet de lire la corrélation de la variable avec deux axes Dim1 et Dim2. Il suffit de lire de gauche à droite et du bas en haut.

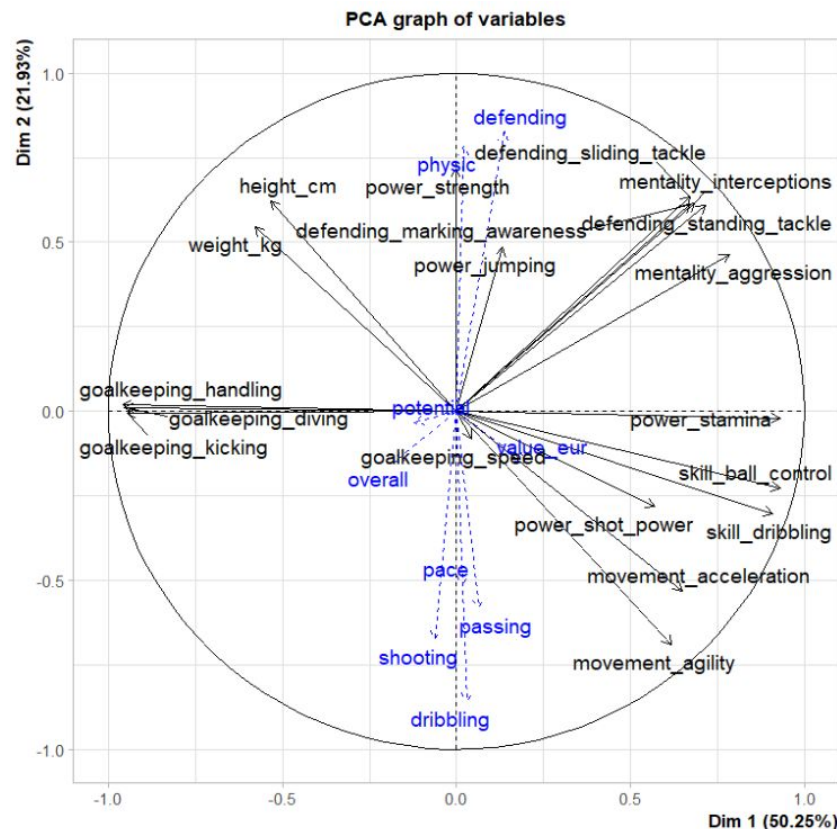




Le graphique des variables (cercle des corrélations)

On peut lire sur ce graphique que weight_kg et heigh_cm évoluent dans le même sens (ils sont corrélés).
weight_kg et height_cm sont négativement corrélés à l'axe 1 (Dim1) mais positivement corrélés à l'axe 2.

Ce graphique se lit de gauche à droite puis de bas en haut.



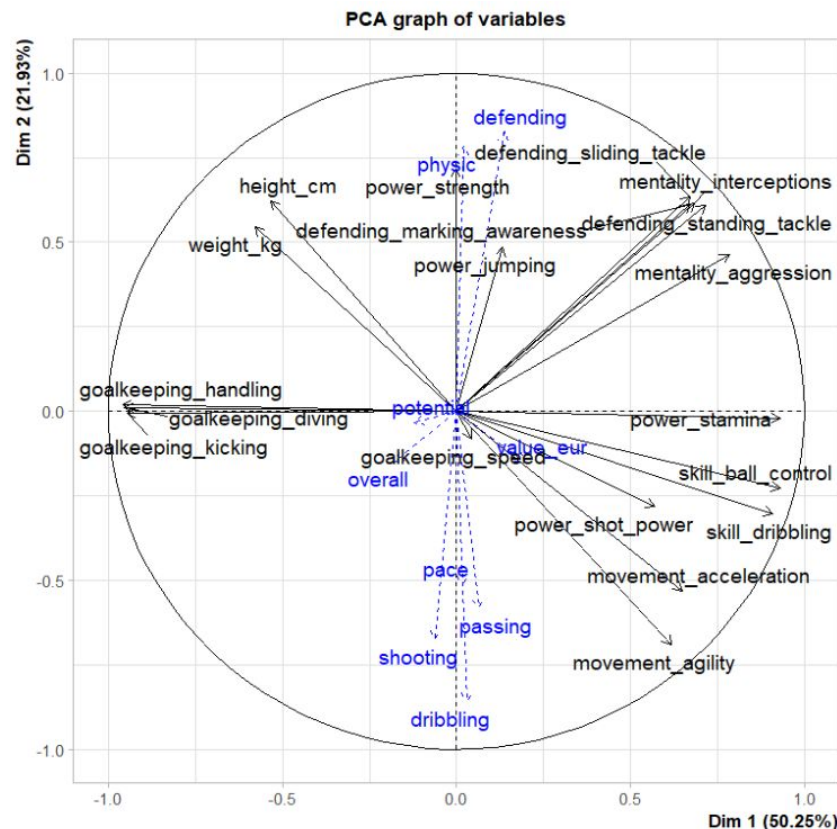


Le graphique des variables (cercle des corrélations)

Les variables les plus corrélés positivement à l'axe 1 sont :
power_stamina, skill_ball_control, skill_dribbling ...

Les variables les plus corrélés positivement à l'axe 2 sont :
power_shot_power, power_strength, weight_kg ...

Nous pouvons afficher les corrélations exactes de chacune
des variables à chaque axe en utilisant la fonction
dimdesc().





Les sorties de la fonction dimdesc

La fonction `dimdesc()` calcule le coefficient de corrélation entre une variable et une dimension et réalise un test de significativité.

