Analyse en Composantes Principales

Jeu de données fifa\_100

Ce jeu de données contient 100 individus et 29 variables, 14 variables quantitatives sont illustratives, 1 variable qualitative est illustrative.

### 1. Observation d’individus extrêmes

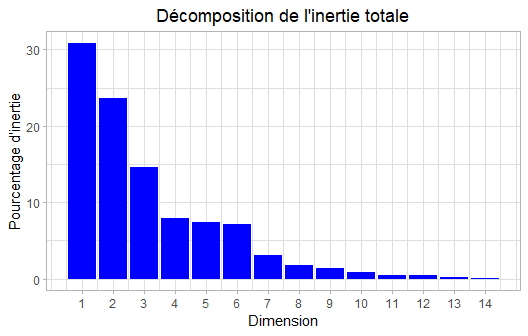
L’analyse des graphes ne révèle aucun individu singulier.

### 2. Distribution de l’inertie

L’inertie des axes factoriels indique d’une part si les variables sont structurées et suggère d’autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l’ analyse expriment **54.53%** de l’inertie totale du jeu de données ; cela signifie que 54.53% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C’est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans une grande part du jeu de données actif. Cette valeur est nettement supérieure à la valeur référence de **24.48%**, la variabilité expliquée par ce plan est donc hautement significative (cette intertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d’inertie obtenue en simulant 2809 jeux de données aléatoires de dimensions comparables sur la base d’une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l’analyse les dimensions supérieures ou égales à la troisième.

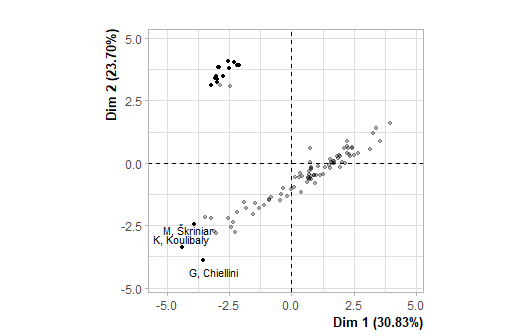


**Figure 2 - Décomposition de l’inertie totale**

Une estimation du nombre pertinent d’axes à interpréter suggère de restreindre l’analyse à la description des 3 premiers axes. Ces composantes révèlent un taux d’inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (69.2% contre 34.61%). Cette observation suggère que seuls ces axes sont porteurs d’une véritable information. En conséquence, la description de l’analyse sera restreinte à ces seuls axes.

### 3. Description du plan 1:2

## Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider increasing  
## max.overlaps



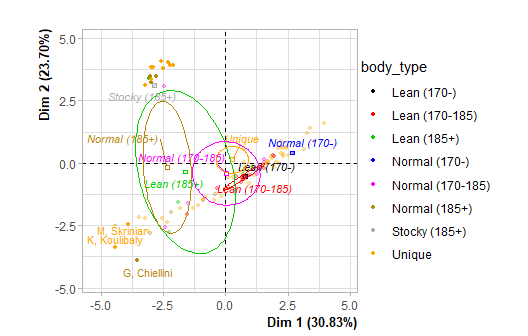
**Figure 3.1 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*

La probabilité critique du test de Wilks indique la variable dont les modalités sépare au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

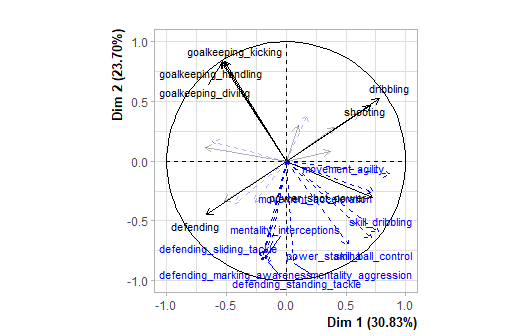
## body\_type   
## 0.02668305

Il n’y a qu’une unique variable qualitative possible pour illustrer les distances entre individus : *body\_type*.

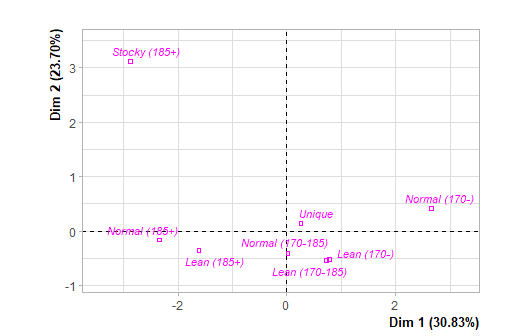
## Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider increasing  
## max.overlaps



**Figure 3.2 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.* *Les individus sont colorés selon leur appartenance aux modalités de la variable* body\_type.



