

# Soccer Manager Assitant

Group 6

April 28, 2017

## Step 0: Load the packages, specify directories

```
packages.used=c("plyr","parallel","plotly","dplyr")

# check packages that need to be installed.
packages.needed=setdiff(packages.used,
                        intersect(installed.packages()[,1],
                                packages.used))

# install additional packages
if(length(packages.needed)>0){
  install.packages(packages.needed, dependencies = TRUE,
                  repos='http://cran.us.r-project.org')
}

library(plotly)
library(dplyr)
library(plyr)
library(parallel)
source("../lib/clean_data.R")
setwd("../data/matches")
matches <- read.csv("matches.csv",header=T)
```

## Part 1: Load and process the data

We first load all data and do the process including data cleaning, player-name matching and saved them as RData.

```
setwd("../data")
Fulldata<-read.csv("Fulldata.csv",header = T)
defender<-read.csv("defence_2015-17.csv",header = T)[,-c(1)]
passer<-read.csv("passing_2015-17.csv",header = T)[,-c(1,3:5)]
scorer<-read.csv("scorer_2015-17.csv",header = T)[,-c(1,3:5)]
```

## Part 2: Clustering and Analysis

We applied k-means cluster method to help soccer managers and coaches find players of certain playing style or strength. We analyzed all k-means cluster results and generated a table named pos\_table (saved in output as pos\_table.RData) which can show players' play styles in different clusters. We analyzed the cluster centers and summarized the characteristics for each cluster.

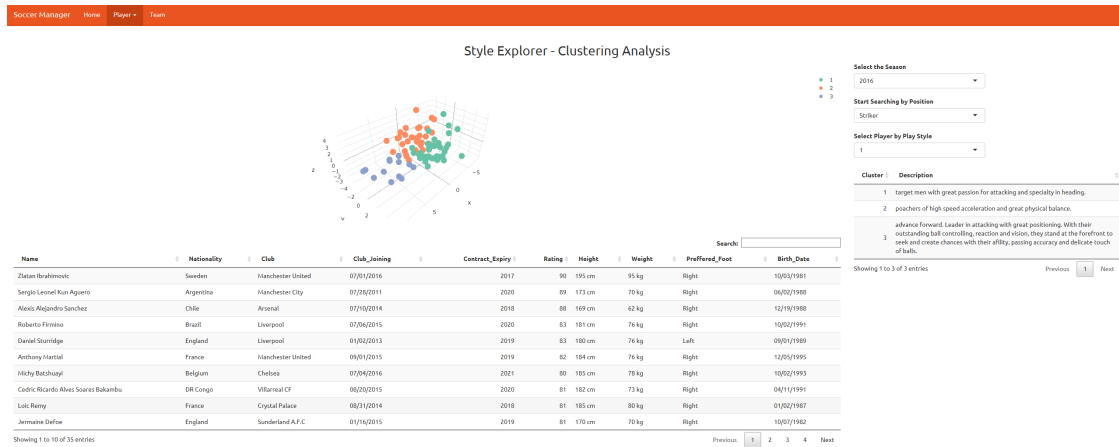
```

cluster_player <- function(Position,year){
  # select rows of a certain position
  select <-
Fulldata[(year==Fulldata$Year&(Fulldata$Pos1==Position)|(Fulldata$Pos2==Position)|(Fulldata$Pos3==Position)),]
  select$Height<-as.numeric(gsub("cm","",select$Height))
  select$Weight<-as.numeric(gsub("kg","",select$Weight))
  select <- t(apply(select[,19:49],1,function(line){line/max(line)}))
  pc <- princomp(select, cor = TRUE, scores=TRUE)
# k means cluster
  set.seed(1)
  KCluster <- kmeans(select, 3, nstart = 20)
  dat1 <- data.frame(cbind(pc$scores[,1:3],KCluster$cluster))
  dat2<-data.frame(cbind(select[,2],select[,6]))
  dat1<-data.frame(cbind(dat1,dat2))
  #clusplot(dat,KCluster$cluster,color = T,shade = T)
  plot_ly(dat1,
    x = dat1[,1],
    y = dat1[,2],
    z = dat1[,3],
    text=paste(dat1$X1,"\t",dat1$X2),
    type = "scatter3d",
    mode = "markers",
    color=factor(dat1$V4),
    hoverinfo="text"
  )
}

