Celtics

2023-02-27

```
# Loading the database
data = read.csv("DS2010_Celtics_Data.csv")
```

```
# Saving the important columns into variable data
data <- data[, -c(1, 3, 5, 25:41)]
```

```
# Making win(1) or loss(0) ---- 1 or 0
data$WL <- ifelse(data$WL == "W", 1, 0)
```

```
# Making X column "HomeAway" into 1 or 0 ---- home(1) / away(0)
colnames(data)[2] <- "HomeAway"
data$HomeAway <- ifelse(data$HomeAway == "@", 0, 1)
```

```
library('corrr')
library(ggplot2)
library(ggcorrplot)
library("FactoMineR")
```

```
# Normalizing the data
numerical_data <- data[,5:21]
head(numerical_data)
```

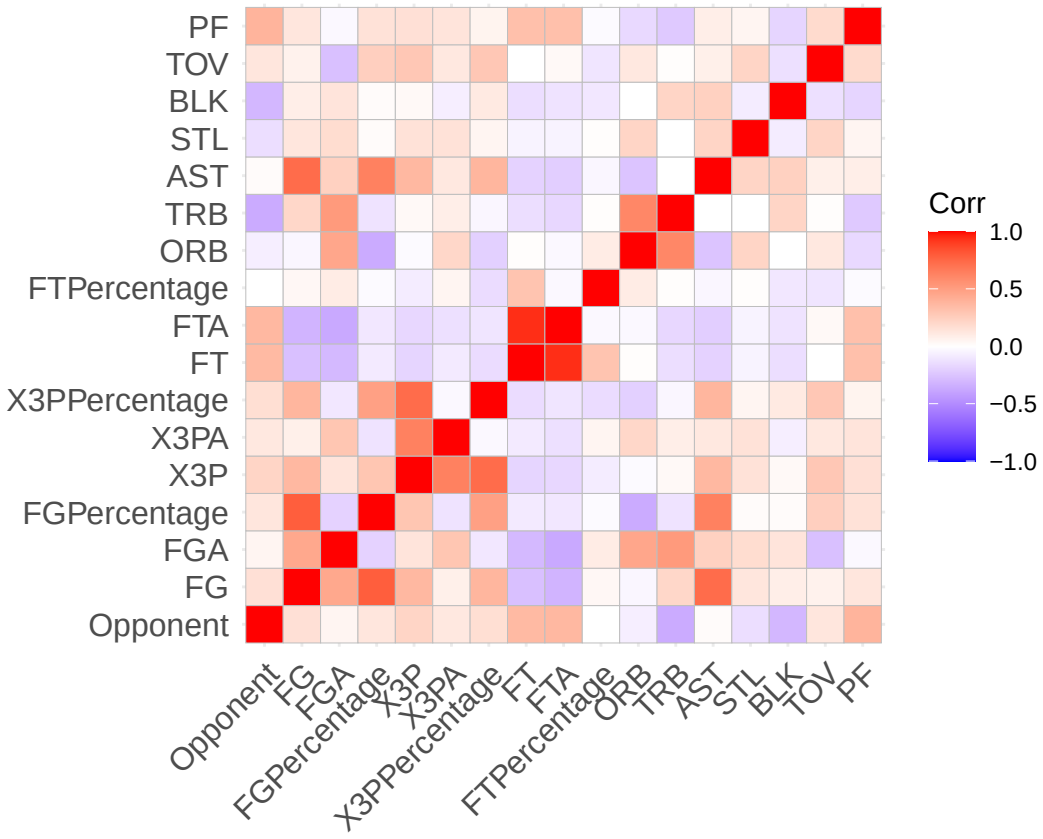
```
##   Opponent FG FGA FGPercentage X3P X3PA X3PPercentage FT FTA FTPercentage ORB
## 1       95 39  85        0.459   8   24         0.333 26  27        0.963  10
## 2      113 32  85        0.376   7   26         0.269 32  41        0.780   7
## 3       95 35  98        0.357   6   29         0.207 11  14        0.786  16
## 4      100 35  83        0.422  10   27         0.370 18  22        0.818  15
## 5       98 44  97        0.454  12   30         0.400 18  23        0.783  15
## 6       83 38  86        0.442   7   30         0.233 16  19        0.842  14
##   TRB AST STL BLK TOV PF
## 1  41  31  10   7  17 23
## 2  38  23  14   3  17 28
## 3  45  21  10   5  16 23
## 4  49  20   7   5  19 22
## 5  53  24  18   5  17 21
## 6  45  21  11   6  12 18
```

