NBA ANALYTICS

ObJECTIVE

The objective of this project is to visualize and understand the performance of the nba player and teams –for NBA dataset for 2007-2008 & 2017-18.

About DATASET

- Source - The raw data for this function is obtained from the website http://dougstats.com
- Variables Used-sliced from original data

```
head(nba1718)
    League             Name Team GamesPlayed TotalMinutesPlayed FieldGoalsMad
e FieldGoalsAttempted
36     NBA    Kent Bazemore  ATL          65               1787            28
4                   676
42     NBA  Deandre' Bembry  ATL          26                453             5
3                   128
74     NBA Nicolas Brussino  ATL           3                 10
0                     2
92     NBA  Tyler Cavanaugh  ATL          37                521             6
7                   152
100    NBA  Anton Cleveland  ATL           4                 41
4                     7
102    NBA      John Collins  ATL          74               1784            31
4                   546
    PersonalFouls Disqualifications TotalPoints
36            147                 1         836
42             38                 0         136
74              0                 0           0
92             60                 0         183
100            12                 1          13
102           215                 4         777
```

i. League-NBA(Name of the League)

ii. Name-Name of the pLayer

iii. Team-Team to which the player belongs

iv. Gamesplayed - Number of games played by the player of the team

v. Totalminutesplayed – Number of minutes a player played.

vi. Feildgoalsmade – number of goals made by a player

vii. Fieldgoalsattempted - number of goals ATTEMPTED by a PLAYER.

viii. PERSONALFOULS – Number of fouls Made BY player.

ix. Disqualifications – NUMBEr of times a player is- disqualified.

x. TOTALpoints – Number of points scored by a player.

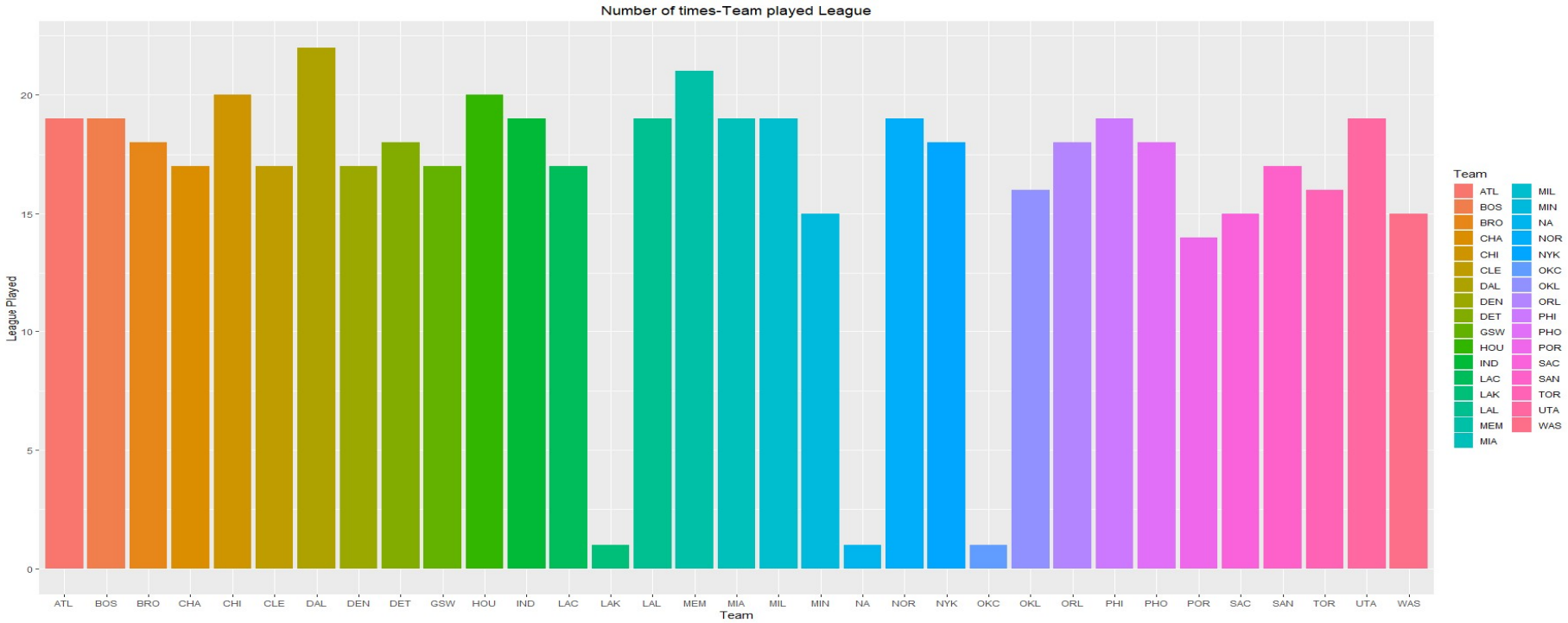
Data preprocessing/cleaning

- For most of the part, the data was pretty clean. Only Few codes are executed to remove blank spaces, punctuation marks etc.

```
#completecases
nba1718 <- nba1718[complete.cases(nba1718),]
nba1718$Name <- sapply(nba1718$Name,as.character)
#Removing unwanted character from Player-name
nba1718$Name <- gsub(".*,","",nba1718$Name)
#remove white-Space
nba1718$Name <- str_trim(nba1718$Name,side = "both")
```

