Warriors

2023-02-27

```
# Loading the database
data = read.csv("DS2010_Warriors_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)
```

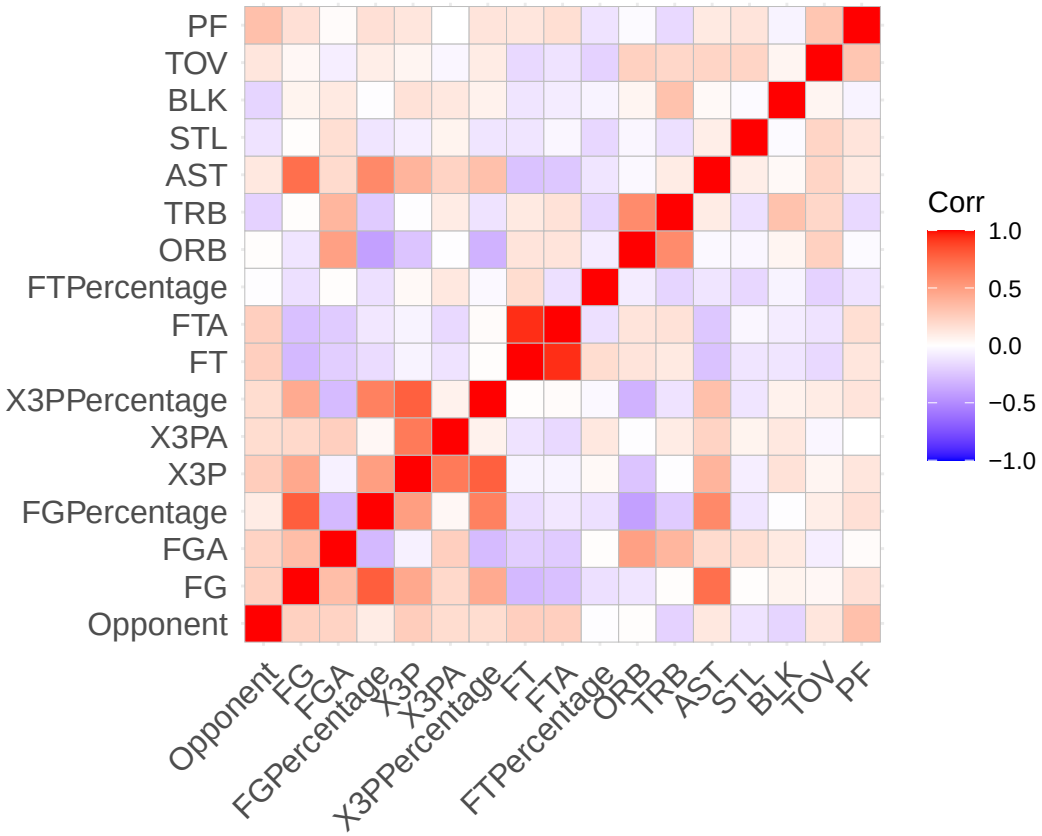
```
##   Opponent FG FGA FGPercentage X3P X3PA X3PPercentage FT FTA FTPercentage ORB
## 1       95 41  96        0.427   9   30         0.300 20  22        0.909  21
## 2       92 43  93        0.462   9   26         0.346 17  25        0.680  11
## 3      120 46  84        0.548  17   30         0.567 25  35        0.714   3
## 4       69 43  84        0.512  11   25         0.440 22  30        0.733  10
## 5      108 39  85        0.459  10   26         0.385 24  31        0.774   9
## 6      104 46  93        0.495  17   37         0.459 10  18        0.556  13
##   TRB AST STL BLK TOV PF
## 1  56  29   8   7  20 29
## 2  47  26   9   4   8 24
## 3  38  26  10   4  10 24
## 4  65  32   8  13  14 15
## 5  42  28   6   5  12 22
## 6  53  40  10   6  20 22
```

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

```
##         Opponent          FG        FGA FGPercentage        X3P       X3PA
## [1,] -0.78734824 -0.33451792  1.3848190   -1.1799669 -1.0313408 -0.2687765
```

```