```
data_normalized <- scale(numerical_data)
head(data_normalized)
```

```
##        Opponent          FG        FGA FGPercentage        X3P        X3PA
```

```
## [1,] -0.6342678 -0.04733376 -0.6422186   0.39772109 -0.2702764 -0.39506536
## [2,]  0.8855037 -1.55675478 -0.6422186  -1.35204603 -0.6335987 -0.02270491
## [3,] -0.6342678 -0.90986006  1.3250372  -1.75259513 -0.9969210  0.53583577
## [4,] -0.2121090 -0.90986006 -0.9448734  -0.38229558  0.4563683  0.16347532
## [5,] -0.3809725  1.03082411  1.1737099   0.29231343  1.1830129  0.72201599
## [6,] -1.6474488 -0.26296533 -0.4908912   0.03933505 -0.6335987  0.72201599
##      X3PPercentage         FT        FTA FTPercentage        ORB         TRB
## [1,]   -0.02800645  1.4134504  0.5529271   2.16461595 -0.4437437 -0.68383339
## [2,]   -0.80566702  2.5497537  2.7801565  -0.10798091 -1.2834434 -1.20789263
## [3,]   -1.55902569 -1.4273078 -1.5152144  -0.03346954  1.2356556  0.01491225
## [4,]    0.42157857 -0.1016206 -0.2425119   0.36392445  0.9557557  0.71365790
## [5,]    0.78610696 -0.1016206 -0.0834241  -0.07072523  0.9557557  1.41240355
## [6,]   -1.24310109 -0.4803884 -0.7197753   0.66196994  0.6758558  0.01491225
##             AST        STL        BLK        TOV          PF
## [1,]  1.26560442  0.2639359  1.1837078  0.9595184  0.22946039
## [2,] -0.21431804  1.5370387 -0.5342398  0.9595184  1.27477997
## [3,] -0.58429865  