Ces tableaux donnent le coefficient de corrélation et la probabilité critique des variables qui sont significativement corrélées aux dimensions principales. Les variables actives et illustratives dont la probabilité critique est inférieure à 0.05 apparaissent.



```
> dimdesc(premiere_acp)
$Dim.1
```

Link between the variable and the continuous variables (R-square)

	correlation	p.value
power_stamina	0.9310608	1.057853e-44
skill_ball_control	0.9296284	2.798423e-44
skill_dribbling	0.9101928	2.692916e-39
mentality_aggression	0.7834201	5.824363e-22
defending_standing_tackle	0.7160741	5.565933e-17
mentality_interceptions	0.6825780	5.238686e-15
defending_sliding_tackle	0.6715655	2.051306e-14
defending_marking_awareness	0.6704033	2.361122e-14
movement_acceleration	0.6492827	2.737652e-13
movement_agility	0.6184369	7.073146e-12
power_shot_power	0.5701057	5.968991e-10
value_eur	0.1972808	4.913873e-02
height_cm	-0.5324573	1.185330e-08
weight_kg	-0.5794330	2.682184e-10
goalkeeping_kicking	-0.9460502	9.222093e-50
goalkeeping_diving	-0.9542690	3.417323e-53
goalkeeping_handling	-0.9598525	6.620247e-56



\$Dim.2

Link between the variable and the continuous variables (R-square)

	correlation	p.value
defending	0.8266670	3.253779e-26
physic	0.7826747	6.758755e-22
power_strength	0.7172355	4.699250e-17
defending_sliding_tackle	0.6344015	1.375379e-12
height_cm	0.6226199	4.647055e-12
mentality_interceptions	0.6147787	1.016163e-11
defending_marking_awareness	0.6110181	1.467626e-11
defending_standing_tackle	0.6102678	1.578390e-11
weight_kg	0.5443635	4.792828e-09
power_jumping	0.4843844	3.285453e-07
mentality_aggression	0.4630878	1.225685e-06
skill_ball_control	-0.2283309	2.232021e-02
power_shot_power	-0.2829131	4.343500e-03
skill_dribbling	-0.3038003	2.120761e-03
pace	-0.5136892	4.610158e-08
movement_acceleration	-0.5320392	1.222851e-08
passing	-0.5840214	1.792800e-10
shooting	-0.6751369	1.326081e-14
movement_agility	-0.6916079	1.634877e-15
dribbling	-0.8601548	2.059502e-30



\$Dim.3

Link between the variable and the continuous variables (R-square)

```
=====
correlation  p.value
power_shot_power  0.5965309 5.786781e-11
power_strength    0.5895215 1.096965e-10
weight_kg         0.4901275 2.268130e-07
shooting          0.4749392 5.956005e-07
power_jumping     0.4214601 1.260597e-05
height_cm         0.3741842 1.255169e-04
physic            0.3677805 1.669526e-04
potential         0.2473631 1.309335e-02
overall           0.2446758 1.415090e-02
pace              0.2335800 1.934134e-02
value_eur         0.2174076 2.979520e-02
passing           -0.2129004 3.344647e-02
defending_standing_tackle -0.2886445 3.586468e-03
defending_sliding_tackle -0.3325388 7.240194e-04
mentality_interceptions -0.3378593 5.864856e-04
defending_marking_awareness -0.3583304 2.516312e-04
defending         -0.5012765 1.083267e-07
```

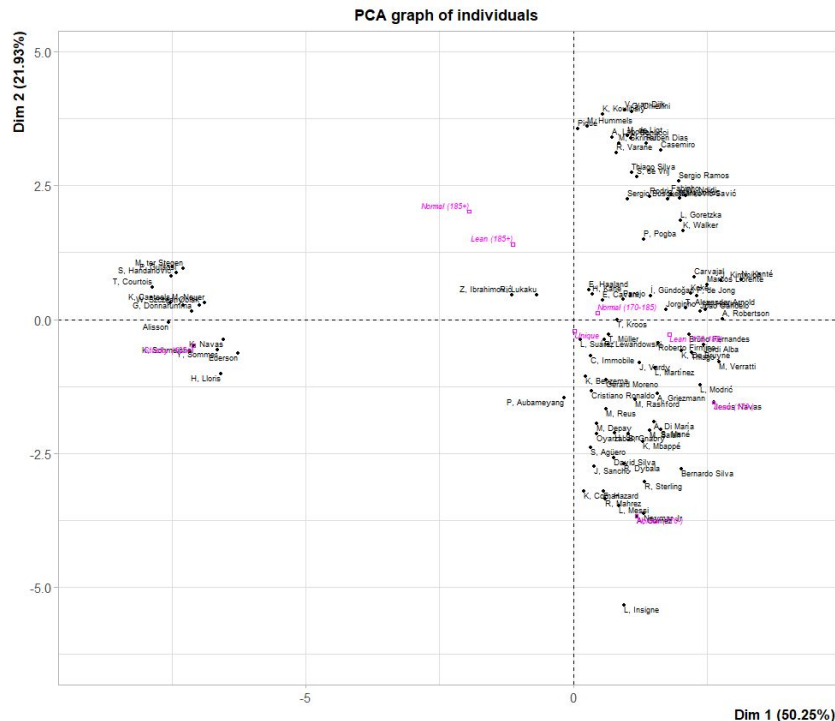


Le graphique des individus

```
plot(premiere_acp, choix = "ind", axes = c(1, 2), cex = 0.5
```

Le graphique des individus est similaire en essence au cercle des corrélations. Il montre comment sont positionnés les individus selon le sens des variables.

Nous avons vu que height_cm et weight_kg étaient négativement corrélés à Dim1 et positivement corrélés à Dim2. Dans ce cas on peut dire Ter stegen est grand et pèse lourd.





```
plot(premiere_acp, choix = "ind", axes = c(1, 2), cex = 0.5)
```