## [2,] -1.04594325  0.09745798  0.9070273   -0.5081746 -1.0313408 -0.9366453
## [3,]  1.36761022  0.74542182 -0.5263478    1.1425151  0.9644101 -0.2687765
## [4,] -3.02850503  0.09745798 -0.5263478    0.4515287 -0.5324031 -1.1036125
## [5,]  0.33323016 -0.76649381 -0.3670839   -0.5657568 -0.7818719 -0.9366453
## [6,] -0.01156319  0.74542182  0.9070273    0.1252296  0.9644101  0.8999939
##      X3PPercentage          FT         FTA FTPercentage         ORB        TRB
## [1,]    -1.2149945  0.55496900  0.02209689    1.3494296  3.08376255  1.5074238
## [2,]    -0.7257280  0.05671216  0.41037074   -0.8397358  0.29271918  0.1237437
## [3,]     1.6248784  1.38539707  1.70461690   -0.5147068 -1.94011551 -1.2599363
## [4,]     0.2740774  0.88714023  1.05749382   -0.3330730  0.01361485  2.8911039
## [5,]    -0.3109151  1.21931146  1.18691843    0.0588736 -0.26548949 -0.6449674
## [6,]     0.4761658 -1.10588714 -0.49560158   -2.0251353  0.85092786  1.0461971
##              AST        STL        BLK        TOV         PF
## [1,]  0.01203658 -0.1329912  0.3604440  1.1900866  2.0771913
## [2,] -0.58016336  0.1974717 -0.8062563 -1.5914935  0.8192172
## [3,] -0.58016336  0.5279346 -0.8062563 -1.1278968  0.8192172
## [4,]  0.60423653 -0.1329912  2.6938445 -0.2007034 -1.4451360
## [5,] -0.18536340 -0.7939169 -0.4173562 -0.6643001  0.3160276
## [6,]  2.18343640  0.5279346 -0.0284561  1.1900866  0.3160276
```

```