0.2639359  0.3247340  0.6882063  0.22946039
## [4,] -0.76928896 -0.6908911  0.3247340  1.5021425  0.02039648
## [5,] -0.02932773  2.8101415  0.3247340  0.9595184 -0.18866743
## [6,] -0.58429865  0.5822116  0.7542209 -0.3970421 -0.81585918
```

```
#Compute the correlation matrix

corr_matrix <- cor(data_normalized)
ggcorrplot(corr_matrix)
```



```r
# Applying PCA

data.pca <- princomp(corr_matrix)

#Cumulative Proportion
summary(data.pca)
```

```
## Importance of components:
##                            Comp.1    Comp.2     Comp.3    Comp.4     Comp.5
## Standard deviation     0.8015295 0.7014848 0.40932778 0.3382005 0.28173363
## Proportion of Variance 0.3789398 0.2902470 0.09882657 0.0674652 0.04681755
## Cumulative Proportion  0.3789398 0.6691868 0.76801338 0.8354786 0.88229613
##                            Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
## Standard deviation     0.24799641 0.23457496 0.17237847 0.15575467 0.13357976
## Proportion of Variance 0.03627622 0.03245597 0.01752659 0.01430914 0.01052477
## Cumulative Proportion  0.91857235 0.95102831 0.96855490 0.98286403 0.99338880
##                           Comp.11     Comp.12     Comp.13      Comp.14
## Standard deviation     0.072289146 0.061468761 0.042088638 0.0207590851
## Proportion of Variance 0.003082318 0.002228641 0.001044867 0.0002541837
## Cumulative Proportion  0.996471122 0.998699763 