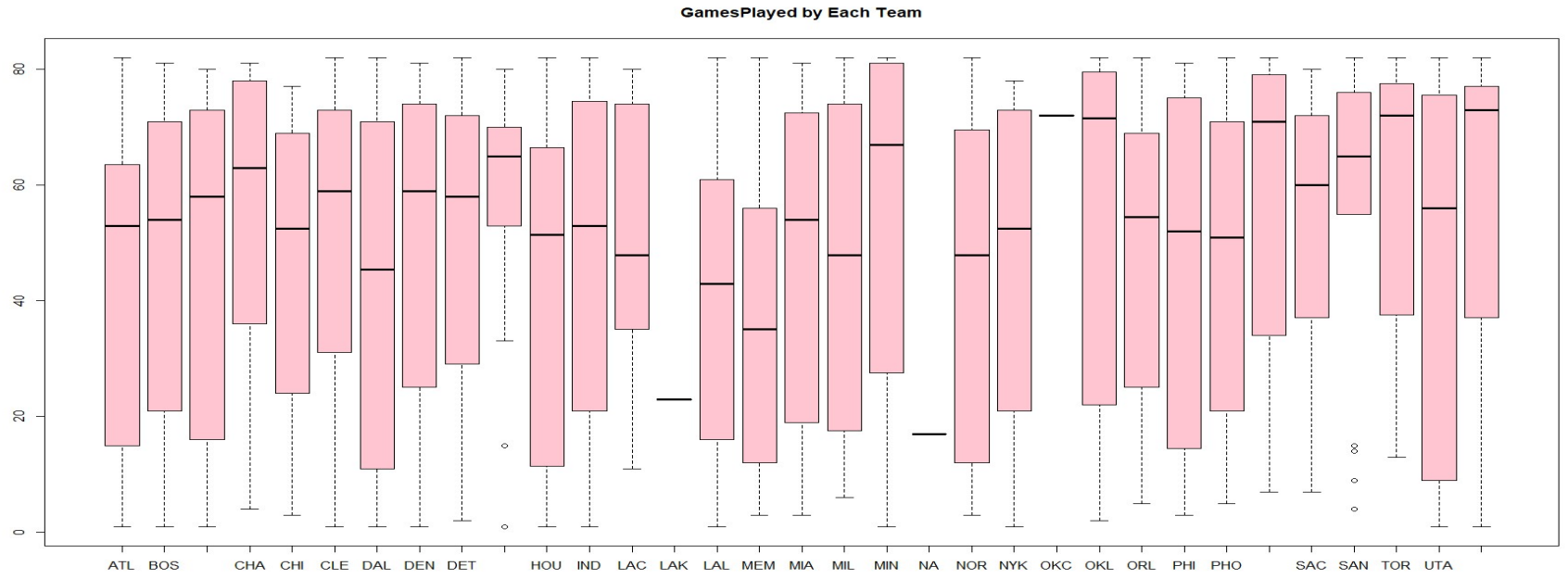
ANALYZING DATA

- As the first Step we Visualize data- here we show the number of times a team played in a nba league.



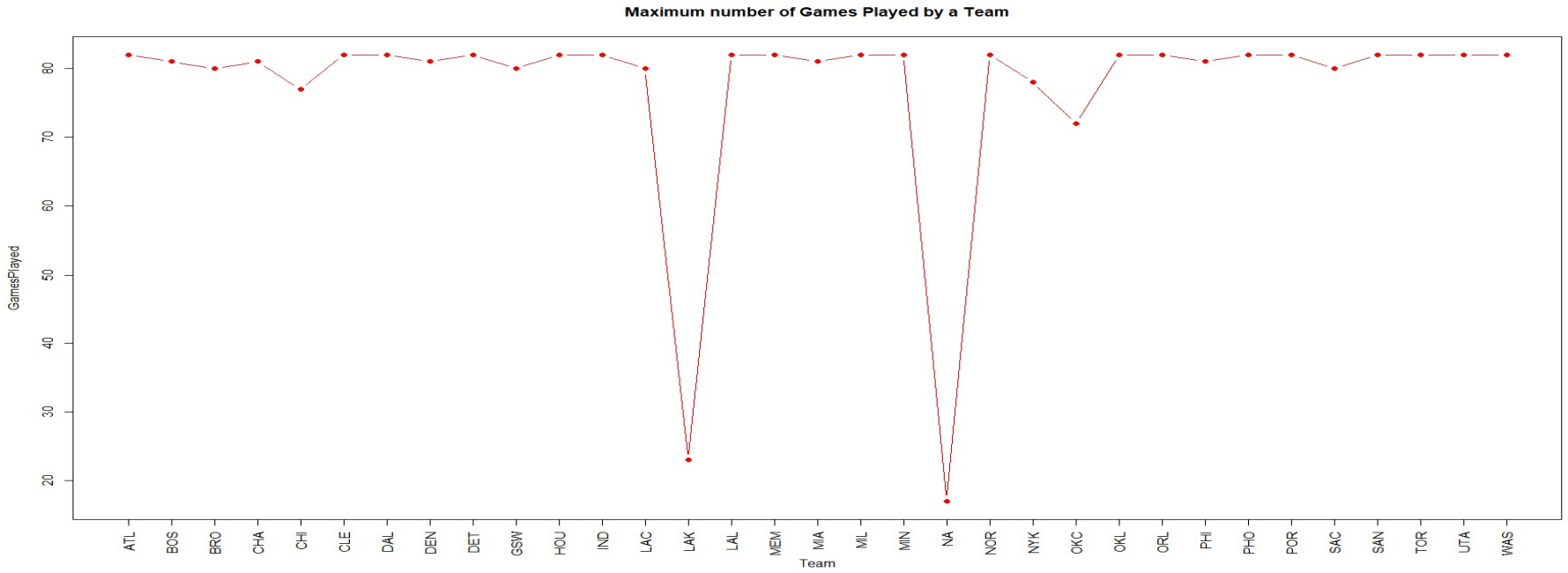
TEAm vs GAMESplayed

- Here team –being a categorical variable and Gamesplayed, a numerical variable we draw their relationship as follows:-



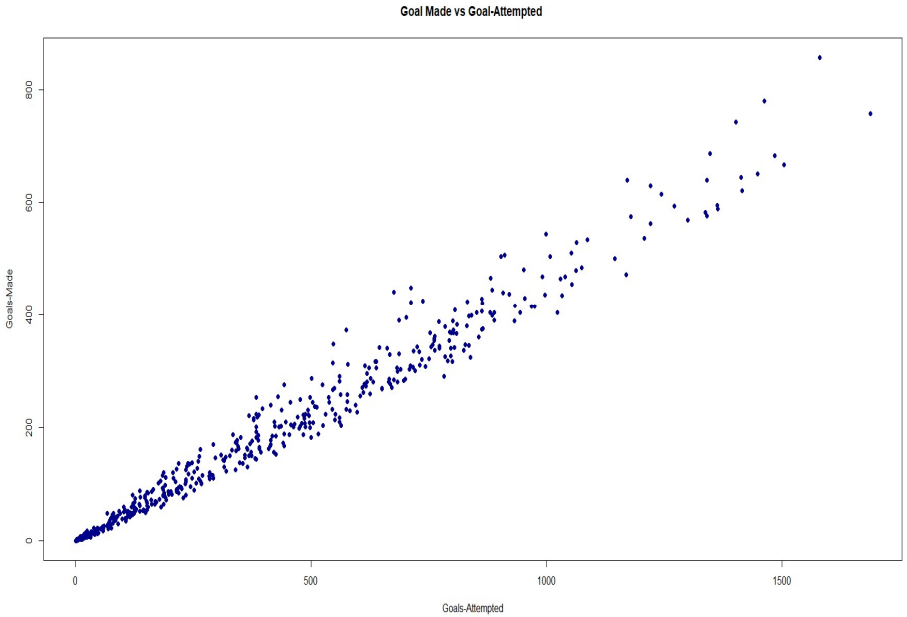
Maximum number of games by a team

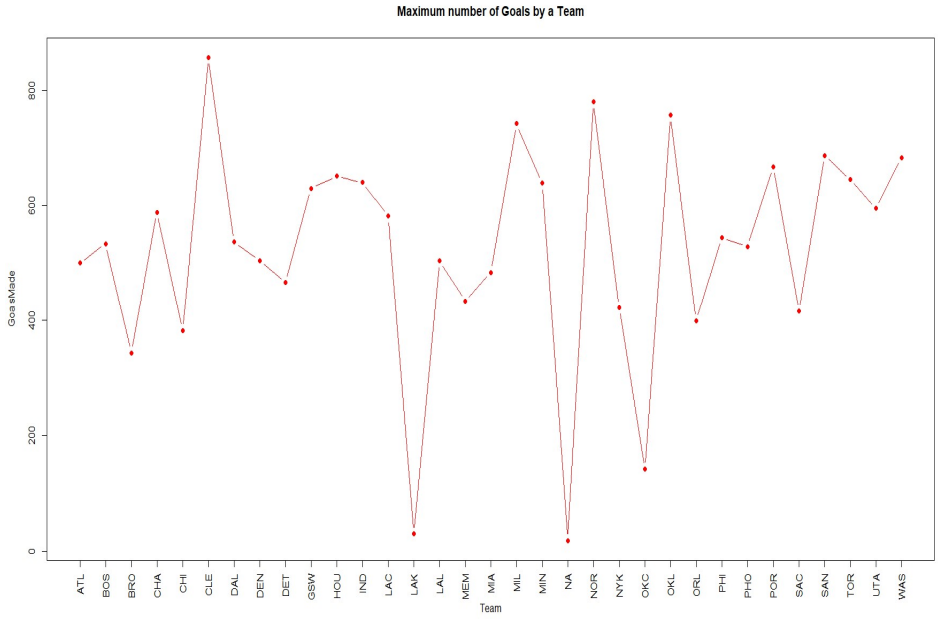
- Here we summarize maximum number of games played by a team



