Projet LSTAT2110(A) – Analyse de données

BARTHELEMI Lauriane , 56281400, BIRE2M, MARTELEE Baudouin , 43191400, DATS2M

Table of Contents

Introduction

La violence est un excellent moyen d'extérioriser tous les tracas que nous gardons profondément au fond de nous. Cependant, il ne faut pas qu'elle soit destructrice et cela implique qu'elle soit canalisée. On peut alors l'exprimer sous différentes formes tels que les sports de combat. Tous deux passionnés de jiu jitsu brésilien et membres du même club, nous avons décidé de nous pencher un peu plus sur le sujet.

Présentation des données, analyse descriptive

Description du dataset et motivation de l'utilisation de cette base de données

Le dataset porte le nom de dataset UFC. L'UFC ou de son nom entier l'Ultimate Fighting Championship est la plus grande organisation au monde de sport de combat Mixed Martial Art (MMA).

Cette base de données contient 5144 combats qui se sont déroulées entre 1993 et 2019. Elle comprend pour chaque combat, 2 combattants qui, pour être distingués portent les noms de : combattant Bleu (B_fighter) et combattant Rouge (R_fighter). On a différentes informations à propos de l'organisation du combat tel que la date du combat, l'endroit où se déroule celui-ci, la catégorie de poids , le nombre de rounds que comporte le combat, l'arbitre qui régulera le combat ou bien encore s'il s'agit d'un combat pour le titre de champion du monde. En plus de cela, on peut trouver, pour chaque combattant, des statistiques personelles physiologiques ou encore des performances réalisées lors du combat.

Avant de commencer le projet, il faut mettre en avant comment se déroule un combat de mma. Ce dernier implique 2 combattants et est régulé par un arbitre. Le combat prend place dans une cage appelé octogone (car cage de 8 cotés) et comporte 3 rounds de 5 minutes excepté pour un combat d'un titre de champion du monde, il est alors de 5 rounds de 5 minutes. Le mma est composé de 70 règles mais celles-ci ne sont pas importantes afin de comprendre ce projet. Cependant, il faut comprendre que le combat peut se dérouler en 3 phases à chaque moment.

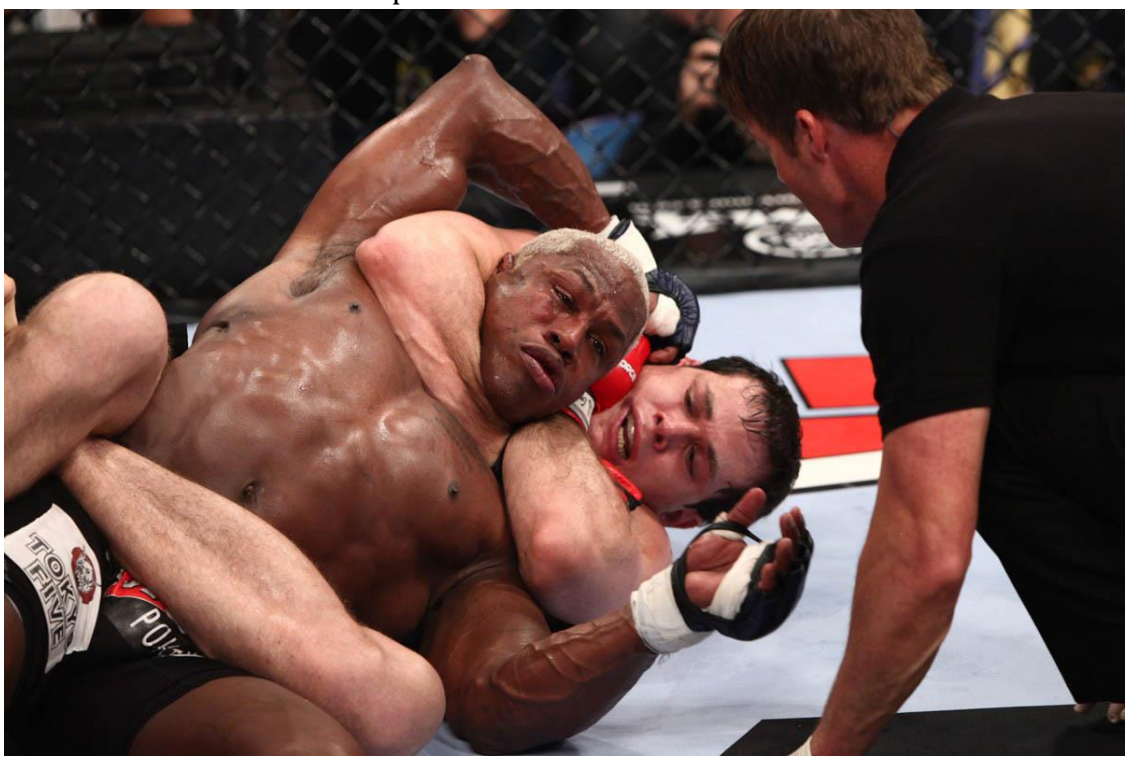
1) Le striking (combat pied, poings, genoux, coup de coude) se pratique en position debout. Il admet la possibilité de mettre un adversaire KO.

Striking

2) La takedown (prises de lutte ou de judo) permet de déstabiliser l'adversaire et de soit l'amener au sol soit le maintenir au sol.

Takedown

3) La submission (prise de jiujitsu brésilien ou luta livre) permet de garder son adversaire au sol et de le soumettre en utilisant par exemple des étranglements, des clefs ou encore des compressions musculaires.



Submission

Il y a plusieurs manières de gagner un combat :

1) Gagner par arrêt du docteur
2) Gagner par KO technique (Soit l'adversaire est ko soit l'arbitre arrête le combat car le combattant est quasi ko).
3) Gagner par soumission
4) Gagner par décision (Il y a 2 types de décisions). 4.1) Gagner par décision unanime. 4.2) Gagner par split décision.

Si le combat ne se termine pas par un ko, une soumission ou un arret du docteur après les 3 ou 5 rounds, ce sont les 3 juges qui prennent la décision de nommer le vainqueur. Les décisions se prennent en fonction du nombre d'actions tentées par chaque adversaire. A chaque action correspond un nombre de points bien précis.

Maintenant que tout est clarifié, nous pouvons commencer le projet en important les données.