#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     0.8394068 0.6014129 0.40218891 0.35713313 0.33529425
## Proportion of Variance 0.4195134 0.2153507 0.09630772 0.07593836 0.06693498
## Cumulative Proportion  0.4195134 0.6348641 0.73117187 0.80711022 0.87404520
##                            Comp.6     Comp.7     Comp.8     Comp.9     Comp.10
## Standard deviation     0.26874147 0.23338426 0.19698307 0.13646704 0.121560427
## Proportion of Variance 0.04300018 0.03242978 0.02310249 0.01108808 0.008798028
## Cumulative Proportion  0.91704538 0.94947517 0.97257765 0.98366574 0.992463766
##                           Comp.11     Comp.12      Comp.13      Comp.14
## Standard deviation     0.078067529 0.068156963 0.0392718681 0.0192080316
## Proportion of Variance 0.003628622 0.002765804 0.0009182565 0.0002196679
## Cumulative Proportion  0.996092389 0.998858193 0.9997764492 0.9999961172
##                            Comp.15      Comp.16 Comp.17
## Standard deviation     2.338123e-03 1.026986e-03       0
## Proportion of Variance 3.254884e-06 6.279568e-07       0
## Cumulative Proportion  9.999994e-01 1.000000e+00       1
```

```r
# Visualization of the principal components

library(factoextra)
```

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

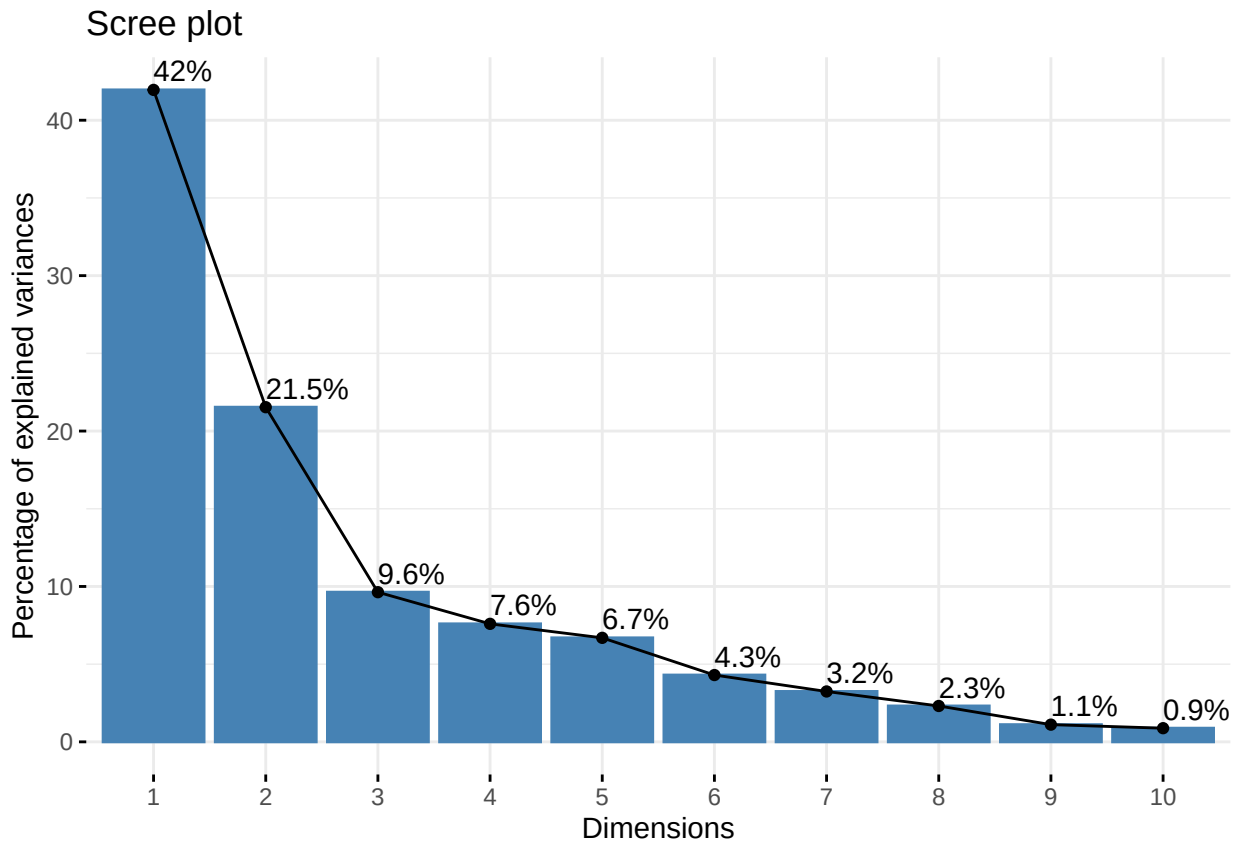
```r
library(vctrs)

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