Success Rate-for GOALs

- He we use both number variable –goals made vs attempted
- Goal MADE VS GOAL ATTeMPTeD MAXIMUM NUMBER OF



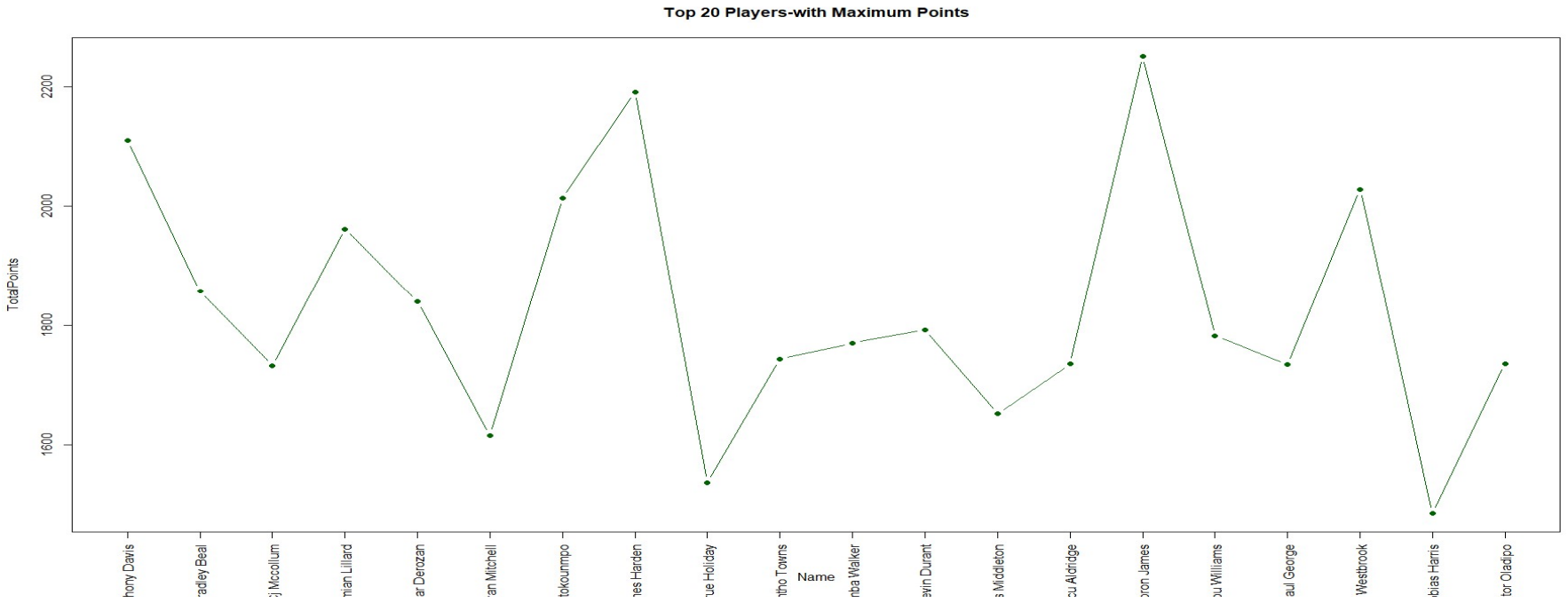



```
average_success_for_goals <- 100*(mean(FieldGoalsMade)/mean(FieldGoalsAttempt
ed))
average_success_for_goals
```

```
## [1] 46.03352
```

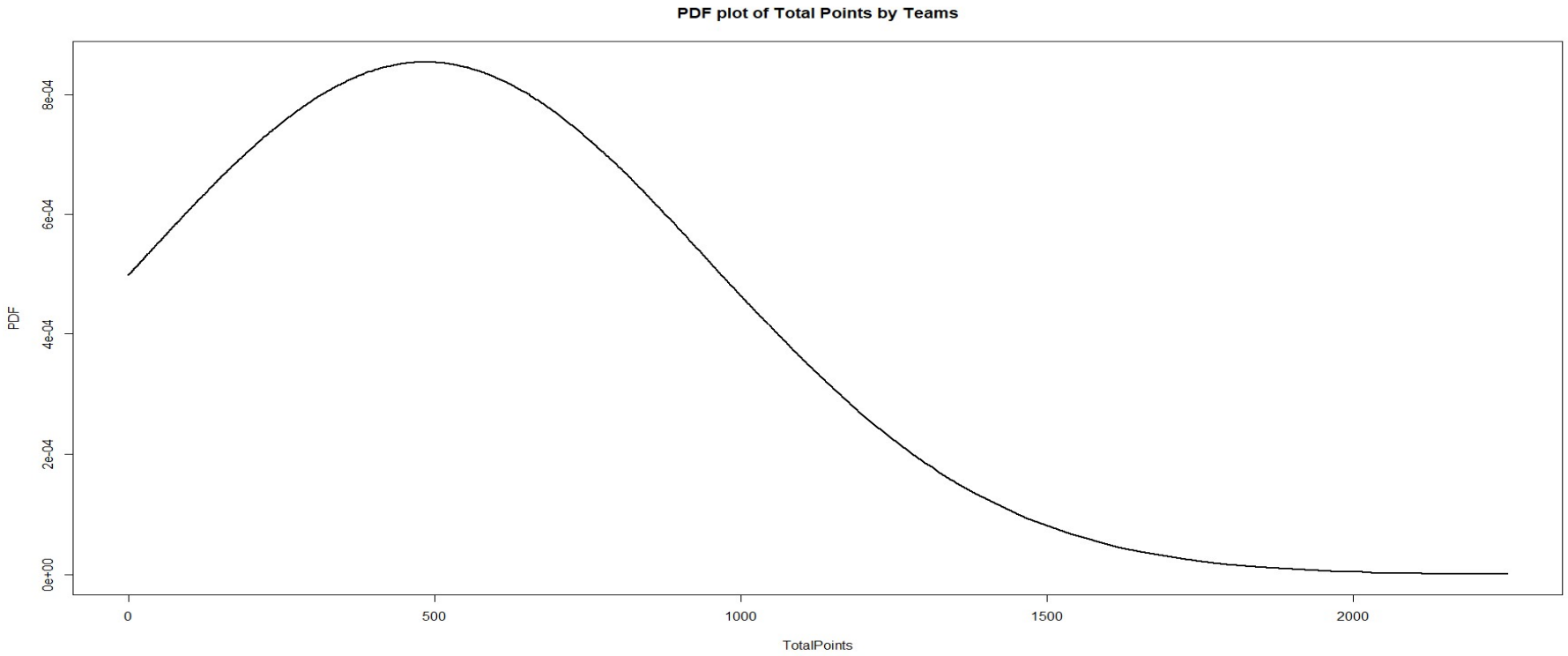
Finding Top 20- Player vs maximum points

- We take player as a categorical variable and total points numerical variable for and find top 20 players on the basis of their total points.