```

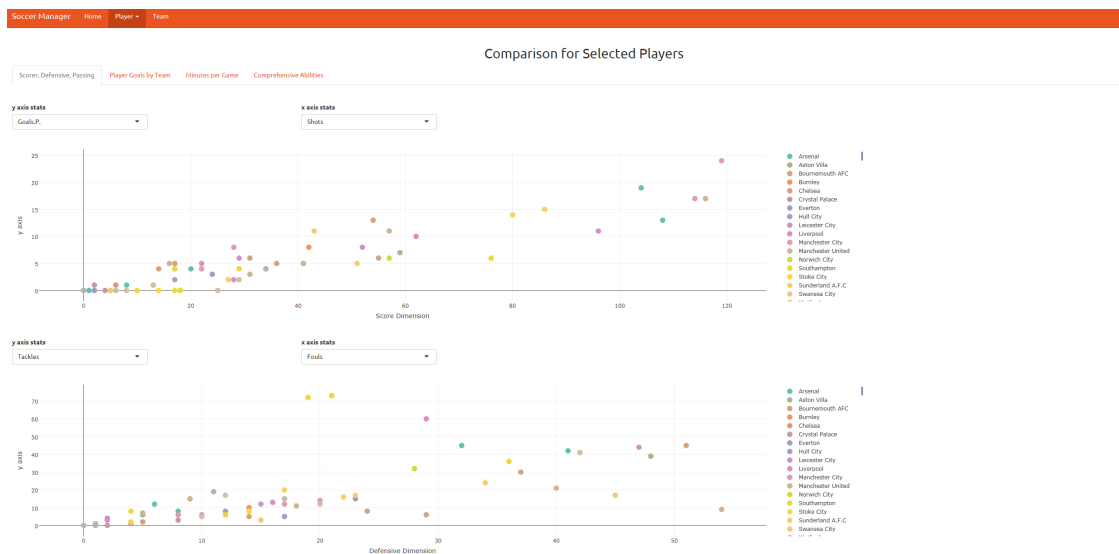
## Part3 Demonstration of our application:

To begin with, Users will select the position of players they want to explore in the "Position and Style" pannel under "Player" tab. The app will perform a K-means clustering based on the technical skills of the player and cluster them into three groups. A summary of each cluster will be given to characterize the style of play for athletes in that group (shown in the table on the right-hand side). furthermore. Some basic info of the players will be given in the table at the bottom.



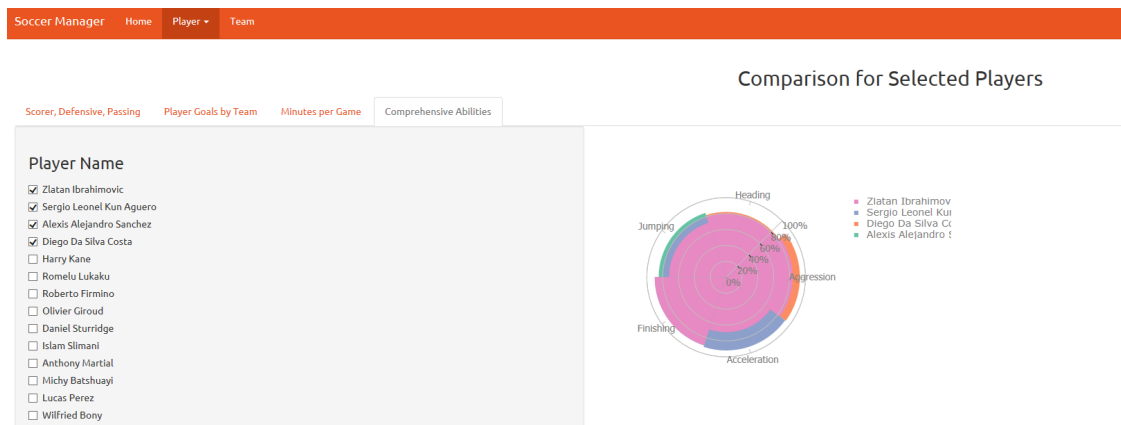
## Cluster

Given selected season, position and cluster number, users can view detailed visualization of player-level data in the "Comparison" pannel. the visualization consists of three sets of plots where users can select Y and X axis from features in socre. defense and passing dimensions.



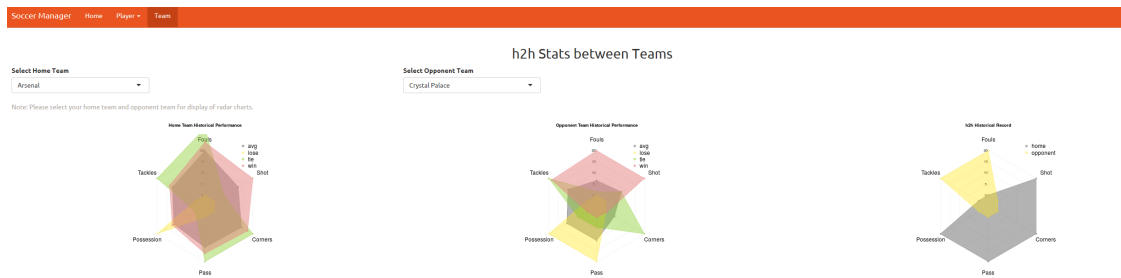
## SDP

Some more visualizations can be found in the next few pannels under Comparison. Finally, users will be able to compare several players of their interest in the "Comprehensive Abilities" by checking their name in the check box.



## Comparison

Finally, our app help coaches to be better prepared for the next game by showing radar charts on their Head-to-Head performance with selected opponents.



## H2H

Moving on, we would like to collect more player-level data for each game so we can optimize the line-up and formations and give recommendations to coaches before each game. Coaches can use the results and decide the players they need and choose the appropriate players against different opponents.