0.999744630 0.9999988139
##                             Comp.15      Comp.16      Comp.17
## Standard deviation     1.247133e-03 6.748941e-04 6.983424e-09
## Proportion of Variance 9.173952e-07 2.686597e-07 2.876525e-17
## Cumulative Proportion  9.999997e-01 1.000000e+00 1.000000e+00
```

```r
# Visualization of the principal components

library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

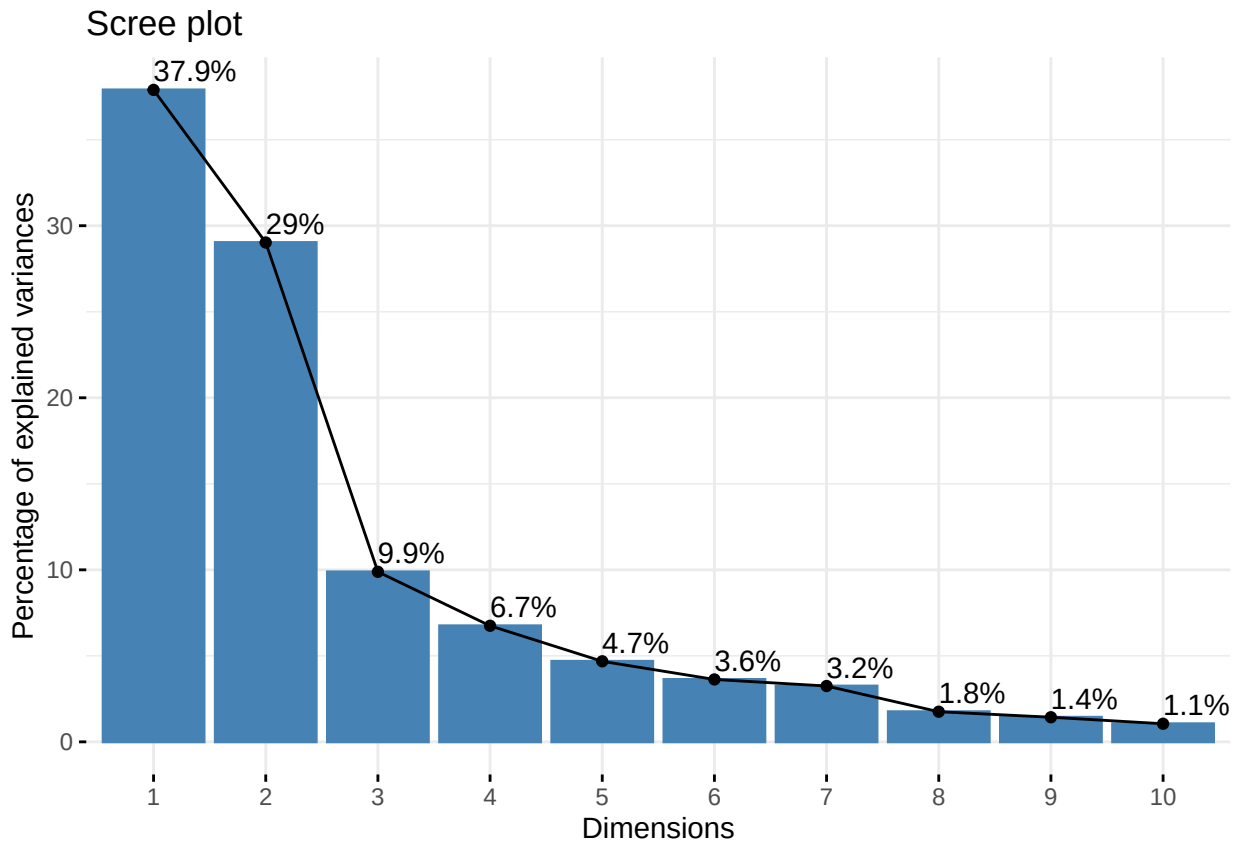
```r
library(vctrs)

data.pca$loadings[, 1:17]
```

```
##                      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
## Opponent       0.15692858  0.26530177  0.22819902  0.36765597  0.15966061
## FG            -0.37035826  0.13491800 -0.15172256  0.30764304 -0.15146263
## FGA           -0.21914112 -0.33862309  0.06491445  0.42938687  0.06830455
## FGPercentage  -0.25986710  0.37718046 -0.22606251  0.03722168 -0.20876512
## X3P           -0.25313820  0.16987759  0.42386320 -0.06922996  0.28814191
## X3PA          -0.09290624 -0.08812895  0.53423576  0.16476024  0.20881984
## X3PPercentage -0.24642198  0.29858050  0.08549276 -0.26292988  0.19915570
## FT             0.44762212  0.12010832 -0.11439650  0.06284053  0.01238131
## FTA            0.43826201  0.15982340 -0.07685153 -0.05703757  0.09526706
## FTPercentage   0.10510142 -0.09451282 -0.13082875  0.32407083 -0.23508582
## ORB            0.03549614 -0.42291086  0.19417699 -0.06971055 -0.13819053
## TRB           -0.13893437 -0.39730752 -0.07208054 -0.11095834  0.03915938
## AST           -0.33929564  0.19897225 -0.18758259  0.14995145 -0.08328168
## STL           -0.07662411 -0.06399118  0.17924699 -0.12583874 -0.57829893
## BLK           -0.15305853 -0.11957694 -0.36732648 -0.26843597  0.44732090
## TOV           -0.02333516  0.16141661  0.29582745 -0.43612940 -0.33160786
```

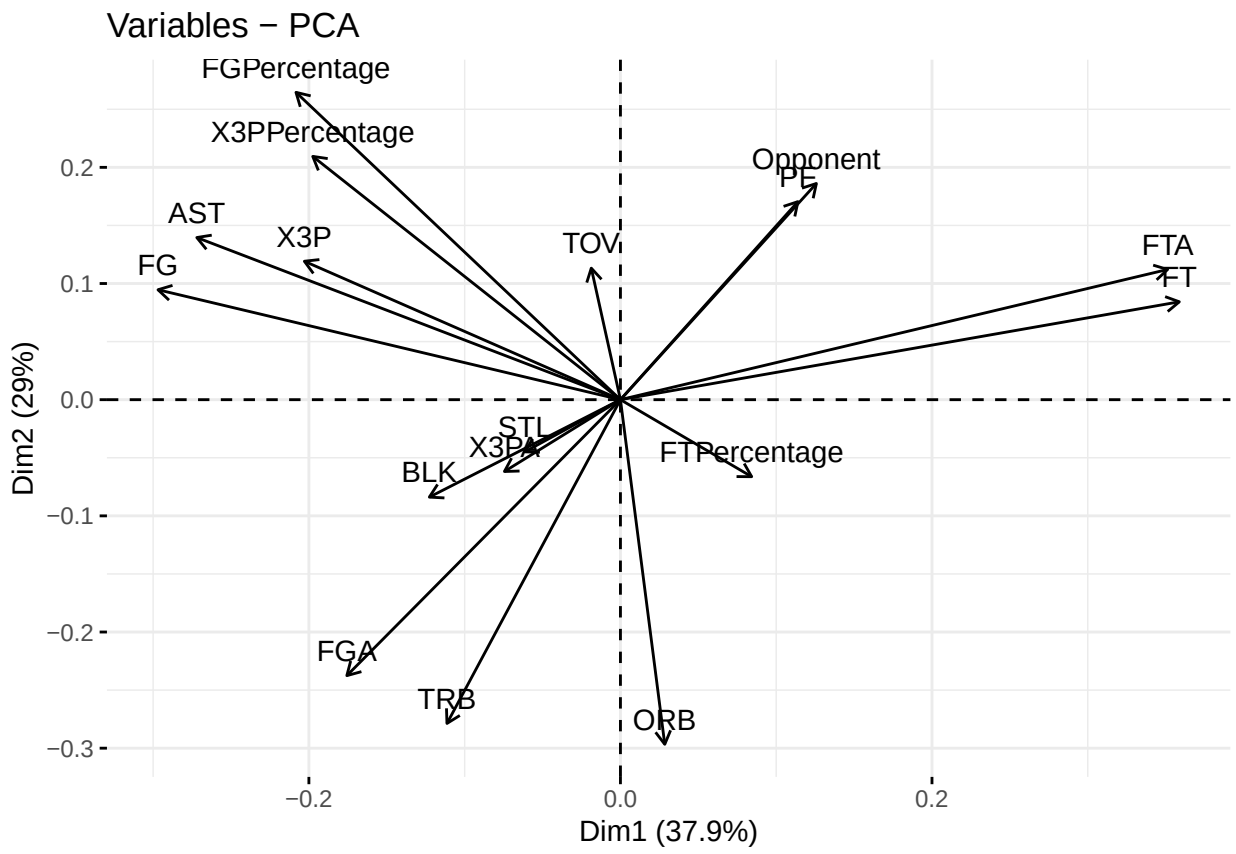
```
## PF             0.14215382  0.24301878  0.16864738  0.23102485 -0.02179241
##                    Comp.6      Comp.7      Comp.8      Comp.9      Comp.10
## Opponent       0.30674437  0.11068503  0.22520706  0.36969242  0.327390921
## FG             0.23220276  0.05936204 -0.01044543 -0.13775394  0.048879033
## FGA            0.20072065 -0.13690997  0.13579488  0.13497701 -0.083099809
## FGPercentage   0.11396235  0.16035623 -0.10698174 -0.24136181  0.120063070
## X3P           -0.16495740  0.10846167  0.16604288 -0.20046196 -0.157147203
## X3PA          -0.30927720 -0.11679366 -0.21996346 -0.38270105  0.275673786
## X3PPercentage  0.03549881  0.25599572  0.40981133  0.08164282 -0.415401346
## FT            -0.01581793  0.03054085  0.15035814 -0.35122070 -0.009773835
## FTA            0.16985436 -0.14792036  0.20718577 -0.38659346  0.042097726
## FTPercentage  -0.56222317  0.50108715 -0.05754925  0.05780198 -0.171519522
## ORB            