Importation des données

```
rm(list = ls())
# Import des données à partir du fichier csv
dataUFC <- read.csv('data.csv',header = TRUE,sep = ',', dec = '.')
# Permet d'afficher les 2 dimensions du dataset, çad le nombre de combats
avec le nombre de variables pour chaque combat.
dim(dataUFC)
```

```
[1] 5144  145
```

```
# Affiche les 5 permières lignes du dataset
head(dataUFC, n = 5)
```

```
             R_fighter      B_fighter         Referee       date
1         Henry Cejudo  Marlon Moraes    Marc Goddard 2019-06-08
2 Valentina Shevchenko    Jessica Eye Robert Madrigal 2019-06-08
3        Tony Ferguson Donald Cerrone Dan Miragliotta 2019-06-08
4        Jimmie Rivera       Petr Yan Kevin MacDonald 2019-06-08
5          Tai Tuivasa  Blagoy Ivanov Dan Miragliotta 2019-06-08
                location Winner title_bout      weight_class no_of_rounds
1 Chicago, Illinois, USA    Red       True      Bantamweight            5
2 Chicago, Illinois, USA    Red       True Women's Flyweight            5
3 Chicago, Illinois, USA    Red      False       Lightweight            3
4 Chicago, Illinois, USA   Blue      False      Bantamweight            3
5 Chicago, Illinois, USA   Blue      False       Heavyweight            3
  B_current_lose_streak B_current_win_streak B_draw B_avg_BODY_att
1                     0                    4      0        9.20000
2                     0                    3      0       14.60000
3                     0                    3      0       15.35484
4                     0                    4      0       17.00000
5                     0                    1      0       17.00000
  B_avg_BODY_landed B_avg_CLINCH_att B_avg_CLINCH_landed B_avg_DISTANCE_att
1           6.00000         0.200000            0.000000           62.60000
```

```
2           9.10000        11.800000            7.300000          124.70000
3          11.32258         6.741935            4.387097           84.74194
4          14.00000        13.750000           11.000000          109.50000
5          14.50000         2.500000            2.000000          201.00000
  B_avg_DISTANCE_landed B_avg_GROUND_att B_avg_GROUND_landed B_avg_HEAD_att
1              20.60000         2.600000             2.000000       48.60000
2              42.10000         2.400000             1.900000      112.00000
3              38.58065         5.516129             3.806452       67.64516
4              48.75000        13.000000            10.500000      116.25000
5              59.50000         0.000000             0.000000      184.50000
  B_avg_HEAD_landed  B_avg_KD B_avg_LEG_att B_avg_LEG_landed B_avg_PASS
1          11.20000 0.8000000           7.6          5.40000  0.4000000
2          32.00000 0.0000000          12.3         10.20000  0.8000000
3          23.25806 0.6451613          14.0         12.19355  0.9354839
4          53.75000 0.5000000           3.0          2.50000  0.5000000
5          45.00000 0.0000000           2.0          2.00000  0.0000000
   B_avg_REV B_avg_SIG_STR_att B_avg_SIG_STR_landed B_avg_SIG_STR_pct
1 0.00000000             65.40             22.60000          0.466000
2 0.00000000            138.90             51.30000          0.399000
3 0.09677419             97.00             46.77419          0.496129
4 0.25000000            136.25             70.25000          0.550000
5 0.00000000            203.50             61.50000          0.310000
  B_avg_SUB_ATT B_avg_TD_att B_avg_TD_landed B_avg_TD_pct B_avg_TOTAL_STR_att
1     0.4000000      0.80000       0.2000000    0.1000000              66.4000
2     0.7000000      1.00000       0.5000000    0.2250000             158.7000
3     0.3548387      2.16129       0.6774194    0.2954839             103.7097
4     0.2500000      2.50000       1.2500000    0.2875000             154.7500
5     0.0000000      0.00000       0.0000000    0.0000000             204.0000
  B_avg_TOTAL_STR_landed B_longest_win_streak B_losses B_avg_opp_BODY_att
1               23.60000                    4        1            6.40000
2               69.60000                    3        6           13.00000
3               52.54839                    8        8           17.90323
4               86.75000                    4        0           12.25000
5               62.00000                    1        1           42.50000
  B_avg_opp_BODY_landed B_avg_opp_CLINCH_att B_avg_opp_CLINCH_landed
1               4.00000             1.000000                 0.60000
2               9.30000            12.800000                 9.60000
3              11.87097             8.419355                 5.83871
4               6.00000             6.000000                 3.75000
5              23.50000             0.500000                 0.50000
  B_avg_opp_DISTANCE_att B_avg_opp_DISTANCE_landed B_avg_opp_GROUND_att
1               51.20000                 17.40000            0.600000
2              101.70000                 32.00000            8.100000
3               84.54839                 38.06452            1.741935
4               94.25000                 26.75000            1.750000
5              205.00000                 89.50000            0.000000
  B_avg_opp_GROUND_landed B_avg_opp_HEAD_att B_avg_opp_HEAD_landed
B_avg_opp_KD
1               0.2000000           39.60000              9.40000
0.2000000
```

```
2               6.9000000           97.70000              30.80000
0.1000000
3               0.9354839           67.64516              25.48387
0.2258065
4               1.2500000           82.50000              21.50000
0.2500000
5               0.0000000          152.50000              56.50000
0.0000000
  B_avg_opp_LEG_att B_avg_opp_LEG_landed B_avg_opp_PASS B_avg_opp_REV
1           6.80000             4.800000     0.00000000    0.00000000
2          11.90000             8.400000     1.40000000    0.00000000
3           9.16129             7.483871     0.03225806    0.03225806
4           7.25000             4.250000     0.00000000    0.00000000
5          10.50000            10.000000     0.00000000    0.00000000
  B_avg_opp_SIG_STR_att B_avg_opp_SIG_STR_landed B_avg_opp_SIG_STR_pct
1              52.80000                 18.20000             0.2360000
2             122.60000                 48.50000             0.4080000
3              94.70968                 44.83871             0.4532258
4             102.00000                 31.75000             0.3375000
5             205.50000                 90.00000             0.4300000
  B_avg_opp_SUB_ATT B_avg_opp_TD_att B_avg_opp_TD_landed B_avg_opp_TD_pct
1        0.00000000         1.000000           0.4000000       0.10000000
2        0.70000000         2.300000           0.9000000       0.23100000
3        0.09677419         2.096774           0.2258065       0.06354839
4        0.00000000         4.500000           0.7500000       0.09750000
5        0.00000000         0.500000           0.0000000       0.00000000
  B_avg_opp_TOTAL_STR_att B_avg_opp_TOTAL_STR_landed B_total_rounds_fought
1                  53.8000                   19.20000                     9
2                 151.5000                   75.40000                    29
3                 100.3871                   49.77419                    68
4                 104.7500                   34.25000                     9
5                 205.5000                   90.00000                     8
  B_total_time_fought.seconds. B_total_title_bouts B_win_by_Decision_Majority
1                      419.400                   0                          0
2                      849.000                   0                          0
3                      581.871                   1                          0
4                      652.000                   0                          0
5                     1200.000                   0                          0
  B_win_by_Decision_Split B_win_by_Decision_Unanimous B_win_by_KO.TKO
1                       1                           0               2
2                       2                           1               0
3                       0                           7              10
4                       0                           2               2
5                       0                           1               0
  B_win_by_Submission B_win_by_TKO_Doctor_Stoppage B_wins B_Stance
B_Height_cms
1                   1                           0      4 Orthodox
167.64
2                   0                           1      4 Orthodox
167.64
```

```
3                   6                            0     23 Orthodox
185.42
4                   0                            0      4   Switch
170.18
5                   0                            0      1 Southpaw
180.34
  B_Reach_cms B_Weight_lbs R_current_lose_streak R_current_win_streak R_draw
1      170.18          135                     0                     4      0
2      167.64          125                     0                     2      0
3      185.42          155                     0                    11      0
4      170.18          135                     1                     0      0
5      185.42          250                     1                     0      0
  R_avg_BODY_att R_avg_BODY_landed R_avg_CLINCH_att R_avg_CLINCH_landed
1       21.90000         16.400000        17.000000           11.000000
2       12.00000          7.714286         9.285714            6.857143
3       13.86667          8.666667         2.866667            1.733333
4       18.25000         10.250000         5.875000            4.125000
5        7.75000          6.750000        11.000000            7.250000
  R_avg_DISTANCE_att R_avg_DISTANCE_landed R_avg_GROUND_att
R_avg_GROUND_landed
1           75.00000              26.50000         9.400000
6.500000
2           88.14286              36.14286        18.428571
16.428571
3          116.13333              49.46667         5.333333
4.266667
4          104.87500              41.00000         1.000000
0.625000
5           50.75000              24.75000         0.500000
0.500000
  R_avg_HEAD_att R_avg_HEAD_landed R_avg_KD R_avg_LEG_att R_avg_LEG_landed
1       74.20000             23.90    0.400       5.30000          3.70000
2       84.57143             37.00    0.000      19.28571         14.71429
3       96.73333             35.60    0.200      13.73333         11.20000