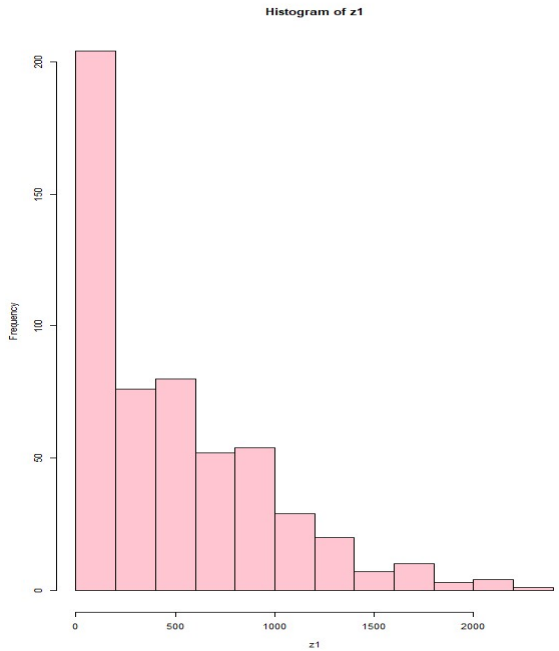
Distribution

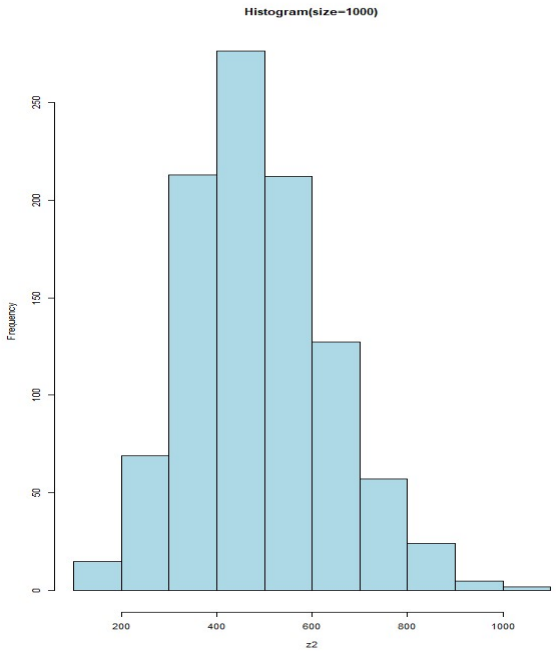
- We take total points as numerical variable and PLOT pDF for it.

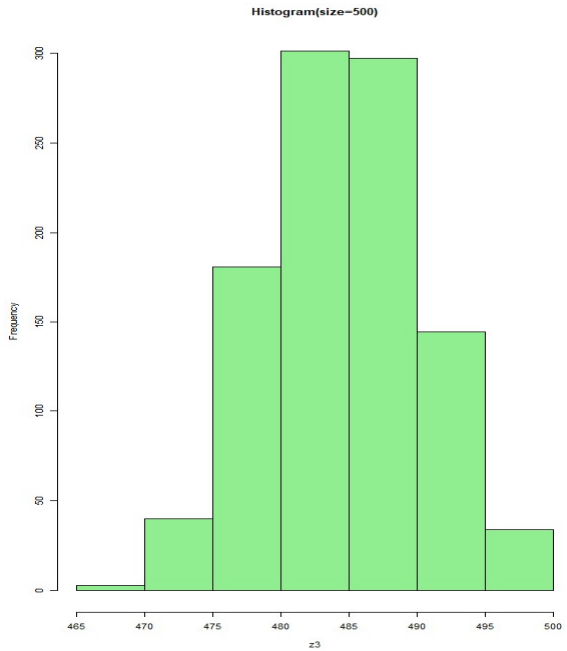


CENTRAL LIMIT THEOREM

- applying central limit theorem approach on totalpoints score by a player –using three sets of samples