0.29981509  0.16318707  0.11175850 -0.03121311  0.061660114
## TRB            0.34629544  0.20328264 -0.08774294 -0.41640105 -0.230767939
## AST           -0.05714825 -0.25093752  0.03085907 -0.24276287  0.194349171
## STL           -0.18262061 -0.48455032  0.41661872  0.02626419 -0.113896264
## BLK           -0.15934933 -0.28238230 -0.13957724  0.20282933  0.102232932
## TOV            0.16352531  0.17853142 -0.38047368  0.13764351  0.239261855
## PF             0.17560154 -0.31241121 -0.49758232  0.06589139 -0.623755943
##                   Comp.11     Comp.12     Comp.13      Comp.14      Comp.15
## Opponent       0.07514922  0.05345925  0.16028067  0.506687919  0.009341795
## FG            -0.15204978  0.11230853  0.29112775 -0.203645812  0.004189242
## FGA            0.32596489 -0.01398318  0.25223606 -0.453609898  0.017604899
## FGPercentage  -0.39241041  0.13966868  0.15746586  0.039666877  0.031510302
## X3P           -0.00335681  0.10457341  0.09691731 -0.017614790 -0.658493423
## X3PA          -0.10299725  0.03704492  0.09301856  0.031857861  0.446426536
## X3PPercentage  0.13495054  0.10853655 -0.04621792 -0.069786861  0.504088423
## FT             0.19308297  0.20976732  0.19115925 -0.105222133  0.235466638
## FTA            0.11240686  0.10219319  0.14763140 -0.161805358 -0.216970280
## FTPercentage   0.20537037  0.23878716  0.15354488  0.133226592 -0.078399848
## ORB           -0.19793181  0.69267657 -0.32411537 -0.007860054 -0.016012220
## TRB            0.19156087 -0.23859229  0.21584654  0.517161588  0.016356756
## AST            0.55098680  0.16603857 -0.49844643  0.188953027 -0.037667513
## STL           -0.06147600  0.04576737  0.29078695  0.247371859  0.023793089
## BLK            0.03843302  0.45098019  0.38493305  0.160254535  0.006475783
## TOV            0.45312922  0.08591660  0.25353066 -0.169664233 -0.015327600
## PF            -0.00531395  0.22829252 -0.04409245  0.119922850  0.011137159
##                   Comp.16      Comp.17
## Opponent       0.003290429  0.017505253
## FG             0.215561454 -0.645497473
## FGA           -0.115236102  0.389921078
## FGPercentage  -0.172795294  0.583610119
## X3P           -0.226203876 -0.052219187
## X3PA           0.157127660  0.040928192
## X3PPercentage  0.165565379  0.039287074
## FT            -0.627531047 -0.206496843
## FTA            0.604293649  0.189637463
## FTPercentage   0.217843076  0.072378595
## ORB            0.009002379  0.003523697
## TRB            0.003400572  0.016416352
## AST           -0.001999951  0.007520967
## STL           -0.005504262  0.006818024
## BLK            0.014886090 -0.002833026
## TOV            0.011620503 -0.002446494
```

```
## PF              0.014508646  0.014176617
```

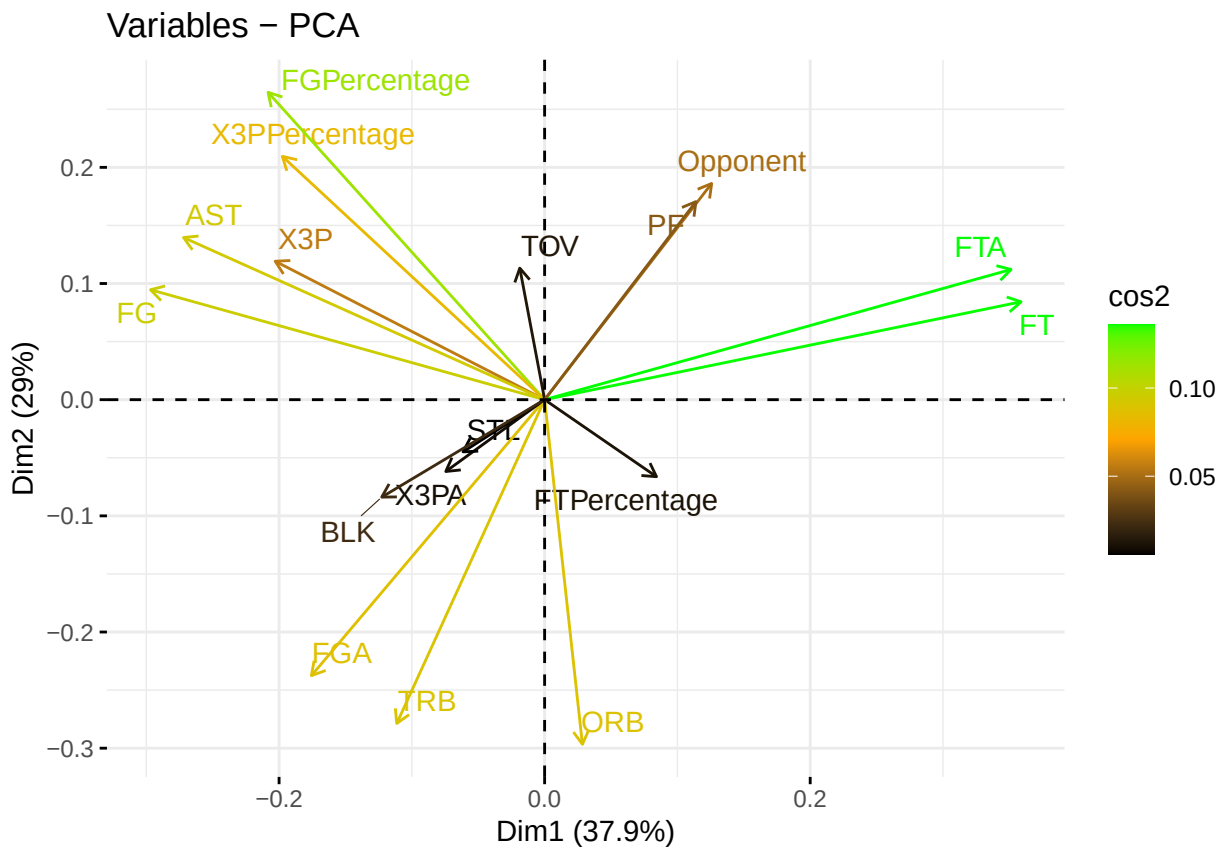
```
# Scree Plot
fviz_eig(data.pca, addlabels = TRUE)