4       80.50000             24.00    0.375      13.00000         11.50000
5       50.75000             22.75    0.500       3.75000          3.00000
  R_avg_PASS R_avg_REV R_avg_SIG_STR_att R_avg_SIG_STR_landed
R_avg_SIG_STR_pct
1  1.2000000 0.0000000          101.4000             44.00000
0.4660000
2  1.7142857 0.1428571          115.8571             59.42857
0.5757143
3  0.3333333 0.1333333          124.3333             55.46667
0.4300000
4  0.1250000 0.0000000          111.7500             45.75000
0.3662500
5  0.2500000 0.0000000           62.2500             32.50000
0.5450000
  R_avg_SUB_ATT R_avg_TD_att R_avg_TD_landed R_avg_TD_pct R_avg_TOTAL_STR_att
1     0.1000000    5.3000000        1.900000    0.4580000            129.9000
```

```
2     0.4285714    5.1428571        2.428571    0.6014286            161.5714
3     1.0000000    0.9333333        0.400000    0.2773333            133.0000
4     0.0000000    2.2500000        0.625000    0.1037500            117.3750
5     0.0000000    0.5000000        0.000000    0.0000000             63.5000
  R_avg_TOTAL_STR_landed R_longest_win_streak R_losses R_avg_opp_BODY_att
1                69.1000                    4        2           13.30000
2               102.8571                    2        2           24.57143
3                63.4000                   11        1           14.46667
4                50.7500                    5        2           20.25000
5                32.7500                    3        1            6.25000
  R_avg_opp_BODY_landed R_avg_opp_CLINCH_att R_avg_opp_CLINCH_landed
1              8.800000              7.50000               5.1000000
2             14.142857             10.57143               7.8571429
3              8.133333              2.80000               0.7333333
4             13.375000              6.87500               5.6250000
5              4.750000              4.50000               3.5000000
  R_avg_opp_DISTANCE_att R_avg_opp_DISTANCE_landed R_avg_opp_GROUND_att
1               90.50000                  26.80000             0.800000
2               98.57143                  32.57143             6.428571
3               91.06667                  32.20000             4.866667
4              103.12500                  38.50000             0.875000
5               42.75000                  16.25000             7.750000
  R_avg_opp_GROUND_landed R_avg_opp_HEAD_att R_avg_opp_HEAD_landed
R_avg_opp_KD
1                0.300000           76.10000              17.30000
0.1000000
2                4.285714           61.85714              12.42857
0.0000000
3                2.800000           78.26667              23.20000
0.2666667
4                0.750000           77.37500              20.37500
0.1250000
5                2.750000           43.25000              14.00000
0.2500000
  R_avg_opp_LEG_att R_avg_opp_LEG_landed R_avg_opp_PASS R_avg_opp_REV
1           9.40000              6.10000      0.0000000     0.0000000
2          29.14286             18.14286      1.1428571     0.0000000
3           6.00000              4.40000      0.3333333     0.1333333
4          13.25000             11.12500      0.0000000     0.0000000
5           5.50000              3.75000      0.7500000     0.0000000
  R_avg_opp_SIG_STR_att R_avg_opp_SIG_STR_landed R_avg_opp_SIG_STR_pct
1              98.80000                 32.20000              0.3360000
2             115.57143                 44.71429              0.4371429
3              98.73333                 35.73333              0.3400000
4             110.87500                 44.87500              0.4462500
5              55.00000                 22.50000              0.3975000
  R_avg_opp_SUB_ATT R_avg_opp_TD_att R_avg_opp_TD_landed R_avg_opp_TD_pct
1        0.00000000         0.900000           0.1000000        0.0500000
2        0.28571429         3.285714           0.8571429        0.1471429
3        0.06666667         2.866667           0.6666667        0.1313333
```

```
4        0.00000000         2.375000           0.0000000        0.0000000
5        0.00000000         1.000000           0.0000000        0.0000000
  R_avg_opp_TOTAL_STR_att R_avg_opp_TOTAL_STR_landed R_total_rounds_fought
1                110.5000                   43.30000                    27
2                158.1429                   82.28571                    25
3                102.1333                   38.60000                    33
4                115.1250                   48.87500                    20
5                 60.5000                   27.75000                     7
  R_total_time_fought.seconds. R_total_title_bouts R_win_by_Decision_Majority
1                       742.60                   3                          0
2                      1062.00                   2                          0
3                       604.40                   2                          0
4                       690.25                   0                          0
5                       440.75                   0                          0
  R_win_by_Decision_Split R_win_by_Decision_Unanimous R_win_by_KO.TKO
1                       2                           4               2
2                       1                           2               0
3                       1                           3               3
4                       1                           4               1
5                       0                           1               2
  R_win_by_Submission R_win_by_TKO_Doctor_Stoppage R_wins R_Stance
R_Height_cms
1                   0                           0      8 Orthodox
162.56
2                   2                           0      5 Southpaw
165.10
3                   6                           1     14 Orthodox
180.34
4                   0                           0      6 Orthodox
162.56
5                   0                           0      3 Southpaw
187.96
  R_Reach_cms R_Weight_lbs B_age R_age
1      162.56          135    31    32
2      167.64          125    32    31
3      193.04          155    36    35
4      172.72          135    26    29
5      190.50          264    32    26
```

```r
# Visualisation des différentes colonnes
colnames(dataUFC)
```

```
  [1] "R_fighter"                   "B_fighter"
  [3] "Referee"                     "date"
  [5] "location"                    "Winner"
  [7] "title_bout"                  "weight_class"
  [9] "no_of_rounds"                "B_current_lose_streak"
 [11] "B_current_win_streak"        "B_draw"
 [13] "B_avg_BODY_att"              "B_avg_BODY_landed"
 [15] "B_avg_CLINCH_att"            "B_avg_CLINCH_landed"
```

```
 [17] "B_avg_DISTANCE_att"           "B_avg_DISTANCE_landed"
 [19] "B_avg_GROUND_att"             "B_avg_GROUND_landed"
 [21] "B_avg_HEAD_att"               "B_avg_HEAD_landed"
 [23] "B_avg_KD"                     "B_avg_LEG_att"
 [25] "B_avg_LEG_landed"             "B_avg_PASS"
 [27] "B_avg_REV"                    "B_avg_SIG_STR_att"
 [29] "B_avg_SIG_STR_landed"         "B_avg_SIG_STR_pct"
 [31] "B_avg_SUB_ATT"                "B_avg_TD_att"
 [33] "B_avg_TD_landed"              "B_avg_TD_pct"
 [35] "B_avg_TOTAL_STR_att"          "B_avg_TOTAL_STR_landed"
 [37] "B_longest_win_streak"         "B_losses"
 [39] "B_avg_opp_BODY_att"           "B_avg_opp_BODY_landed"
 [41] "B_avg_opp_CLINCH_att"         "B_avg_opp_CLINCH_landed"
 [43] "B_avg_opp_DISTANCE_att"       "B_avg_opp_DISTANCE_landed"
 [45] "B_avg_opp_GROUND_att"         "B_avg_opp_GROUND_landed"
 [47] "B_avg_opp_HEAD_att"           "B_avg_opp_HEAD_landed"
 [49] "B_avg_opp_KD"                 "B_avg_opp_LEG_att"
 [51] "B_avg_opp_LEG_landed"         "B_avg_opp_PASS"
 [53] "B_avg_opp_REV"                "B_avg_opp_SIG_STR_att"
 [55] "B_avg_opp_SIG_STR_landed"     "B_avg_opp_SIG_STR_pct"
 [57] "B_avg_opp_SUB_ATT"            "B_avg_opp_TD_att"
 [59] "B_avg_opp_TD_landed"          "B_avg_opp_TD_pct"
 [61] "B_avg_opp_TOTAL_STR_att"      "B_avg_opp_TOTAL_STR_landed"
 [63] "B_total_rounds_fought"        "B_total_time_fought.seconds."
 [65] "B_total_title_bouts"          "B_win_by_Decision_Majority"
 [67] "B_win_by_Decision_Split"      "B_win_by_Decision_Unanimous"
 [69] "B_win_by_KO.TKO"              "B_win_by_Submission"
 [71] "B_win_by_TKO_Doctor_Stoppage" "B_wins"
 [73] "B_Stance"                     "B_Height_cms"
 [75] "B_Reach_cms"                  "B_Weight_lbs"
 [77] "R_current_lose_streak"        "R_current_win_streak"
 [79] "R_draw"                       "R_avg_BODY_att"
 [81] "R_avg_BODY_landed"            "R_avg_CLINCH_att"
 [83] "R_avg_CLINCH_landed"          "R_avg_DISTANCE_att"
 [85] "R_avg_DISTANCE_landed"        "R_avg_GROUND_att"
 [87] "R_avg_GROUND_landed"          "R_avg_HEAD_att"
 [89] "R_avg_HEAD_landed"            "R_avg_KD"
 [91] "R_avg_LEG_att"                "R_avg_LEG_landed"
 [93] "R_avg_PASS"                   "R_avg_REV"
 [95] "R_avg_SIG_STR_att"            "R_avg_SIG_STR_landed"
 [97] "R_avg_SIG_STR_pct"            "R_avg_SUB_ATT"
 [99] "R_avg_TD_att"                 "R_avg_TD_landed"
[101] "R_avg_TD_pct"                 "R_avg_TOTAL_STR_att"
[103] "R_avg_TOTAL_STR_landed"       "R_longest_win_streak"
[105] "R_losses"                     "R_avg_opp_BODY_att"
[107] "R_avg_opp_BODY_landed"        "R_avg_opp_CLINCH_att"
[109] "R_avg_opp_CLINCH_landed"      "R_avg_opp_DISTANCE_att"
[111] "R_avg_opp_DISTANCE_landed"    "R_avg_opp_GROUND_att"
[113] "R_avg_opp_GROUND_landed"      "R_avg_opp_HEAD_att"
[115] "R_avg_opp_HEAD_landed"        "R_avg_opp_KD"
```

```
[117] "R_avg_opp_LEG_att"             "R_avg_opp_LEG_landed"
[119] "R_avg_opp_PASS"                "R_avg_opp_REV"
[121] "R_avg_opp_SIG_STR_att"         "R_avg_opp_SIG_STR_landed"
[123] "R_avg_opp_SIG_STR_pct"         "R_avg_opp_SUB_ATT"
[125] "R_avg_opp_TD_att"              "R_avg_opp_TD_landed"
[127] "R_avg_opp_TD_pct"              "R_avg_opp_TOTAL_STR_att"
[129] "R_avg_opp_TOTAL_STR_landed"    "R_total_rounds_fought"
[131] "R_total_time_fought.seconds." "R_total_title_bouts"
[133] "R_win_by_Decision_Majority"    "R_win_by_Decision_Split"
[135] "R_win_by_Decision_Unanimous"   "R_win_by_KO.TKO"
[137] "R_win_by_Submission"           "R_win_by_TKO_Doctor_Stoppage"
[139] "R_wins"                        "R_Stance"
[141] "R_Height_cms"                  "R_Reach_cms"
[143] "R_Weight_lbs"                  "B_age"
[145] "R_age"
```