**Figure 3.3 - Graphe des variables (ACP)** *Les variables en noir sont les variables actives, celles en bleu sont les variables quantitatives supplémentaires (illustratives).* *Les variables libellées sont celles les mieux représentées sur le plan.*



**Figure 3.4 - Graphe des modalités (ACP)** *Les facteurs libellés sont ceux les mieux représentés sur le plan.*

La **dimension 1** oppose des individus caractérisés par une coordonnée fortement positive sur l’axe (à droite du graphe) à des individus comme *Alisson*, *W, Szczęsny*, *G, Chiellini*, *K, Navas*, *J, Oblak*, *T, Courtois*, *G, Donnarumma*, *S, Handanovič*, *K, Koulibaly* et *K, Casteels* (à gauche du graphe, caractérisés par une coordonnée fortement négative sur l’axe).

Le groupe 1 (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour des variables telles que *movement\_agility*, *movement\_acceleration*, *dribbling*, *skill\_dribbling*, *power\_shot\_power*, *shooting*, *skill\_ball\_control*, *pace*, *passing* et *power\_stamina* (de la plus extrême à la moins extrême).
* de faibles valeurs pour des variables telles que *defending*, *height\_cm*, *weight\_kg*, *physic*, *power\_strength*, *goalkeeping\_handling*, *goalkeeping\_diving*, *goalkeeping\_kicking*, *power\_jumping* et *defending\_sliding\_tackle* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *G, Chiellini*, *K, Koulibaly* et *M, Škriniar* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de fortes valeurs pour des variables telles que *mentality\_interceptions*, *defending\_sliding\_tackle*, *defending\_marking\_awareness*, *defending\_standing\_tackle*, *defending*, *mentality\_aggression*, *power\_strength*, *physic*, *power\_jumping* et *height\_cm* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *dribbling*, *shooting*, *passing*, *pace*, *goalkeeping\_kicking*, *goalkeeping\_diving*, *goalkeeping\_handling*, *movement\_agility* et *overall* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *Alisson*, *W, Szczęsny*, *K, Navas*, *J, Oblak*, *T, Courtois*, *G, Donnarumma*, *S, Handanovič*, *K, Casteels*, *M, Neuer* et *M, ter Stegen* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de fortes valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_handling*, *goalkeeping\_kicking*, *weight\_kg* et *height\_cm* (de la plus extrême à la moins extrême).
* de faibles valeurs pour des variables telles que *skill\_ball\_control*, *skill\_dribbling*, *power\_stamina*, *mentality\_aggression*, *movement\_acceleration*, *defending\_standing\_tackle*, *power\_shot\_power*, *defending\_marking\_awareness*, *mentality\_interceptions* et *defending\_sliding\_tackle* (de la plus extrême à la moins extrême).

Notons que la variable *Normal (185+)* est extrêmement corrélée à cette dimension (corrélation de 0.91). Cette variable pourrait donc résumer à elle seule la dimension 1.

La **dimension 2** oppose des individus tels que *Alisson*, *W, Szczęsny*, *K, Navas*, *J, Oblak*, *T, Courtois*, *G, Donnarumma*, *S, Handanovič*, *K, Casteels*, *M, Neuer* et *M, ter Stegen* (en haut du graphe, caractérisés par une coordonnées fortement positive sur l’axe) à des individus comme *G, Chiellini*, *K, Koulibaly* et *M, Škriniar* (en bas du graphe, caractérisés par une coordonnées fortement négative sur l’axe).

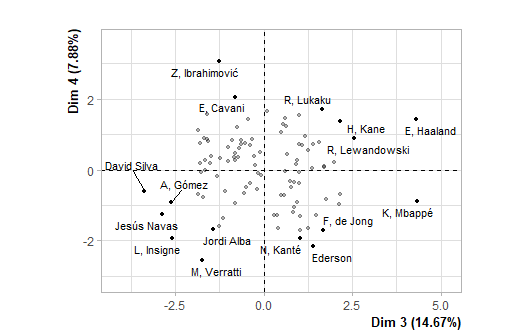
Le groupe auquel les individus *Alisson*, *W, Szczęsny*, *K, Navas*, *J, Oblak*, *T, Courtois*, *G, Donnarumma*, *S, Handanovič*, *K, Casteels*, *M, Neuer* et *M, ter Stegen* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_handling*, *goalkeeping\_kicking*, *weight\_kg* et *height\_cm* (de la plus extrême à la moins extrême).
* de faibles valeurs pour des variables telles que *skill\_ball\_control*, *skill\_dribbling*, *power\_stamina*, *mentality\_aggression*, *movement\_acceleration*, *defending\_standing\_tackle*, *power\_shot\_power*, *defending\_marking\_awareness*, *mentality\_interceptions* et *defending\_sliding\_tackle* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *G, Chiellini*, *K, Koulibaly* et *M, Škriniar* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de fortes valeurs pour des variables telles que *mentality\_interceptions*, *defending\_sliding\_tackle*, *defending\_marking\_awareness*, *defending\_standing\_tackle*, *defending*, *mentality\_aggression*, *power\_strength*, *physic*, *power\_jumping* et *height\_cm* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *dribbling*, *shooting*, *passing*, *pace*, *goalkeeping\_kicking*, *goalkeeping\_diving*, *goalkeeping\_handling*, *movement\_agility* et *overall* (de la plus extrême à la moins extrême).