```
##                      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
## Opponent       0.052929243  0.17900824  0.09629093  0.29092777  0.51805016
## FG             0.381336456 -0.15179167  0.06854424 -0.03367810  0.28722454
## FGA           -0.100543894 -0.42459581 -0.05684944  0.23337627  0.38747503
## FGPercentage   0.445715412  0.13082936  0.10888539 -0.19298117  0.04421305
## X3P            0.343915152  0.08458039 -0.26806239 -0.07982128  0.07548031
## X3PA           0.133449772 -0.14186934 -0.37493202  0.19731563  0.12052772
## X3PPercentage  0.358717348  0.22263074 -0.05225707 -0.24045752 -0.02525081
## FT            -0.266727129  0.44794980 -0.06227432 -0.11327231  0.17165235
## FTA           -0.249773752  0.43110479  0.07450335 -0.21190934  0.21619546
## FTPercentage  -0.048633701  0.09087532 -0.44494127  0.34330404 -0.17418432
## ORB           -0.305996014 -0.27312818  0.08562102 -0.16424463  0.28795258
## TRB           -0.200275814 -0.28237152 -0.07604756 -0.49701430  0.16860799
## AST            0.323674377 -0.18318700  0.12139346 -0.07718115  0.16617807
## STL           -0.020059811 -0.12859480  0.32441446  0.30456857 -0.35023579
## BLK           -0.005686957 -0.18572685 -0.19889147 -0.35371343 -0.27049849
## TOV            0.024972581 -0.12629311  0.45868956 -0.10687417 -0.16749802
```