Selection des variables intéressantes.

Comme on peut le constater, le nombre de variable est conséquent (145). Une des demandes de ce projet est l'analyse en composantes principales. L'ACP est basée sur de la visualisation. Donc avoir un trop grand nombre de variables est contre productif. Il va donc nous falloir restreindre le nombre de variables et sélectionner des variables intéressantes.

Dans un premier temps, Nous allons nous concentrer sur l'analyse d'un combattant pour chaque combat, dans ce cas-ci, nous avons choisi de se concentrer sur le combattant Blue. On sélectionne donc toutes les variables en rapport avec notre combattant Blue. L'objectif dans ce projet étant de constater quelles compétences ou distinctions physiques permettent de remporter un combat mais aussi de voir les liens entre les différentes techniques de combats et comment celles-ci peuvent influencer le résultat du combat.

```
dataUFC1 = dataUFC[1:1050,c(1:76,144)]
colnames(dataUFC1)
```

```
 [1] "R_fighter"                   "B_fighter"
 [3] "Referee"                     "date"
 [5] "location"                    "Winner"
 [7] "title_bout"                  "weight_class"
 [9] "no_of_rounds"                "B_current_lose_streak"
[11] "B_current_win_streak"        "B_draw"
[13] "B_avg_BODY_att"              "B_avg_BODY_landed"
[15] "B_avg_CLINCH_att"            "B_avg_CLINCH_landed"
[17] "B_avg_DISTANCE_att"          "B_avg_DISTANCE_landed"
[19] "B_avg_GROUND_att"            "B_avg_GROUND_landed"
[21] "B_avg_HEAD_att"              "B_avg_HEAD_landed"
[23] "B_avg_KD"                    "B_avg_LEG_att"
[25] "B_avg_LEG_landed"            "B_avg_PASS"
[27] "B_avg_REV"                   "B_avg_SIG_STR_att"
[29] "B_avg_SIG_STR_landed"        "B_avg_SIG_STR_pct"
[31] "B_avg_SUB_ATT"               "B_avg_TD_att"
[33] "B_avg_TD_landed"             "B_avg_TD_pct"
```

```
[35] "B_avg_TOTAL_STR_att"           "B_avg_TOTAL_STR_landed"
[37] "B_longest_win_streak"          "B_losses"
[39] "B_avg_opp_BODY_att"            "B_avg_opp_BODY_landed"
[41] "B_avg_opp_CLINCH_att"          "B_avg_opp_CLINCH_landed"
[43] "B_avg_opp_DISTANCE_att"        "B_avg_opp_DISTANCE_landed"
[45] "B_avg_opp_GROUND_att"          "B_avg_opp_GROUND_landed"
[47] "B_avg_opp_HEAD_att"            "B_avg_opp_HEAD_landed"
[49] "B_avg_opp_KD"                  "B_avg_opp_LEG_att"
[51] "B_avg_opp_LEG_landed"          "B_avg_opp_PASS"
[53] "B_avg_opp_REV"                 "B_avg_opp_SIG_STR_att"
[55] "B_avg_opp_SIG_STR_landed"      "B_avg_opp_SIG_STR_pct"
[57] "B_avg_opp_SUB_ATT"             "B_avg_opp_TD_att"
[59] "B_avg_opp_TD_landed"           "B_avg_opp_TD_pct"
[61] "B_avg_opp_TOTAL_STR_att"       "B_avg_opp_TOTAL_STR_landed"
[63] "B_total_rounds_fought"         "B_total_time_fought.seconds."
[65] "B_total_title_bouts"           "B_win_by_Decision_Majority"
[67] "B_win_by_Decision_Split"       "B_win_by_Decision_Unanimous"
[69] "B_win_by_KO.TKO"               "B_win_by_Submission"
[71] "B_win_by_TKO_Doctor_Stoppage" "B_wins"
[73] "B_Stance"                      "B_Height_cms"
[75] "B_Reach_cms"                   "B_Weight_lbs"
[77] "B_age"
```

