IPL(2008-2020) Team Analysis

Using R

By-Kamleshwar Dubey

IPL(2008-2020) Team analysis Using R

Kamleshwar

2023-04-23

Loading library

first we load library dplyr to work on data and ggplot2 for data visualisartion

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
```

Reding Data file

First we read csv file

```
setwd("C:/Users/Admin/Downloads/")
d <- read.csv("IPL Matches 2008-2020.csv", header = TRUE, sep = ",")
tbl <- as_tibble(d)
```

taking overview of this data

```
#taking overview of data
head(tbl)

## # A tibble: 6 × 17
##       id city     date  playe…¹ venue neutr…² team1 team2 toss_…³ toss_…⁴
winner
##    <int> <chr>    <chr> <chr>   <chr>   <int> <chr> <chr> <chr>   <chr>
<chr>
## 1 335982 Bangalo… 4/18… BB McC… M Ch…       0 Roya… Kolk… Royal … field
Kolka…
## 2 335983 Chandig… 4/19… MEK Hu… Punj…       0 King… Chen… Chenna… bat
Chenn…
## 3 335984 Delhi    4/19… MF Mah… Fero…       0 Delh… Raja… Rajast… bat
Delhi…
```

```
## 4 335985 Mumbai   4/20… MV Bou… Wank…       0 Mumb… Roya… Mumbai… bat
Royal…
## 5 335986 Kolkata  4/20… DJ Hus… Eden…       0 Kolk… Decc… Deccan… bat
Kolka…
## 6 335987 Jaipur   4/21… SR Wat… Sawa…       0 Raja… King… Kings … bat
Rajas…
## # … with 6 more variables: result <chr>, result_margin <int>, eliminator
<chr>,
## #   method <chr>, umpire1 <chr>, umpire2 <chr>, and abbreviated variable
names
## #   ¹player_of_match, ²neutral_venue, ³toss_winner, ⁴toss_decision
```

Total match

Now we check how many match played by each team in decsending order i.e.whichh team played most match and which played least match.

```
# Grouping team1 column
t1 <- group_by(d,team1)
# Grouping team1 column
t2 <- group_by(d,team2)

#count total match of any team as team1
c_t1 <- count(t1)
# renaming n
c_t1 <- rename(c_t1, `Total Match as Team 1` = n,)
#count total match of any team as team1
c_t2 <- count(t2)
# renaming n
c_t2 <- rename(c_t2, `Total Match as Team 2` = n,)

# adding total no of match played by any team (as sum of team1 and team2 )
Sum <- c_t1$`Total Match as Team 1` + c_t2$`Total Match as Team 2`

#creating new data frame for total match played by team
total_match <- data.frame(`Team`=c_t1$team1,`match`=Sum)

# sorting total match played by team in descending order
total_match_sorted <- arrange(total_match,desc(match))
print(total_match_sorted)
```