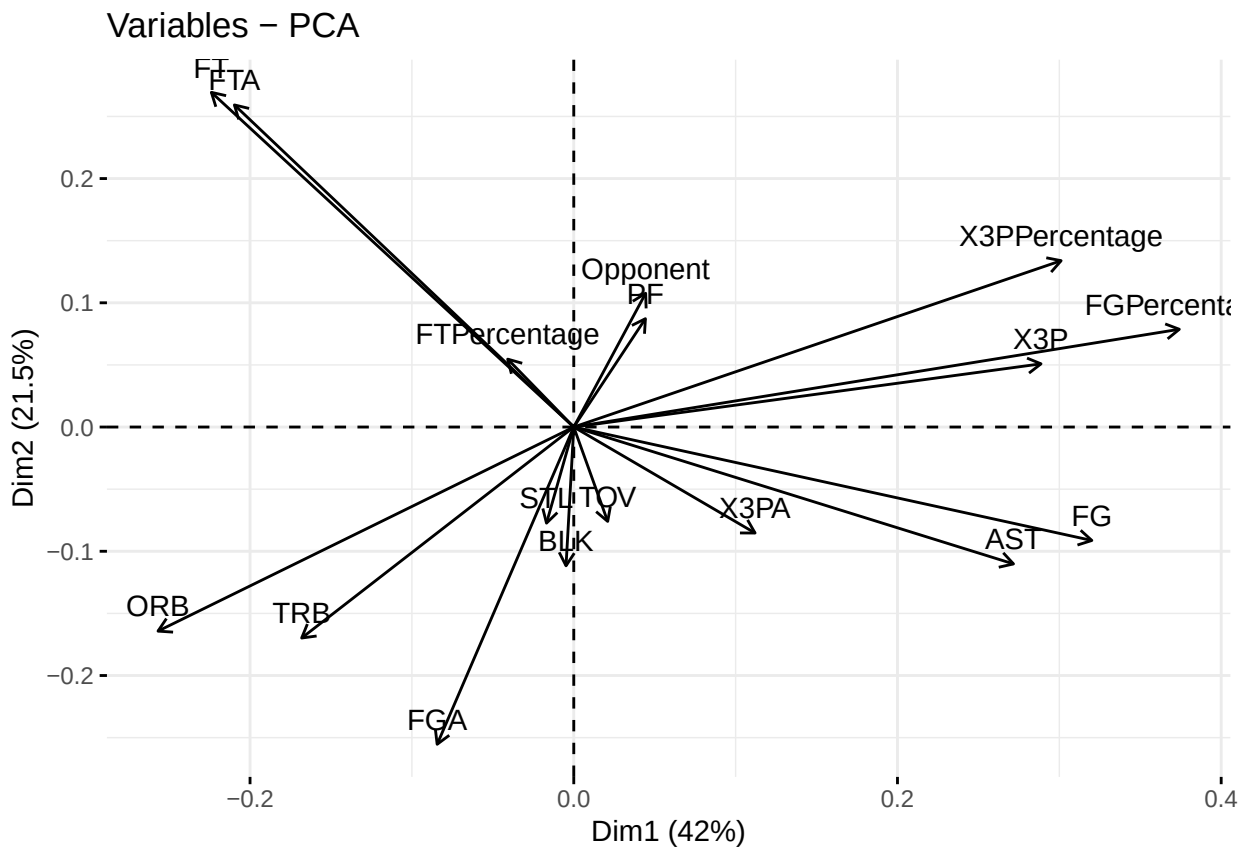
```
## PF             0.052745923  0.14502533  0.40660643  0.18587176  0.08232121
##                      Comp.6      Comp.7      Comp.8      Comp.9     Comp.10
## Opponent       0.188000750  0.18630797  0.20217627  0.46865873  0.29048395
## FG            -0.271350058 -0.16309988  0.09355973 -0.02858673 -0.02984266
## FGA           -0.051383579 -0.16556342  0.17581591  0.11961787 -0.23778895
## FGPercentage  -0.245511219 -0.04849925 -0.02649479 -0.10651732  0.13158727
## X3P            0.434225340  0.01422524 -0.12852208  0.01603344 -0.21700773
## X3PA           0.543112862 -0.09318367 -0.21166058 -0.31820237  0.25488102
## X3PPercentage  0.109684203  0.09796765 -0.05110100  0.30331712 -0.53491266
## FT            -0.028324357 -0.14295179 -0.09961844 -0.10276323  0.09430228
## FTA            0.079768255 -0.28703052 -0.05805202 -0.05464185  0.10266055
## FTPercentage  -0.321846070  0.39388151 -0.07058520 -0.13676495 -0.03531662
## ORB           -0.006835469  0.19765459 -0.17717028 -0.02802966 -0.24593132
## TRB            0.087471453  0.03134113 -0.15042477 -0.14008520 -0.07951298
## AST           -0.200199199 -0.08192145 -0.22437455 -0.23699517  0.34649835
## STL            0.173414298 -0.53236134 -0.22054568  0.15716749 -0.10890482