#Selection des variables intéressantes

Pour que vous vérifiez que les données ainsi que les analyses soient cohérentes, on ne va séletionner que les combattants participant à un combat pour le titre de champion du monde. Ces combattants étant connus, leurs compétences sont souvent connus du grand public (même le public hors sport de combat, example : Conor McGregor). Il est donc plus facile pour vous de comprendre les résultats des analyses.

```
typeof(dataUFC1$title_bout)

[1] "character"

dataUFC2 <- dataUFC1[which(dataUFC1$title_bout ==  "True"),]
```

Cependant, il nous reste encore 77 variables ce qui reste trop conséquent pour la visualisation lors de l'ACP. Nous allons donc choisir les variables suivantes : nom du combattant Blue, la catégorie de poids, la séquence de victoires actuelles, la plus longue séquence de victoire, l'âge du combattant, le nombre de victoire par soumission , le nombre de vicoitre par TKO (Technical KO), le nombre moyen d'attaque à la tête, le nombre moyen d'attaque à la tête atteint, pareil pour les attaques au corps et les attaques de takedowns.

Pour info : un takedown est un mouvement de judo, jiujitsu brésilien ou de lutte qui permet de mettre un adversaire à terre.

```
dataFinal =
dataUFC2[,c("B_fighter","weight_class","B_current_win_streak","B_longest_win_
streak","B_age","B_win_by_Submission","B_win_by_KO.TKO","B_avg_HEAD_landed","
```

```
B_avg_LEG_landed","B_avg_GROUND_landed","B_avg_HEAD_att","B_avg_LEG_att","B_avg_GROUND_att")]
```

Nous allons nous concentrer sur les combattants masculins car d'une part les combattants masculins représentent une grande majorité des combats et d'autres part, cela permet de n'avoir qu'une seule catégorie d'individus.

```
dataFinalWithoutWomen <- dataFinal[!grepl("^Women's.+",
dataFinal$weight_class),]
dataCleaned <- dataFinalWithoutWomen[complete.cases(dataFinalWithoutWomen), ]
```

Pour pouvoir mettre en place l'analyse en composante principale, nous allons mettre le nom de combattants en nom de ligne et nous allons en même temps supprimer les doublons de combattants car on ne veut pas prendre en compte plusieurs combats d'un même combattant pour cette analyse.

```
#Mettre  les noms de combattants en noms de lignes
dataCleaned = dataCleaned[!duplicated(dataCleaned$B_fighter), ]
rownames(dataCleaned) = dataCleaned$B_fighter
dataCleaned$B_fighter = NULL
head(dataCleaned)
```

```
                    weight_class B_current_win_streak B_longest_win_streak
Marlon Moraes        Bantamweight                    4                     4
Dustin Poirier        Lightweight                    3                     4
Israel Adesanya      Middleweight                    5                     5
Anthony Smith   Light Heavyweight                    3                     3
Kamaru Usman         Welterweight                    9                     9
TJ Dillashaw            Flyweight                    4                     4
                B_age B_win_by_Submission B_win_by_KO.TKO B_avg_HEAD_landed
Marlon Moraes      31                    1               2          11.20000
Dustin Poirier     30                    3               8          40.66667
Israel Adesanya    29                    0               2          32.60000
Anthony Smith      30                    1               5          24.10000
Kamaru Usman       31                    1               1          39.11111
TJ Dillashaw       32                    1               7          45.73333
                B_avg_LEG_landed B_avg_GROUND_landed B_avg_HEAD_att
Marlon Moraes           5.400000            2.000000       48.60000
Dustin Poirier          6.619048            5.095238       90.47619
Israel Adesanya        17.200000            1.600000       77.40000
Anthony Smith           2.900000            2.200000       54.70000
Kamaru Usman            7.555556           23.777778       90.22222
TJ Dillashaw           11.133333           12.400000      133.06667
                B_avg_LEG_att B_avg_GROUND_att
Marlon Moraes        7.600000         2.600000
Dustin Poirier       8.095238         8.095238
Israel Adesanya     20.200000         2.000000
Anthony Smith        3.200000         2.900000
Kamaru Usman         8.888889        33.222222
TJ Dillashaw        14.533333        16.533333
```

Maintenant notre dataset est prêt pour les différentes analyses. Il comporte 31 observations différentes et 12 variables.

##Analyse descriptive

###Statistique descriptive

Nous allons maintenant faire une analyse descriptive des données afin de pouvoir avoir une meilleure compréhension du dataset. Dans un premier temps, nous allons montrer une matrice de corrélation afin de voir les liens entre les différentes variables continues de notre analyse. install.packages("corrplot")

```
summary(dataCleaned)
```

```
 weight_class       B_current_win_streak B_longest_win_streak     B_age
 Length:31          Min.   : 0.000       Min.   : 0.000       Min.   :24.00
 Class :character   1st Qu.: 2.000       1st Qu.: 4.000       1st Qu.:28.50
 Mode  :character   Median : 3.000       Median : 5.000       Median :31.00
                    Mean   : 4.226       Mean   : 5.645       Mean   :30.52
                    3rd Qu.: 5.500       3rd Qu.: 7.000       3rd Qu.:32.50
                    Max.   :13.000       Max.   :13.000       Max.   :39.00
 B_win_by_Submission B_win_by_KO.TKO  B_avg_HEAD_landed B_avg_LEG_landed
 Min.   :0.000       Min.   : 0.000   Min.   : 2.00     Min.   : 0.7143
 1st Qu.:0.000       1st Qu.: 2.000   1st Qu.:21.75     1st Qu.: 2.1833
 Median :1.000       Median : 3.000   Median :27.43     Median : 5.3750
 Mean   :1.516       Mean   : 3.774   Mean   :27.74     Mean   : 5.4296
 3rd Qu.:2.000       3rd Qu.: 5.000   3rd Qu.:37.52     3rd Qu.: 6.7875
 Max.   :9.000       Max.   :11.000   Max.   :45.73     Max.   :17.2000
 B_avg_GROUND_landed B_avg_HEAD_att   B_avg_LEG_att     B_avg_GROUND_att
 Min.   : 1.000      Min.   : 11.00   Min.   : 0.7273   Min.   : 1.667
 1st Qu.: 2.636      1st Qu.: 51.30   1st Qu.: 2.5667   1st Qu.: 5.410
 Median : 5.688      Median : 75.50   Median : 7.0000   Median : 8.455
 Mean   : 7.449      Mean   : 72.39   Mean   : 6.8331   Mean   :10.797
 3rd Qu.:10.980      3rd Qu.: 90.35   3rd Qu.: 8.4545   3rd Qu.:15.133
 Max.   :23.778      Max.   :133.07   Max.   :20.2000   Max.   :33.222
```