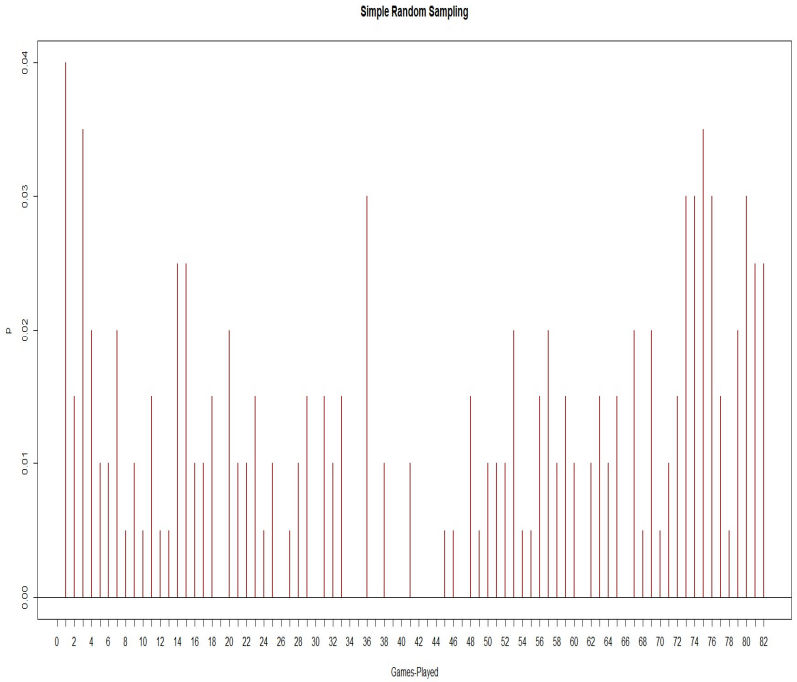




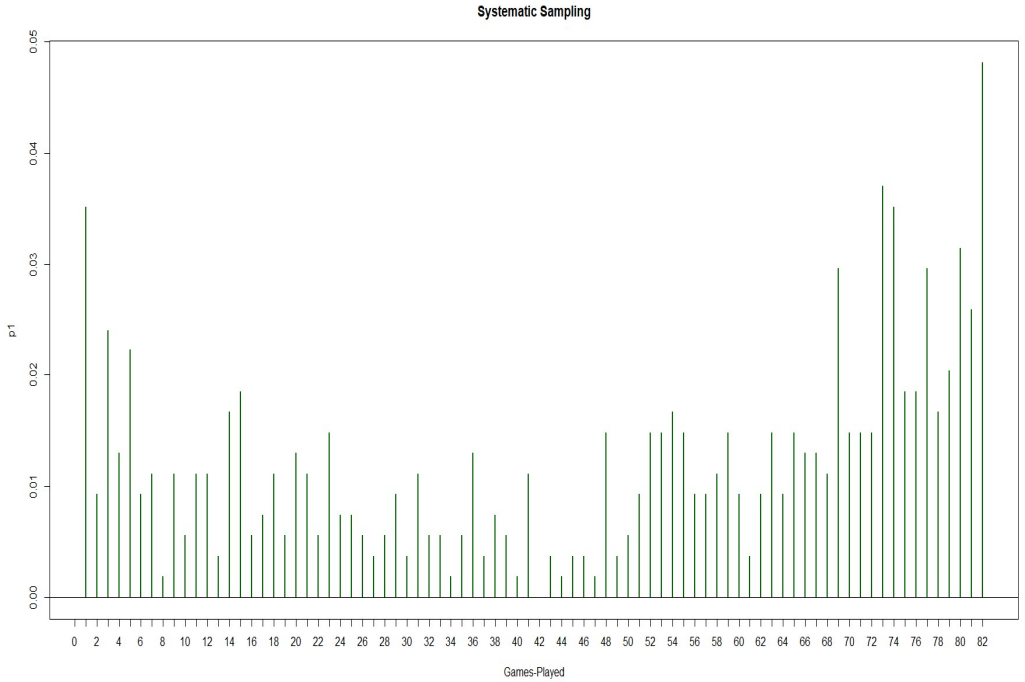


SAmplING

- Showing various sampling methods- on number of Gamesplayed
- SIMPLE RANDOM SAMPLING