## BLK            0.169176610 -0.08822076  0.67412948  0.09603720  0.26683201
## TOV            0.260190163  0.52029594 -0.21715929  0.12868437  0.25810506
## PF             0.218254398  0.15432645  0.42654727 -0.63329692 -0.29233310
##                     Comp.11       Comp.12      Comp.13      Comp.14      Comp.15
## Opponent       0.09608870  0.0085257585  0.100078397  0.409906675  0.007348904
## FG            -0.15258306  0.1921385260 -0.393428021  0.003580074 -0.068547959
## FGA            0.24966988  0.1353070203 -0.233154590 -0.408413095  0.041956555
## FGPercentage  -0.30648844  0.1069349788 -0.253131691  0.240062869  0.070535921
## X3P            0.06323650 -0.0869921473 -0.079320935 -0.017629191 -0.367081210
## X3PA          -0.19705011  0.0212017817 -0.106798615 -0.031784980  0.229937957
## X3PPercentage  0.25122441 -0.1290203524 -0.009381868 -0.049047581  0.266220936
## FT             0.19824935 -0.0383383906 -0.336377574 -0.059827483  0.590291777
## FTA            0.08578558 -0.0266835584 -0.204963411 -0.163263137 -0.583673952
## FTPercentage   0.31498597 -0.0001979964 -0.380982503  0.248770453 -0.200013367
## ORB           -0.32103875 -0.6082027552 -0.226640784  0.240866741 -0.011626637
## TRB            0.29385082  0.5029345795  0.103011394  0.437393253 -0.003717588
## AST            