Matrice de corrélation

Voici 2 manières de visualiser la corrélations entre les variables:

```
library(corrplot)
```

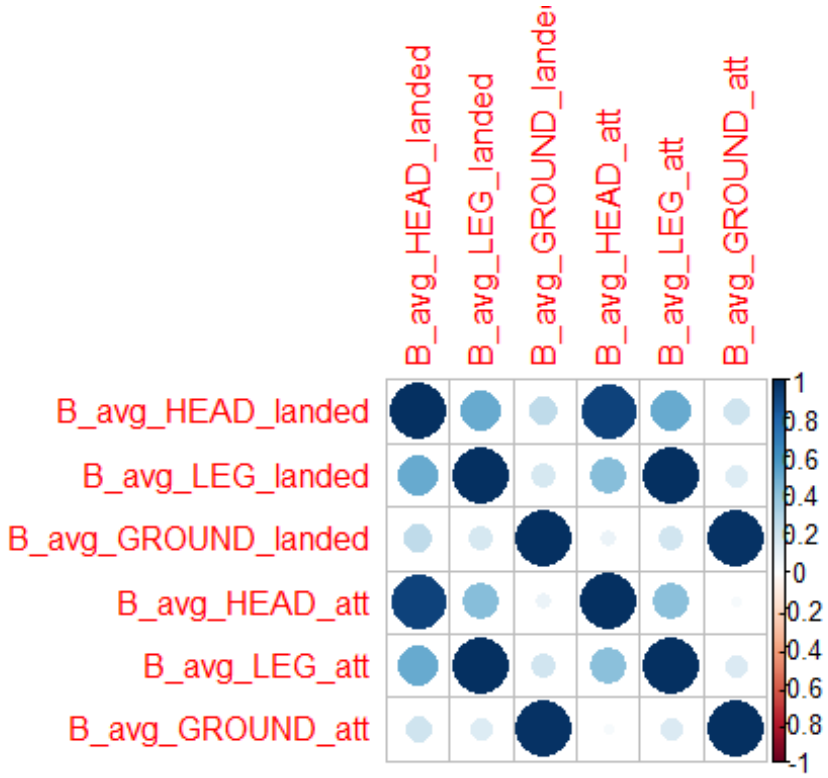
```
corrplot 0.84 loaded
```

```
quantitativeVariables <- dataCleaned[7:12]
dataUFC.cor <- cor(quantitativeVariables)
dataUFC.cor
```

```
                    B_avg_HEAD_landed B_avg_LEG_landed B_avg_GROUND_landed
B_avg_HEAD_landed           1.0000000        0.5089706          0.25839373
B_avg_LEG_landed            0.5089706        1.0000000          0.17848006
B_avg_GROUND_landed         0.2583937        0.1784801          1.00000000
```

```
B_avg_HEAD_att               0.9262812        0.4213820          0.08338993
B_avg_LEG_att                0.5019298        0.9928699          0.19203369
B_avg_GROUND_att             0.2057809        0.1469704          0.98927955
                    B_avg_HEAD_att B_avg_LEG_att B_avg_GROUND_att
B_avg_HEAD_landed       0.92628123     0.5019298       0.20578088
B_avg_LEG_landed        0.42138200     0.9928699       0.14697042
B_avg_GROUND_landed     0.08338993     0.1920337       0.98927955
B_avg_HEAD_att          1.00000000     0.4170185       0.03396246
B_avg_LEG_att           0.41701849     1.0000000       0.15985547
B_avg_GROUND_att        0.03396246     0.1598555       1.00000000

corrplot(dataUFC.cor)
```



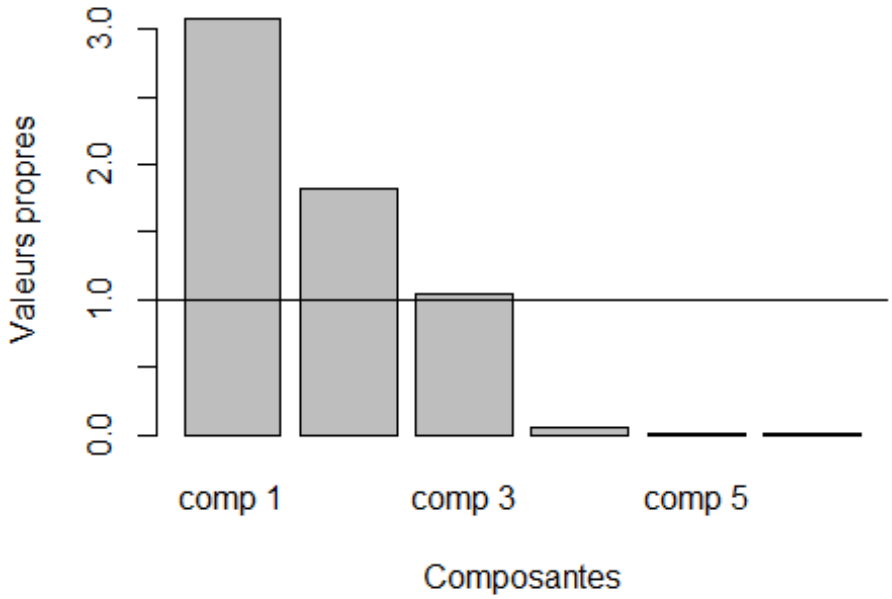
On peut constater qu'une variable _att est fortement correlée à une variable _landed ce qui est logique car un combattant qui essaye d'effectuer un mouvement de combat (_att) a une chance de toucher son adversaire. On peut aussi remarquer que la corrélation est faible pour certaines variables : si on prends les attaques de mise au sol par exemple (_att et _landed) on remarque que les corrélations avec les techniques de striking sont faibles. Ce qui est normal car quand un combattant tente un takedown, il se concentre moins à donner des coups à son adversaire car il doit le déstabiliser pour l'amener au sol. Il n'y a cependant pas de corrélations négatives entre les variables analysées.