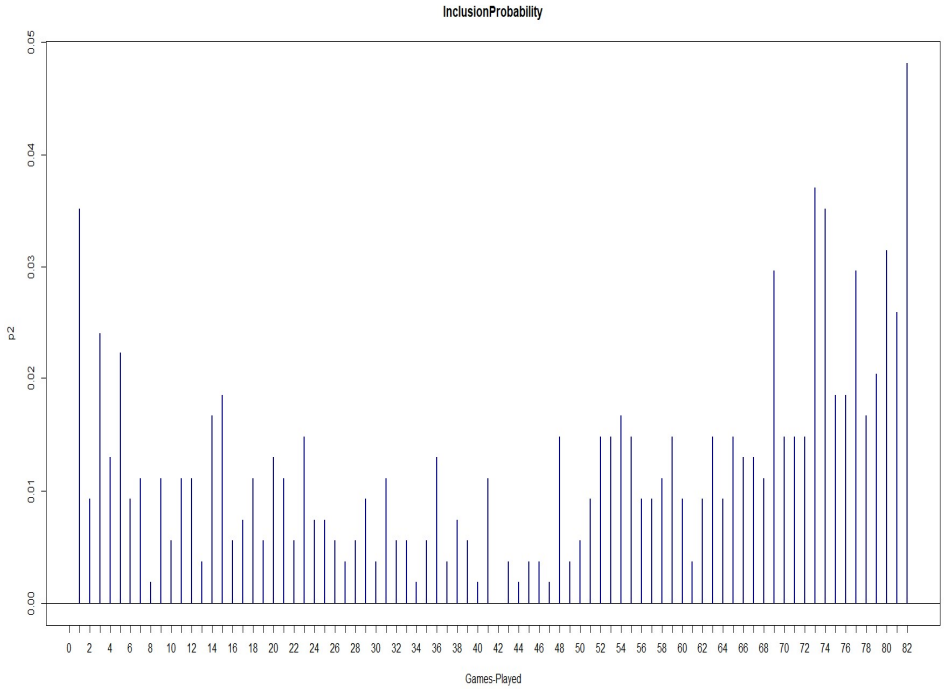
SYSTEMATIC SAMPLING

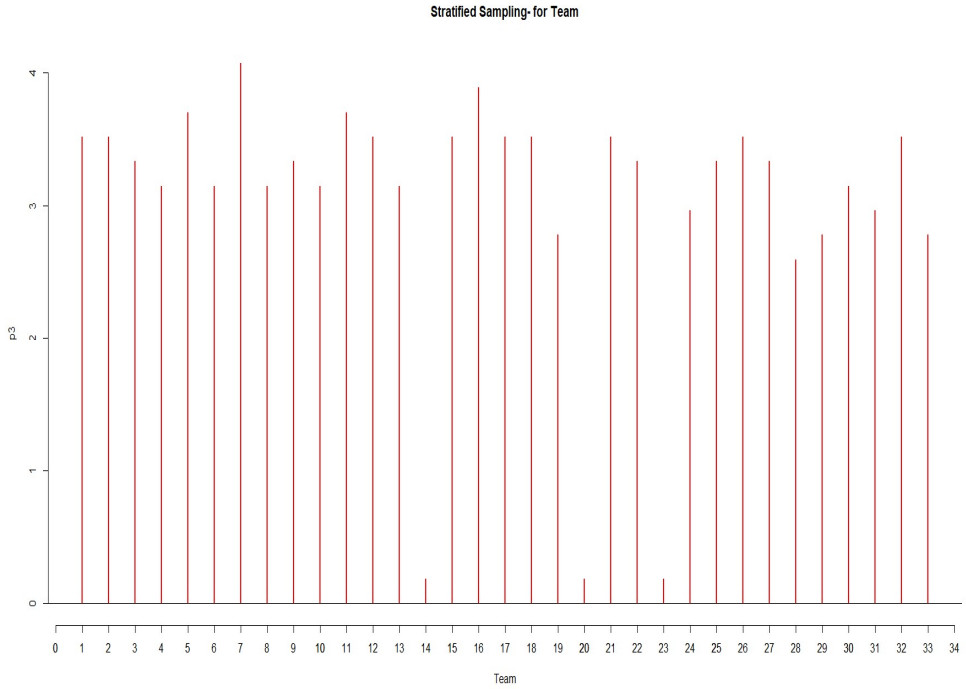


SAMPLING…cntd

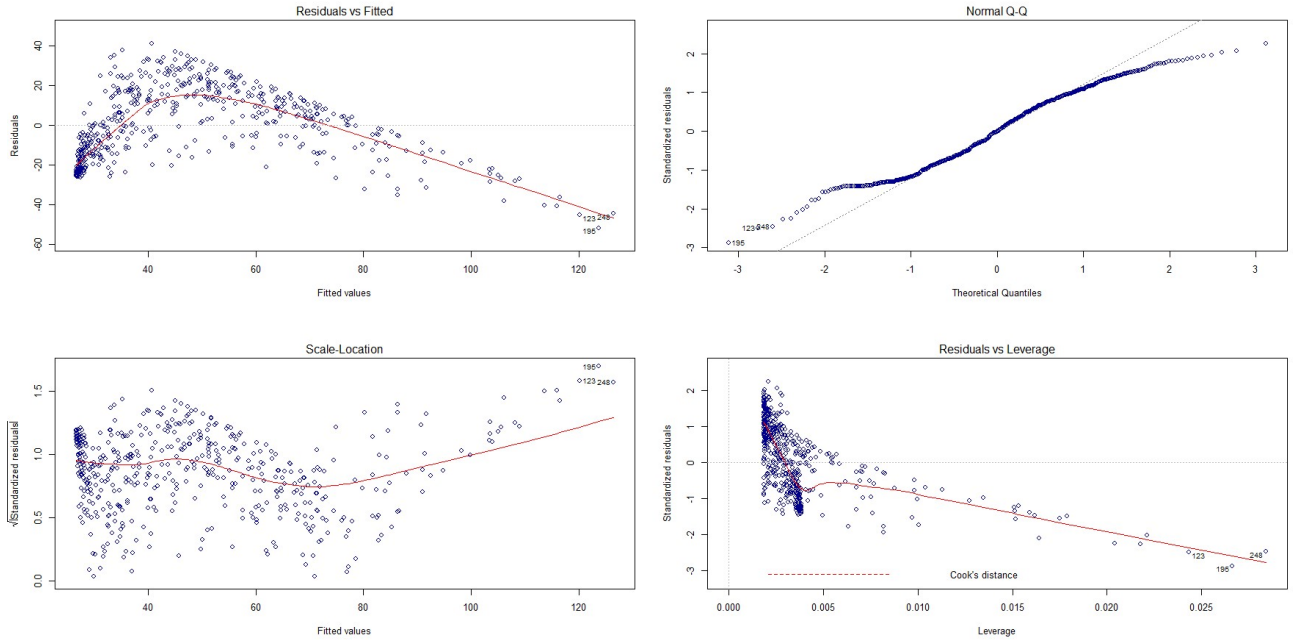
- Inclusion probabilities

STRATIFIED SAMPLING ON TEAM





Correlation and Residuals plot- Gamesplayed vs TotalPOINTS



Preparing Data aND Loading LIBRARIES

- AS the FIRST STEP I load all the required libraries

```
> library(rvest)
> library(SportsAnalytics)
> library(tidyverse)
> library(dplyr)
> library(plyr)
> library(RCurl)
> library(RJSONIO)
>
> nba0708 <- fetch_NBAPlayerStatistics("07-08")
```

Subset DATA BY TEAm(BOSTON) AND PROCESS

- We CHOSE and subset data for *BOSTON* .
- *Apply filters for FINDINGS-"RAY ALLEN "-has best three point %,"Paul Pierce" has largest number of minutes and " Rajon Rondo" has most "ste*

```
> nba0708.bos <- subset(nba0708, Team == 'BOS')
>
> #Q1
>
> nba0708.bos%>%
+   mutate(three.point.percentage=ThreesMade/ThreesAttempted)%>%
+   filter(three.point.percentage==max(three.point.percentage,na.rm=T))%>%
+   select(Name,three.point.percentage)
       Name three.point.percentage
1 Ray Allen              0.3982301
>
> #Q2
> nba0708.bos%>%
+   filter(TotalMinutesPlayed==max(TotalMinutesPlayed))%>%
+   select(Name,TotalMinutesPlayed)
          Name TotalMinutesPlayed
1 Paul Pierce               2873
>
> #Q3
> nba0708.bos%>%
+   filter(Steals==max(Steals))%>%
+   select(Name,Steals)
          Name Steals
1 Rajon Rondo    129
```

Five 5 –with most wins

- Here I have used totalpoints to decide for the wins for each team.

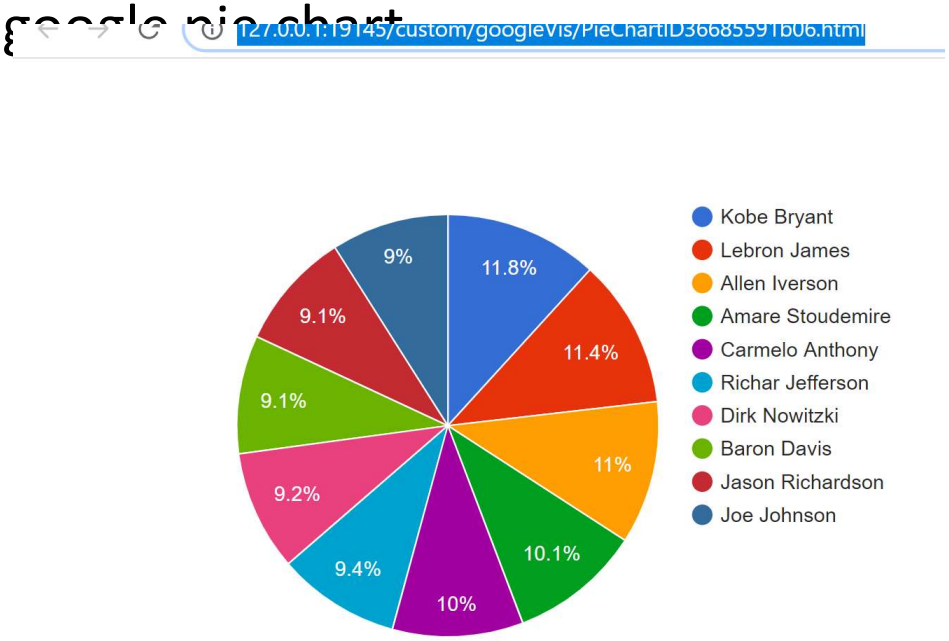
```
> Top.5.Team <- nba0708 %>% select(Team,TotalPoints)
> Top.5.Team <- Top.5.Team[order(Top.5.Team$TotalPoints,decreasing = T),]
> head(Top.5.Team,5)
    Team TotalPoints
67   LAL         2323
211  CLE         2250
203  DEN         2164
385  PHO         1989
15   DEN         1978
```

- Alternate-for other data- I have
- An experimented on A different
- to extract number of wins.

```
> url <- read_html("https://www.basketball-reference.com/leagues/NBA_2008.htm
l")
> data<-html_nodes(x=url,css=".full_table .left+ .right , .full_table .left")
%>%html_text()
>
> team<-data[c(TRUE,FALSE)] # extract teams' names
> wins<-data[c(FALSE,TRUE)]
> #clean data
> team<-gsub(pattern =  "\\([^()]+\\)","",team)
> team<-gsub("*", "", team)
> team<-gsub("\\*", "", team)
> team<-gsub(intToUtf8(160),'',team)
> teamwins<-data.frame(team,wins)
>
> teamwins$wins<-as.numeric(as.character(teamwins$wins))
> Top.5<-teamwins%>%top_n(5,wins)
> Top.5
                team wins
1      Boston Celtics   66
2     Detroit Pistons   59
3  Los Angeles Lakers   57
4 New Orleans Hornets   56
5   San Antonio Spurs   56
```