0.53924781 -0.4416773073  0.226141877 -0.052910259 -0.012302645
## STL            0.18438342 -0.0507166704 -0.260122090  0.399154232 -0.003424513
## BLK            0.13123506 -0.2504526524 -0.259972542  0.088824497  0.017936053
## TOV            0.11048669  0.1522832476 -0.387503430 -0.281039882 -0.012326189
## PF             0.12556328  0.0004570489  0.011958407  0.115577508  0.010349251
##                     Comp.16      Comp.17
## Opponent       0.0214072528  0.007640458
## FG             0.1036415722 -0.634097841
## FGA           -0.0757237129  0.400826044
## FGPercentage  -0.0888319168  0.633431389
## X3P           -0.6179709051 -0.056703359
## X3PA           0.3838314249  0.038278743
## X3PPercentage  0.4509975734  0.040874545
## FT            -0.3313665047 -0.125139451
## FTA            0.3456511022  0.109701835
## FTPercentage   0.1085171579  0.038887239
## ORB            0.0150274144  0.007206852
## TRB            0.0204699595  0.008060575
## AST            0.0063537138 -0.002820216
## STL           -0.0047588560  0.010691819
## BLK            0.0241769168  0.005020483
## TOV           -0.0003993502 -0.010857481
```

```
## PF              0.0146489059  0.004958333
```

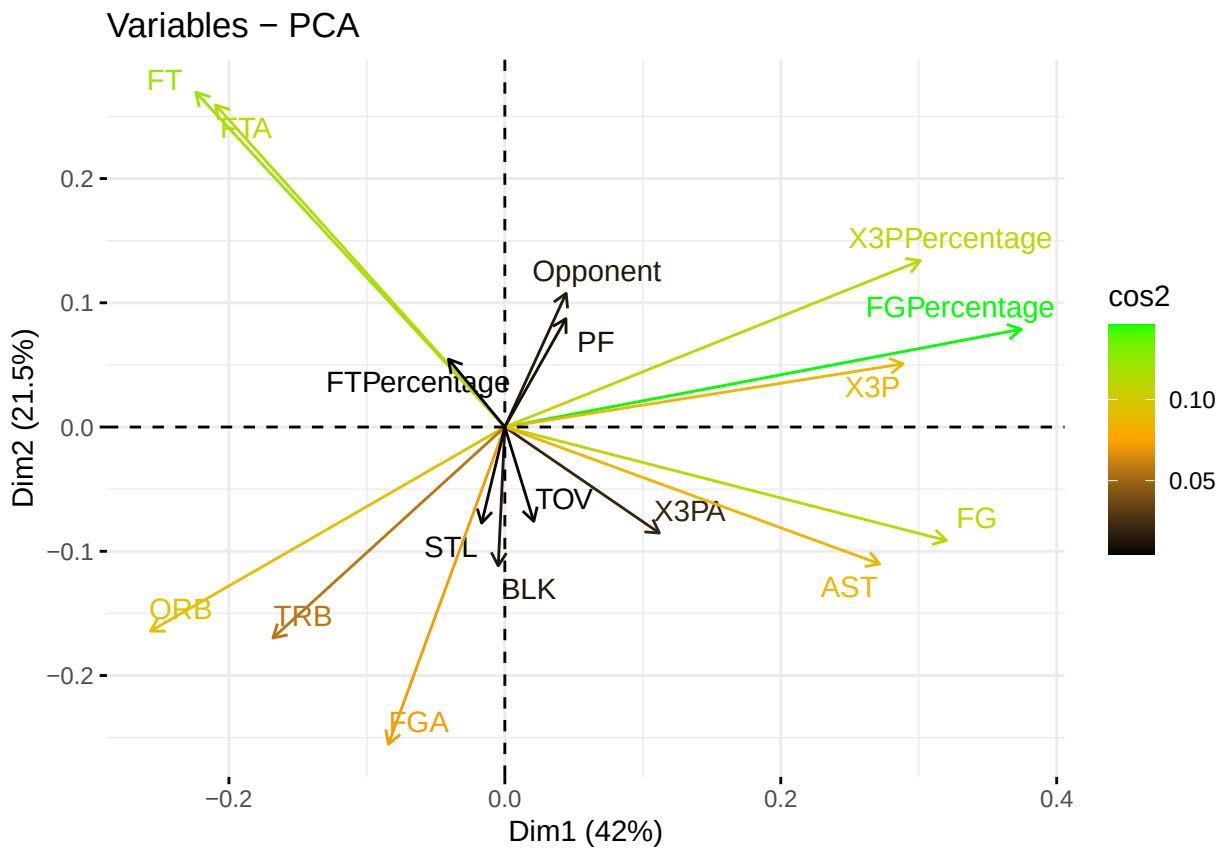
```
# Scree Plot
fviz_eig(data.pca, addlabels = TRUE)
```



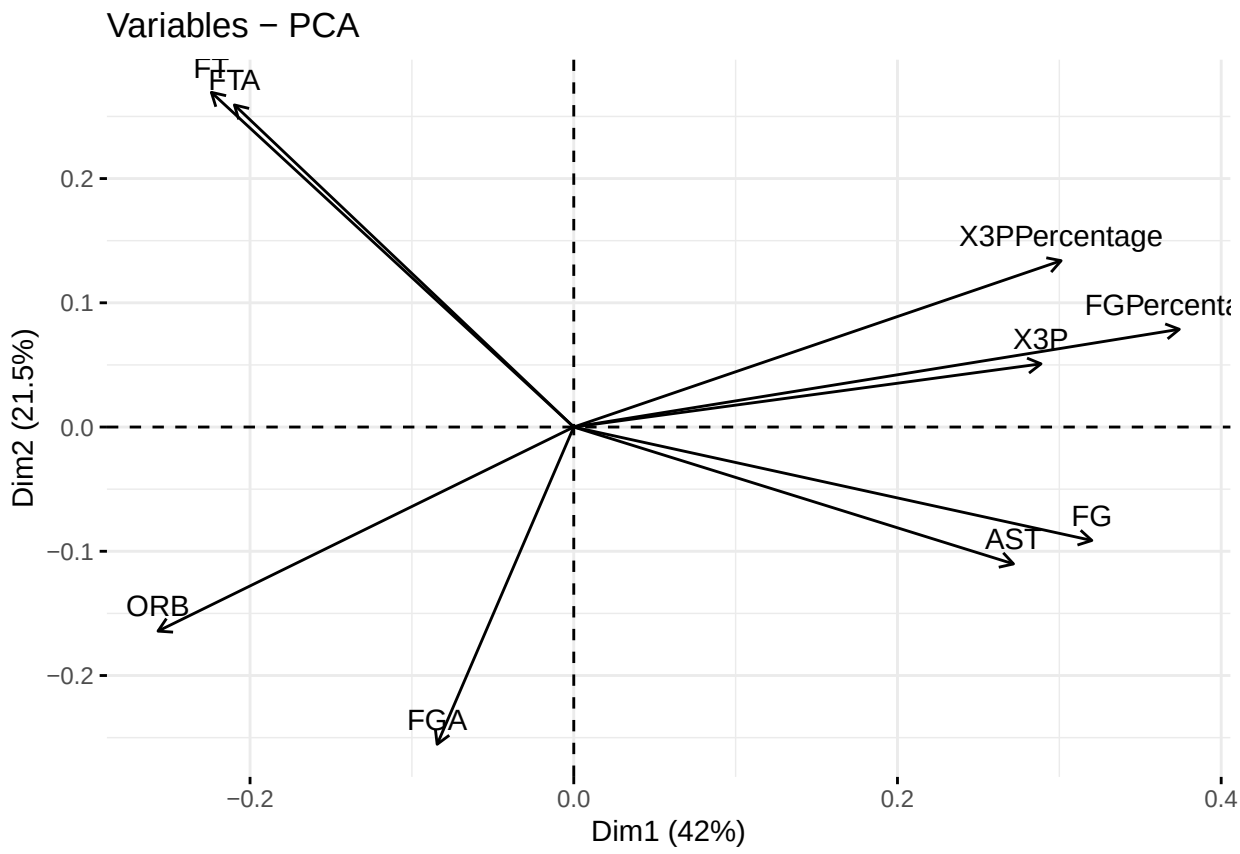