Analyse en composantes principales

PCA

Pour pouvoir effectuer une analyse en composante principale, on va devoir utiliser le package FactoMineR. Nous devons sélectionner les variables continues de notre dataset final. Après cela, nous allons visualiser les valeurs propres en fonction des composantes afin de sélectionner quelles composantes nous devons choisir pour notre analyse.

```
library(FactoMineR)
res <- PCA(quantitativeVariables, graph = FALSE, ncp = 6)
barplot(res$eig[,"eigenvalue"],xlab = "Composantes",ylab = "Valeurs propres")
abline(h = 1)
```



Par une règle générale, on sélectionne les composantes telles que leurs valeurs propres soient plus grande que 0. On prendra donc les 3 premières composantes pour l'ACP.

```
lapply(res$var,round,3)

$coord
                    Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6
B_avg_HEAD_landed   0.840 -0.206  0.473 -0.170  0.003 -0.004
B_avg_LEG_landed    0.827 -0.249 -0.501 -0.001  0.017  0.056
B_avg_GROUND_landed 0.499  0.863  0.039  0.007 -0.067  0.015
B_avg_HEAD_att      0.737 -0.356  0.552  0.158  0.000  0.003
B_avg_LEG_att       0.828 -0.234 -0.505  0.017 -0.017 -0.056
B_avg_GROUND_att    0.456  0.887  0.020  0.021  0.066 -0.013
```

```
$cor
                     Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6
B_avg_HEAD_landed    0.840 -0.206  0.473 -0.170  0.003 -0.004
B_avg_LEG_landed     0.827 -0.249 -0.501 -0.001  0.017  0.056
B_avg_GROUND_landed 0.499  0.863  0.039  0.007 -0.067  0.015
B_avg_HEAD_att       0.737 -0.356  0.552  0.158  0.000  0.003
B_avg_LEG_att        0.828 -0.234 -0.505  0.017 -0.017 -0.056
B_avg_GROUND_att     0.456  0.887  0.020  0.021  0.066 -0.013

$cos2
                     Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6
B_avg_HEAD_landed    0.705 0.043 0.224 0.029 0.000 0.000
B_avg_LEG_landed     0.684 0.062 0.251 0.000 0.000 0.003
B_avg_GROUND_landed 0.249 0.744 0.002 0.000 0.004 0.000
B_avg_HEAD_att       0.543 0.127 0.305 0.025 0.000 0.000
B_avg_LEG_att        0.686 0.055 0.255 0.000 0.000 0.003
B_avg_GROUND_att     0.208 0.787 0.000 0.000 0.004 0.000

$contrib
                      Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6
B_avg_HEAD_landed    22.922  2.339 21.574 52.754  0.125  0.284
B_avg_LEG_landed     22.250  3.398 24.173  0.004  3.196 46.979
B_avg_GROUND_landed  8.107 40.963  0.150  0.094 47.313  3.373
B_avg_HEAD_att       17.654  6.986 29.425 45.808  0.000  0.126
B_avg_LEG_att        22.312  3.015 24.640  0.521  2.990 46.523
B_avg_GROUND_att      6.756 43.298  0.038  0.818 46.376  2.714
```

Représentation des variables sur les 3 premiers composantes

En prenant les 3 premières composantes, on aimerait voir la représentation des variables sur ces composantes afin d'évaluer la perte d'information.

```
round(sort(rowSums(res$var$cos2[,1:2])), digits = 3)

     B_avg_HEAD_att       B_avg_LEG_att    B_avg_LEG_landed
B_avg_HEAD_landed
              0.670               0.741               0.746
0.747
B_avg_GROUND_landed    B_avg_GROUND_att
              0.994               0.995

round(sort(rowSums(res$var$cos2[,1:3])), digits = 3)

  B_avg_HEAD_landed      B_avg_HEAD_att    B_avg_GROUND_att
B_avg_GROUND_landed
              0.971               0.975               0.995
0.995
      B_avg_LEG_att    B_avg_LEG_landed
              0.996               0.997
```

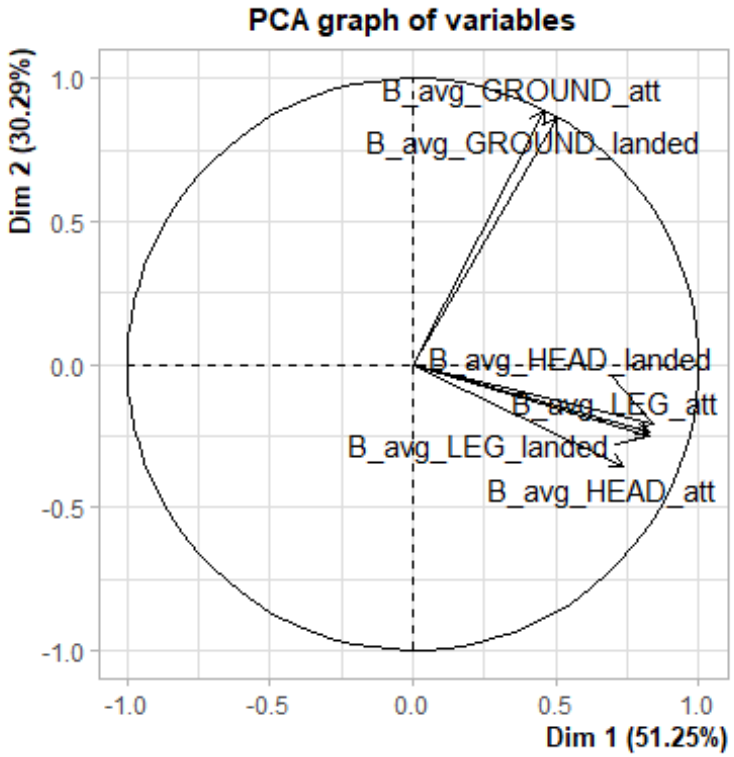
On peut constater que les variables en relation avec les attaques aux sols sont très bien représenter sur le 1er plan factoriel. En prenant en compte les 3 premiers axes factoriels, on remarque qu'une grand quantité de l'information est captée.

Visualisation

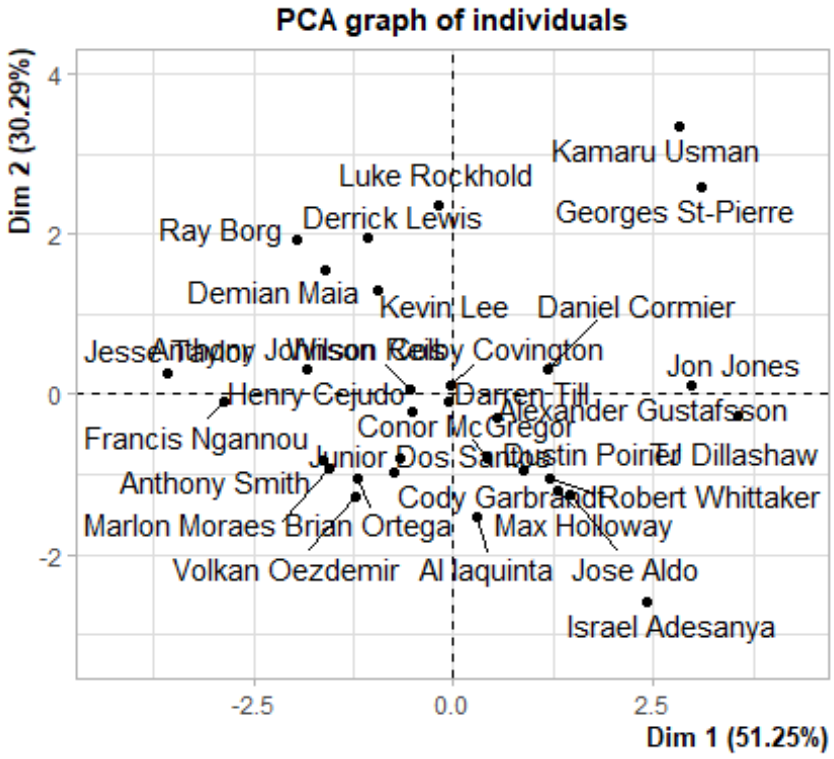
Pour pouvoir faire une conclusion sur l'analyse en composantes principales, il faut pouvoir visualiser les différents graphe tels que le plan factoriel ainsi que le cercle de corrélation. Nous avons vu cependant que 3 composantes factorielles avaient des valeurs propres >=1 , c'est-à-dire qu'il faut prendre en compte ces 3 composantes. La question à se poser est alors, comment visualiser le plan factoriel si on a 3 composantes. Il faut alors visualiser la chose sur 2 plans factoriels ainsi que 2 cercles de corrélation.