### 4. Description de la dimension 3

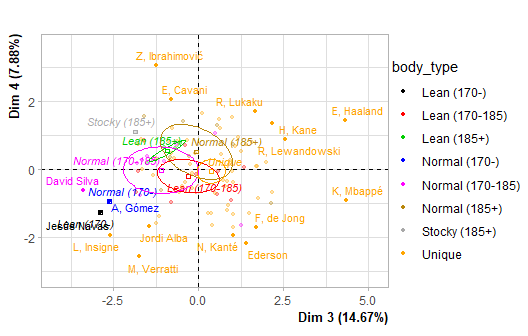


**Figure 4.1 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*

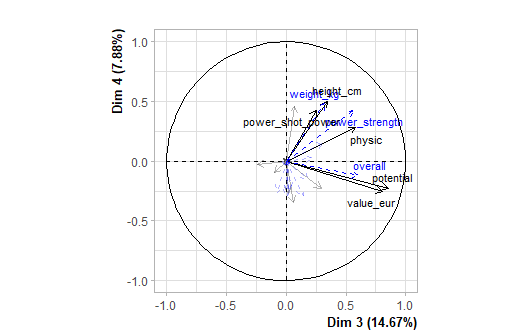
La probabilité critique du test de Wilks indique la variable dont les modalités sépare au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

## body\_type   
## 0.04189557

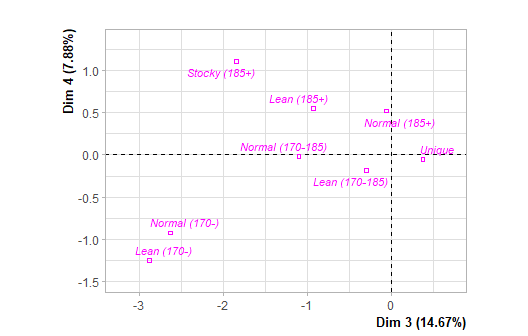
Il n’y a qu’une unique variable qualitative possible pour illustrer les distances entre individus : *body\_type*.



**Figure 4.2 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.* *Les individus sont colorés selon leur appartenance aux modalités de la variable* body\_type.



**Figure 4.3 - Graphe des variables (ACP)** *Les variables en noir sont les variables actives, celles en bleu sont les variables quantitatives supplémentaires (illustratives).* *Les variables libellées sont celles les mieux représentées sur le plan.*



**Figure 4.4 - Graphe des modalités (ACP)** *Les facteurs libellés sont ceux les mieux représentés sur le plan.*

La **dimension 3** oppose des individus tels que *E, Haaland*, *R, Lukaku*, *R, Lewandowski*, *F, de Jong*, *H, Kane*, *K, Mbappé* et *N, Kanté* (à droite du graphe, caractérisés par une coordonnée fortement positive sur l’axe) à des individus comme *Z, Ibrahimović*, *Jesús Navas*, *David Silva*, *Jordi Alba*, *M, Verratti*, *E, Cavani*, *A, Gómez* et *L, Insigne* (à gauche du graphe, caractérisés par une coordonnée fortement négative sur l’axe).

Le groupe auquel les individus *F, de Jong* et *N, Kanté* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour les variables *defending\_marking\_awareness*, *defending*, *power\_stamina*, *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *mentality\_interceptions*, *potential* et *value\_eur* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *shooting* et *power\_shot\_power* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *R, Lukaku*, *R, Lewandowski* et *H, Kane* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour les variables *goalkeeping\_speed*, *power\_strength*, *power\_shot\_power*, *height\_cm*, *physic*, *weight\_kg*, *overall*, *power\_jumping* et *shooting* (de la plus extrême à la moins extrême).
* de faibles valeurs pour la variable *movement\_agility*.

Le groupe auquel les individus *E, Haaland* et *K, Mbappé* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour les variables *value\_eur*, *potential* et *pace* (de la plus extrême à la moins extrême).
* de faibles valeurs pour la variable *passing*.

Le groupe auquel les individus *Jesús Navas*, *David Silva*, *Jordi Alba*, *M, Verratti*, *A, Gómez* et *L, Insigne* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de fortes valeurs pour les variables *movement\_agility*, *dribbling*, *skill\_ball\_control* et *skill\_dribbling* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *height\_cm*, *power\_strength*, *physic*, *weight\_kg*, *power\_shot\_power*, *potential*, *value\_eur*, *shooting*, *power\_jumping* et *overall* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *Z, Ibrahimović* et *E, Cavani* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de faibles valeurs pour les variables *potential*, *pace*, *overall*, *movement\_acceleration* et *value\_eur* (de la plus extrême à la moins extrême).

## Annexes