```
# Biplot of the attributes
fviz_pca_var(data.pca, col.var = "black")
```




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



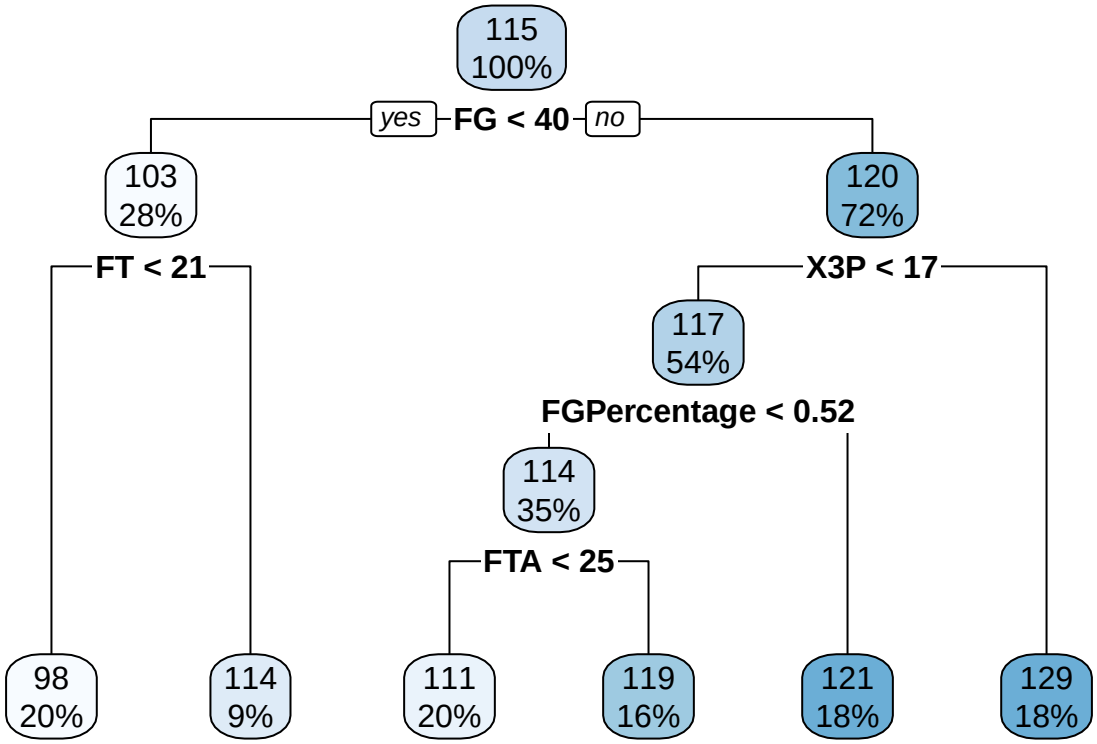

```
# Select the top 9 contributing variable
fviz_pca_var(data.pca, select.var = list(contrib = 9))
```



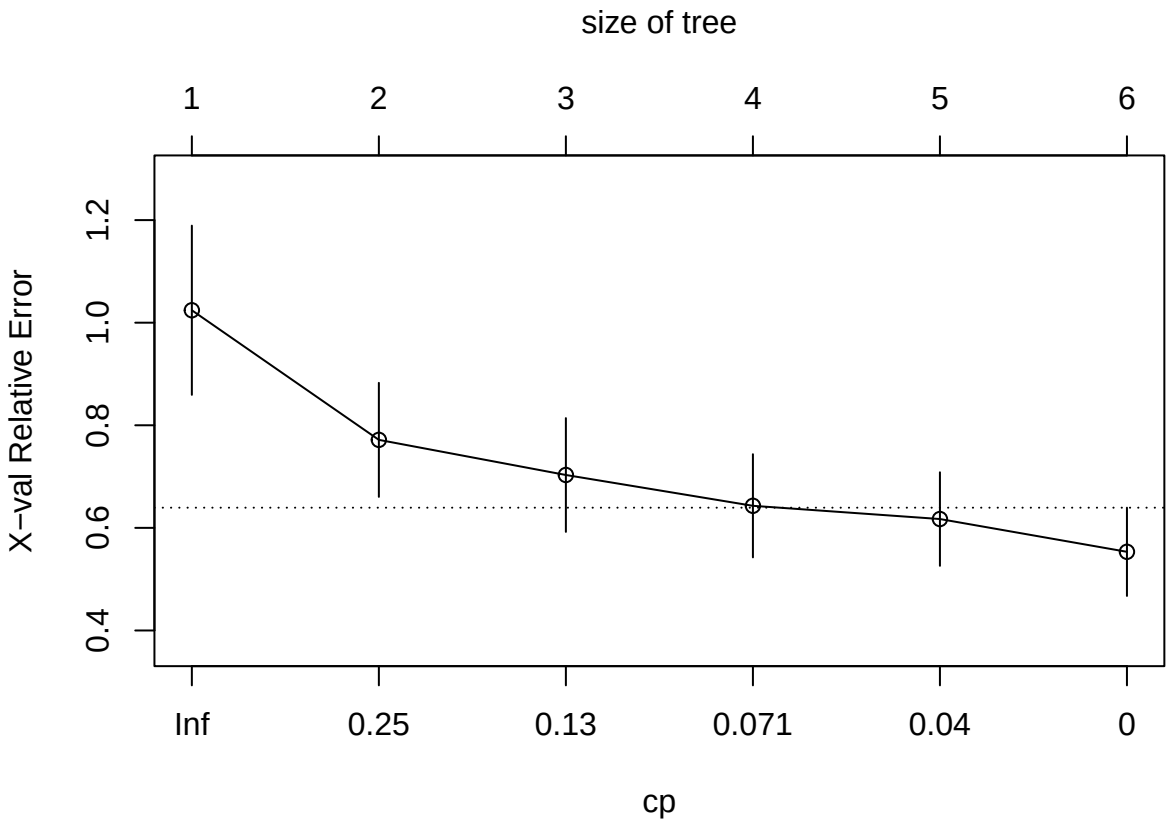

```
# Regression Tree Part

library(rpart)
library(rpart.plot)
fit.tree = rpart(Tm ~ FG+AST+X3PPercentage+X3P+FGPercentage+FT+FTA+FGA+ORB, data=data, method="anova", 
#summary(fit.tree)
fit.tree
```

```
## n= 82
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 82 10952.0100 114.8902
##    2) FG< 39.5 23  2361.4780 102.6087
##      4) FT< 20.5 16   793.9375  97.5625 *
##      5) FT>=20.5 7   228.8571 114.1429 *
##    3) FG>=39.5 59  3768.8810 119.6780
##      6) X3P< 16.5 44  1792.1820 116.6364
##       12) FGPercentage< 0.5175 29   818.5517 114.3448
##         24) FTA< 24.5 16   161.7500 110.8750 *
##         25) FTA>=24.5 13   227.0769 118.6154 *
##       13) FGPercentage>=0.5175 15   526.9333 121.0667 *
##      7) X3P>=16.5 15   375.6000 128.6000 *
```

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




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


size of tree
1
2
3
4
5
6
X−val Relative Error
0.4
0.6
0.8
1.0
1.2
Inf
0.25
0.13
0.071
0.04
0
cp