###Répresentation sur le premier plan factoriel

```
plot.PCA(res, choix = "var", axes=c(1,2))
```




```
plot.PCA(res, choix = "ind", axes=c(1,2))
```



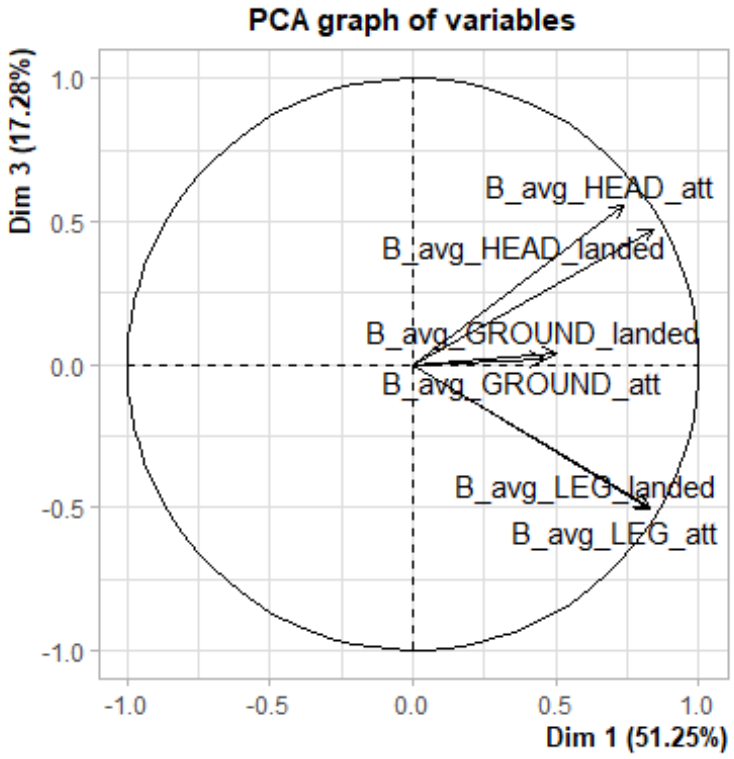
On peut constater que les variables sont plutôt bien représentées par les 2 premières composantes (51.25% + 30.29% = 81.54%). On ne perd donc pas beaucoup d'informations sur le premier plan factoriel ce qui nous permet de faire une interprétation fiable de la map factorielle des individus ainsi que sa map factorielle des individus associés.

On remarque que sur le premier plan factoriel les variables en relations avec les attaques au sols sont mieux représentées que les attaques en striking car les flèches sont plus proches du cercle de corrélation.

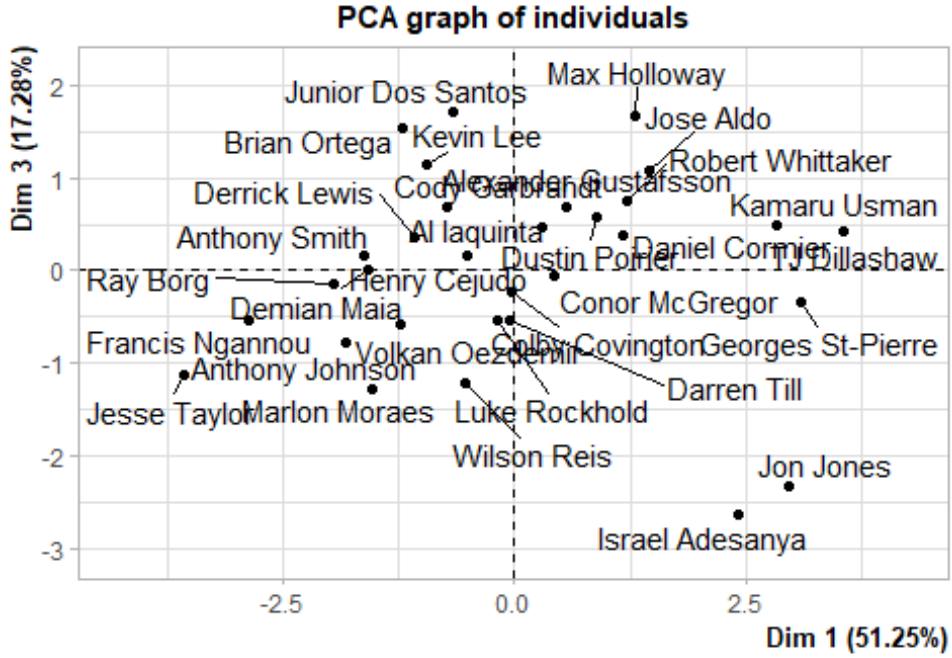
On peut donc voir, en haut à droite, Georges Saint Pierre, qui est considéré comme le meilleur de tous les temps, être un excellent lutteur tout comme Kamaru Usman qui a été Champion de lutte en Amérique avant de commencer sa carrière à l'UFC. A l'inverse en bas à droite, on peut voir Israel Adesanya, qui est l'actuel champion du monde de la catégorie poids léger, avec un excellent striking.

###Répresentation sur le deuxième plan factoriel

```
plot.PCA(res, choix = "var", axes=c(1,3))
```




```
plot.PCA(res, choix = "ind", axes=c(1,3))
```



Sur le 2ème plan factoriel (Dim 1 et Dim 3), on remarque que les variables sont représentées à 68.53% (51.25% + 17.28%). On perd donc plus d'informations sur le premier plan factoriel mais celà reste raisonnable. On peut constater à l'inverse que les attaques en striking sont mieux représentées que les attaques au sol car les flèches sont plus proches du cercle de corrélation.

Sur le 2eme plan factoriel, on peut encore améliorer notre interprétation à propos d'Israel Adesanya. On peut constater qu'il utilise énormément ces jambes afin de kicker son ardversaire aux jambes et le déstabiliser tout comme Jon Jones qui fut l'ancien champion du monde , poids lourd-léger.

Clustering

Analyse des correspondances

Conclusions

Annexes