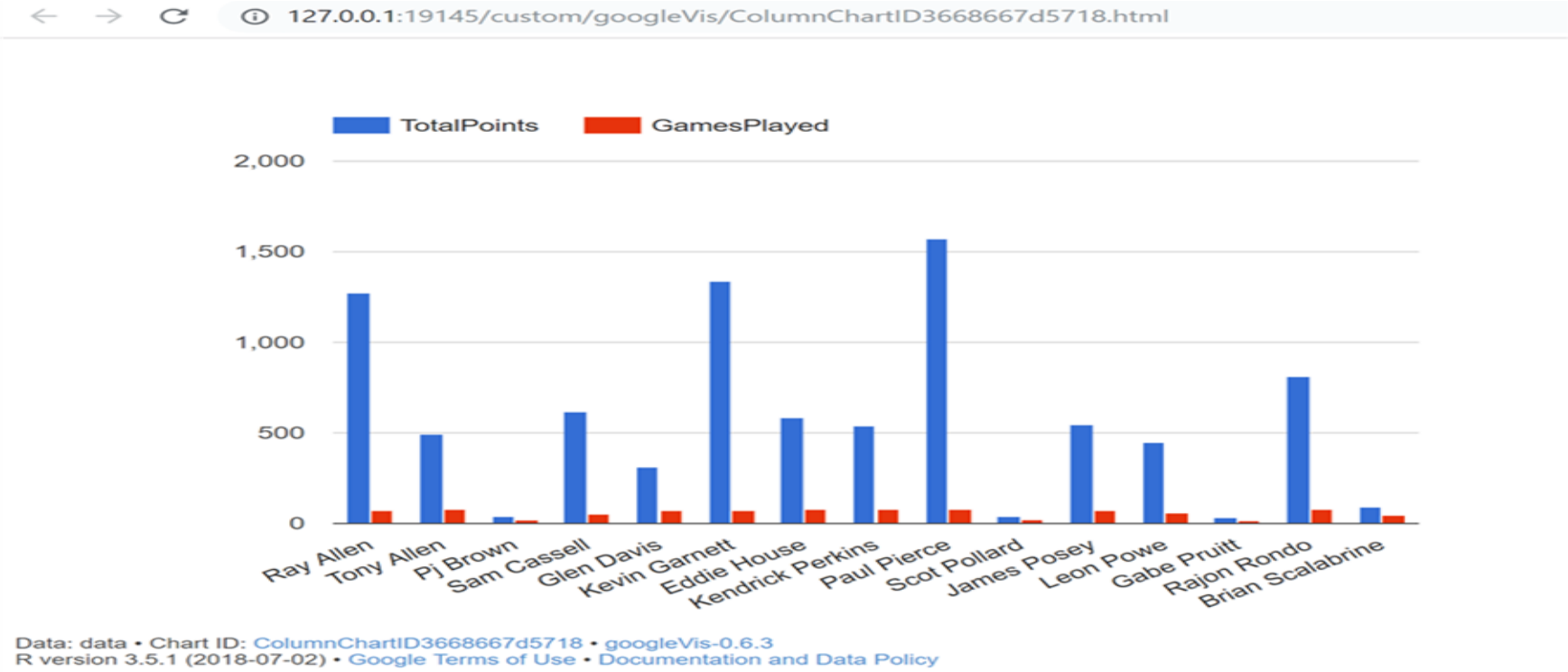
Five Google CHArts-CHART1(PIE-CHART)

- CHART-1(PIE-CHART):On the BASIS of TOTalpoints I have retrieved Top 10 players and plot a google pie chart



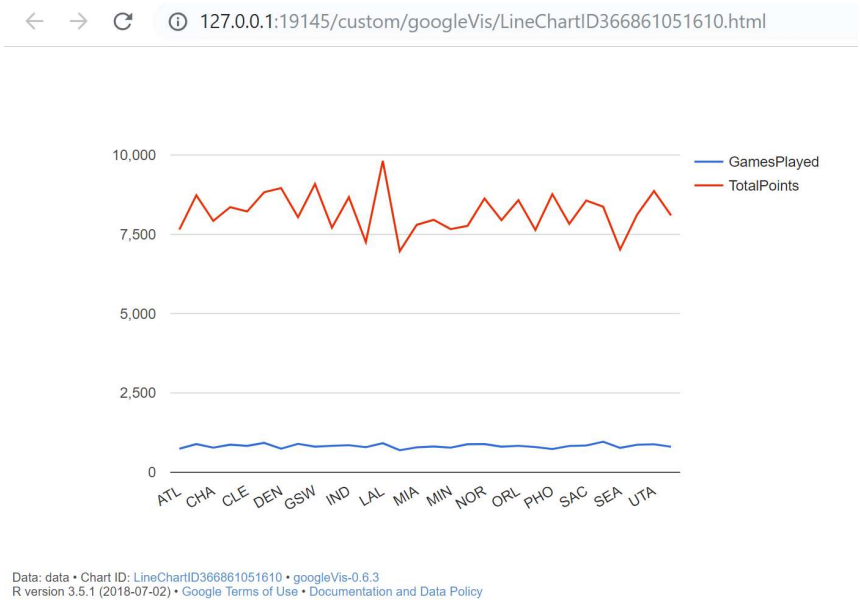
Five Google CHArts-CHART2 COLUmn CHART

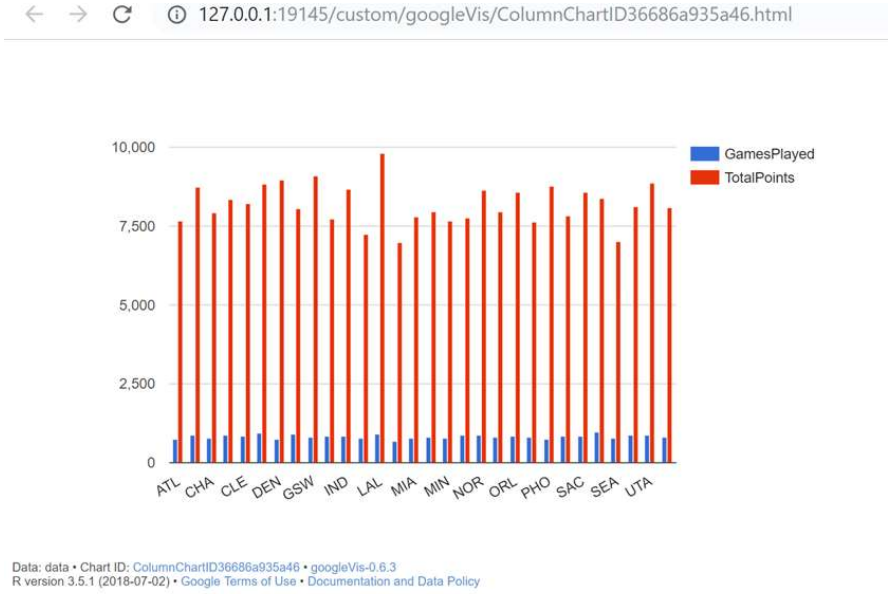
- CHART-2(COLUMN CHART):Here I have used TOTAL POINTS AND GAMES PLAYED to reflect performance FOR EACH PLAYER.



Five Google CHArts-CHART3 LINE/COLUMN CHART

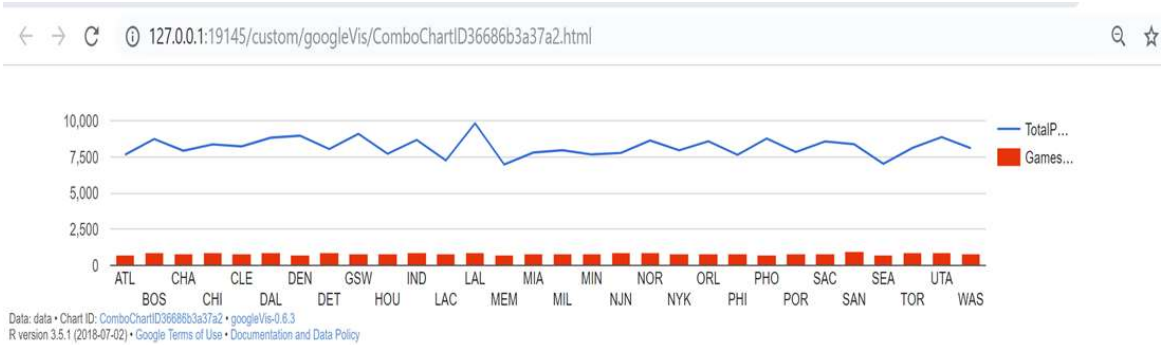
- CHART-3(Line/COLUMN CHART): Here I have used TOTAL POINTS AND GAMES PLAYED to reflect performance FOR EACH TEAMS.