```



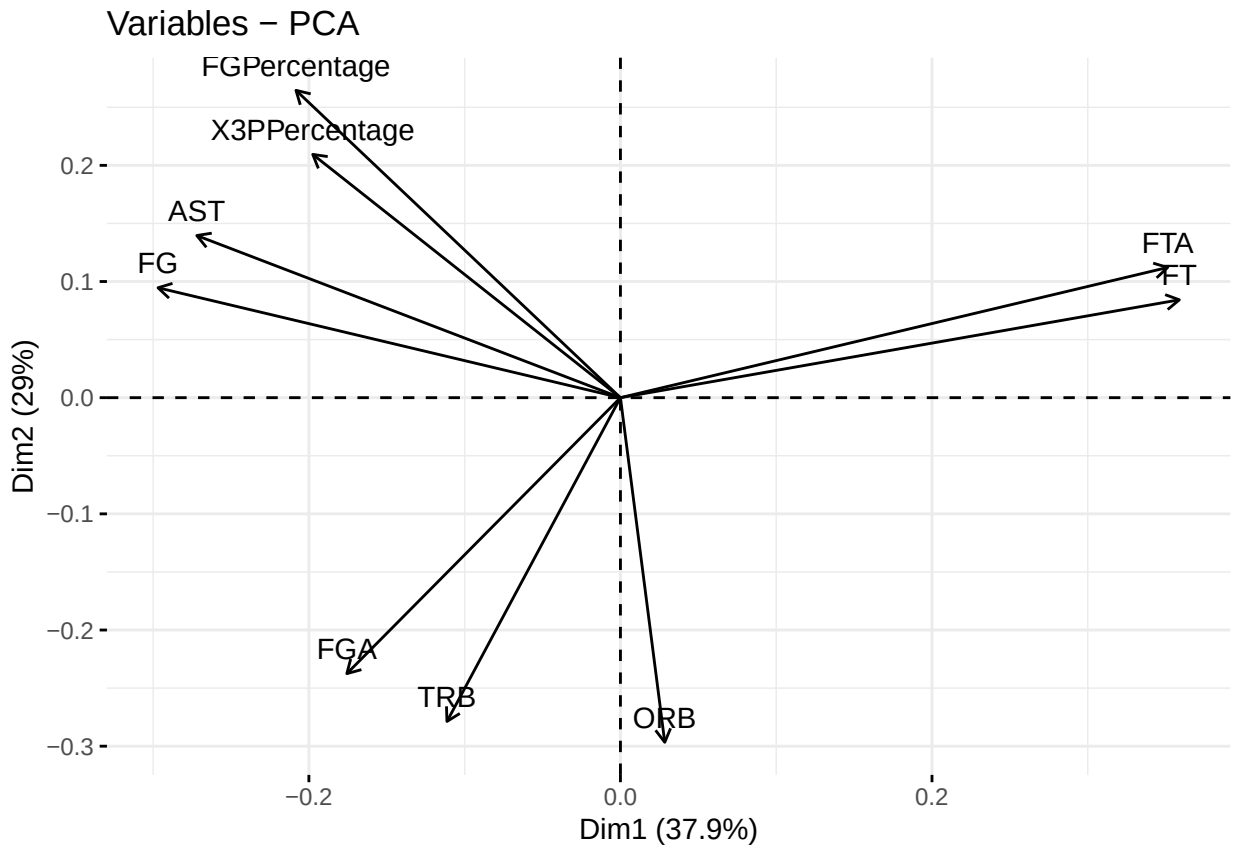

```
# Biplot of the attributes
fviz_pca_var(data.pca, col.var = "black")
```




```
# Biplot combined with cos2
fviz_pca_var(data.pca, col.var = "cos2",
             gradient.cols = c("black", "orange", "green"),
             repel = TRUE)
```



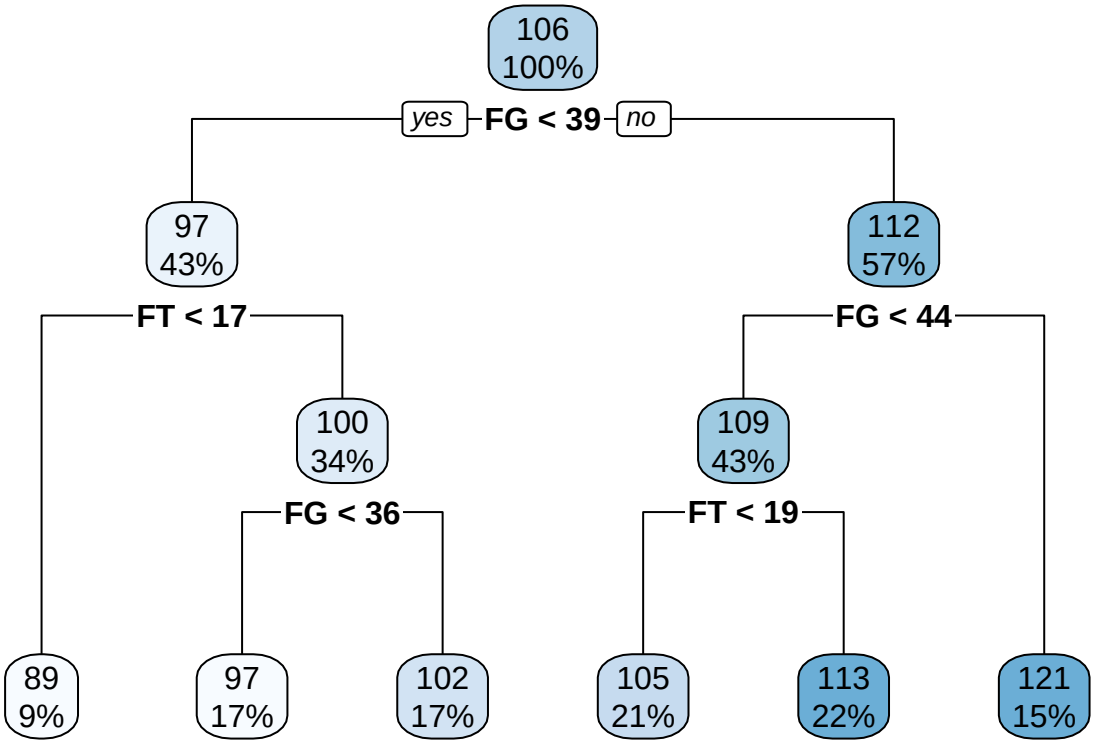

```
# Select the top 9 contributing variable
fviz_pca_var(data.pca, select.var = list(contrib = 9))
```



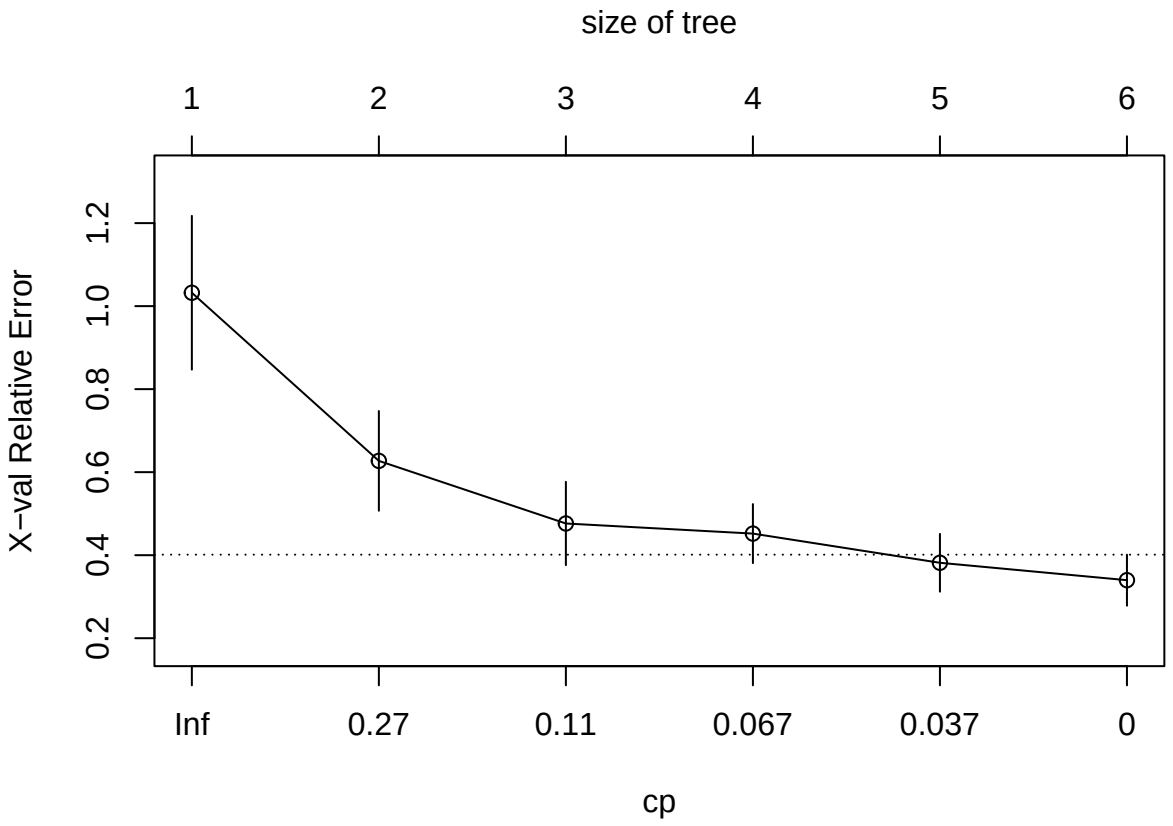

```
# Regression Tree Part

library(rpart)
library(rpart.plot)
fit.tree = rpart(Tm ~ FG+AST+X3PPercentage+FGPercentage+FT+FTA+FGA+TRB+ORB, data=data, method="anova", c
#summary(fit.tree)
fit.tree
```

```
## n= 82
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 82 8804.5490 105.71950
##    2) FG< 38.5 35 1540.4000  97.40000
##      4) FT< 16.5 7  283.7143  88.57143 *
##      5) FT>=16.5 28  574.6786  99.60714
##       10) FG< 35.5 14  238.9286  96.92857 *
##       11) FG>=35.5 14  134.8571 102.28570 *
##    3) FG>=38.5 47 3037.6600 111.91490
##      6) FG< 43.5 35 1144.9710 108.82860
##       12) FT< 18.5 17  405.7647 104.88240 *
##       13) FT>=18.5 18  224.4444 112.55560 *
##      7) FG>=43.5 12  586.9167 120.91670 *
```

```
rpart.plot(fit.tree)
```




```
plotcp(fit.tree)
```


size of tree
1
2
3
4
5
6
X−val Relative Error
0.2
0.4
0.6
0.8
1.0
1.2
Inf
0.27
0.11
0.067
0.037
0
cp