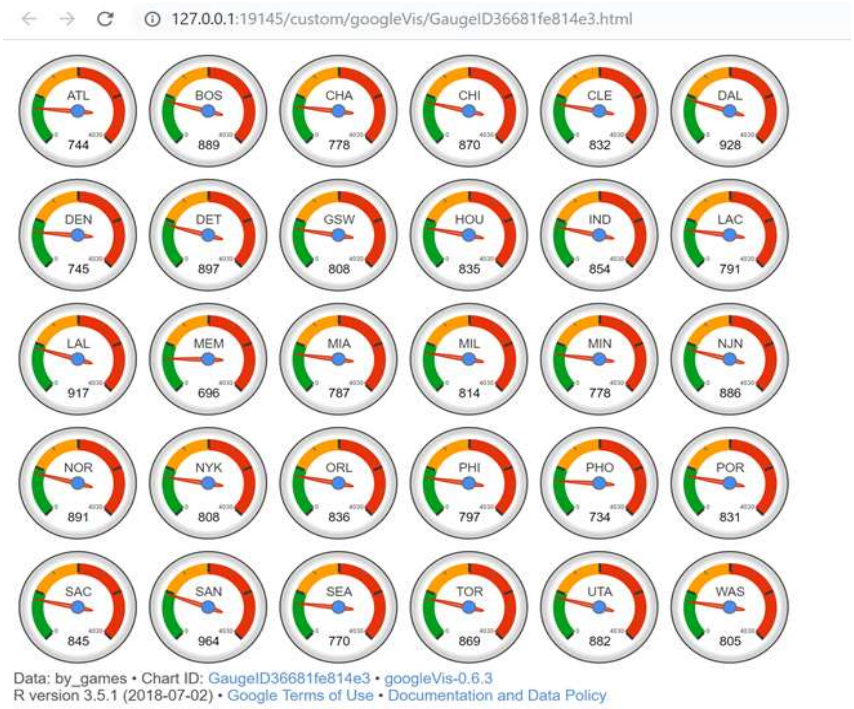
Five Google CHArts-CHART4 COMBO/BUBBLE CHART

- CHART-4(COMBO/BUBBLE CHART): Here I have used TOTAL POINTS AND GAMES PLAYED to reflect performance FOR EACH TEAMS.



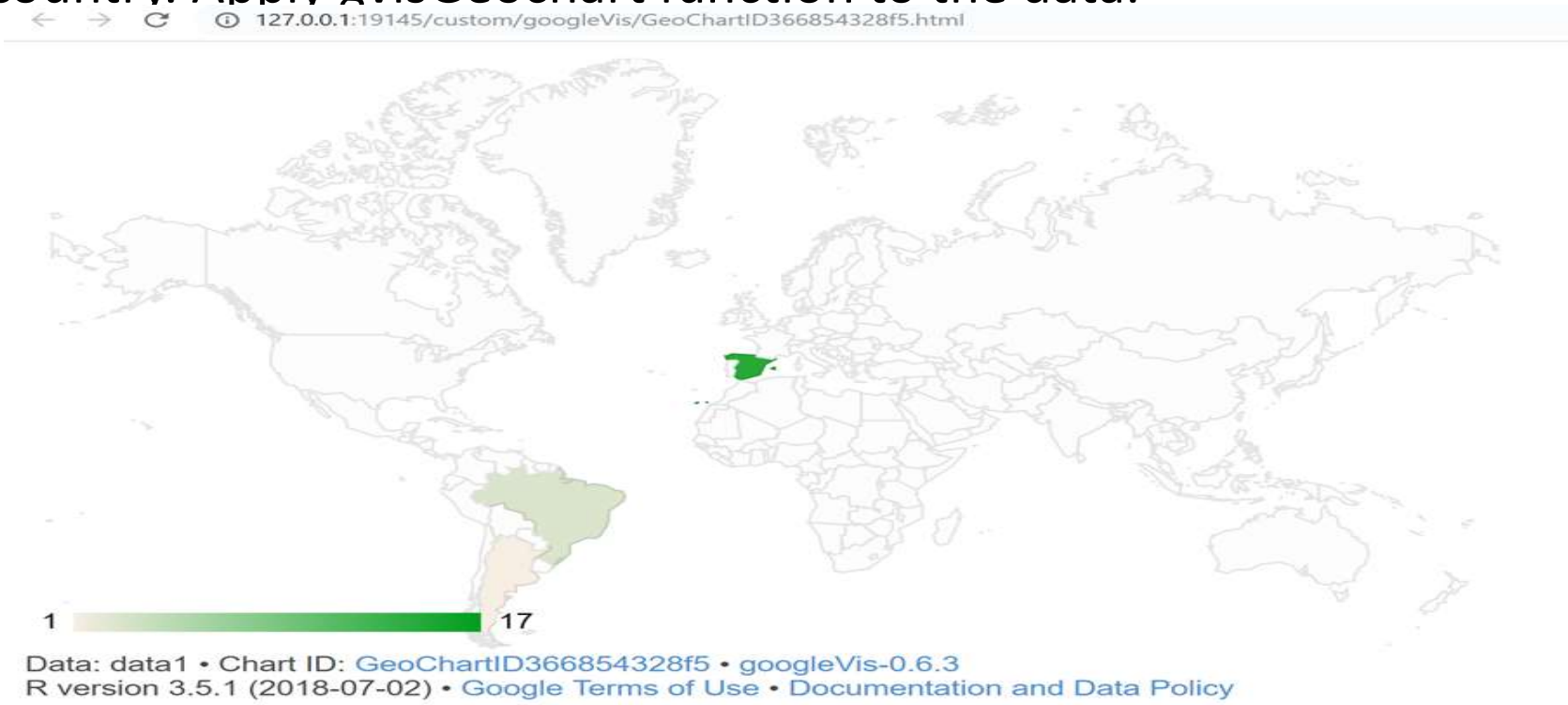
Five Google CHArt5-CHART GAUGE CHART

- CHART-5(gAUGE CHART): Here I have used TOTAL POINTS AND GAMES PLAYED to reflect performance FOR EACH TEAMS.



GOOGLE GEOCHART-ALL BASKETBALL CHAMPION

- GEOCHART CHART:CLEAN the data- form a dataframe for eyar and country. Apply gvisGeochart function to the data.



SUMMARY

- The analysis conducted to determine the performance of the player in the nba 2017-18.
- Also on the basis of total point scored by a player-top 20 players is determined.
- Success rate of the goals is calculated to be 46%~.
- IF THE SAMPLING METHODS-on totalpoints ARE USED Instead of whole-data set then the objective of analyzing the performance of the player is served as it gives the visualization of probability distribution for a specified sample.