```
##                            Team match
## 1                Mumbai Indians   203
## 2   Royal Challengers Bangalore   195
## 3         Kolkata Knight Riders   192
## 4               Kings XI Punjab   190
## 5           Chennai Super Kings   178
## 6              Delhi Daredevils   161
## 7              Rajasthan Royals   161
## 8           Sunrisers Hyderabad   124
## 9              Deccan Chargers     75
## 10               Pune Warriors    46
## 11              Delhi Capitals    33
## 12               Gujarat Lions    30
## 13      Rising Pune Supergiant    30
## 14        Kochi Tuskers Kerala    14
```

ploting graph

now we plot bar graph with the help of ggplot for team by number of match played by them. ggplot comes under ggplot2 package which is comes with tidyverse package . ggplot is used for data visualisation

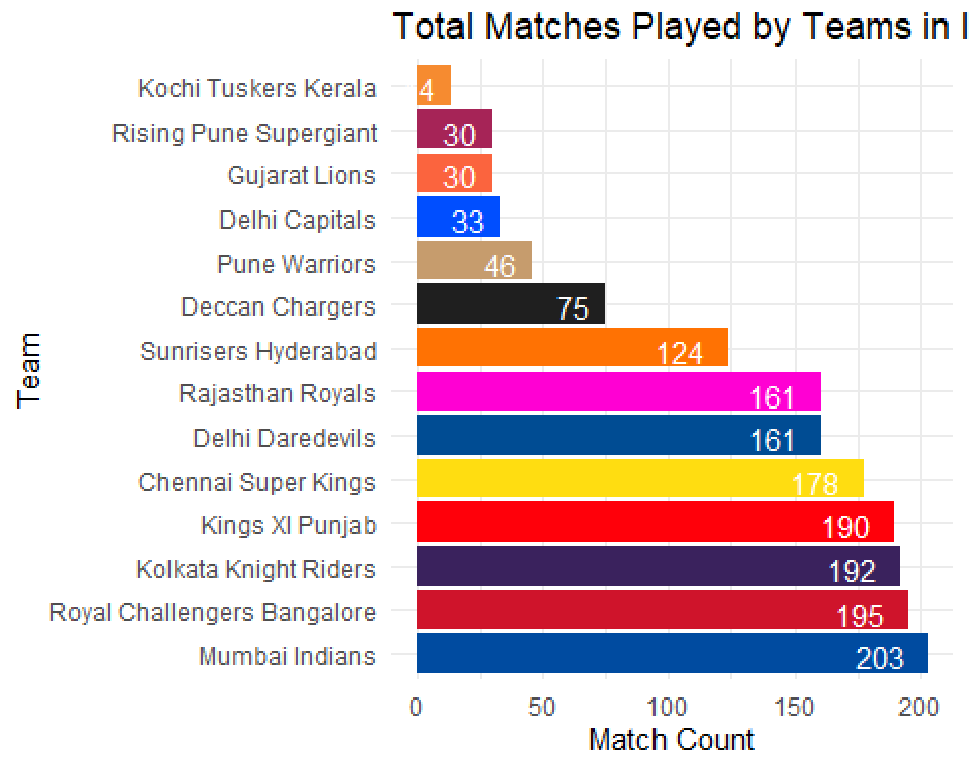
```
team_colors <- c("Mumbai Indians" = "#004ba0",
                 "Royal Challengers Bangalore" = "#CF142B",
                 "Kolkata Knight Riders" = "#3a225d",
                 "Kings XI Punjab" = "#ff000a",
                 "Chennai Super Kings" = "#FFDD11",
                 "Delhi Daredevils" = "#004c93",
                 "Rajasthan Royals" = "#ff00d3",
                 "Sunrisers Hyderabad" = "#ff7203",
                 "Deccan Chargers" = "#1f1f1f",
                 "Pune Warriors" = "#c69c6d",
                 "Delhi Capitals" = "#004eff",
                 "Gujarat Lions" = "#fb643e",
                 "Rising Pune Supergiant" = "#a62457",
                 "Kochi Tuskers Kerala" = "#f68b30")
```

first we create a variable contain hex color code for each team

we sort bar in descending order of match played and for doing this we use “reoder()”

The reorder() function takes two arguments: the variable to be reordered (winner), and the variable to use for ordering (-n in this case, which orders the bars in descending order of the n variable).

~~~

```r
# Create a bar plot with team names on y-axis and match count on x-axis
ggplot(total_match_sorted, aes(x = match, y = reorder(Team, -match))) +
  geom_bar(stat = "identity", fill = team_colors,)+
  scale_fill_manual(values = team_colors) +
  labs(title = "Total Matches Played by Teams in IPL (2008-2020)",
       x = "Match Count", y = "Team",)+geom_text(aes(label = match), hjust =
1.5, vjust = 0.6, , color = "white") +
  theme_minimal()
~~~



So we can cleary see that mumbai indians have played te maximum match

## Winning

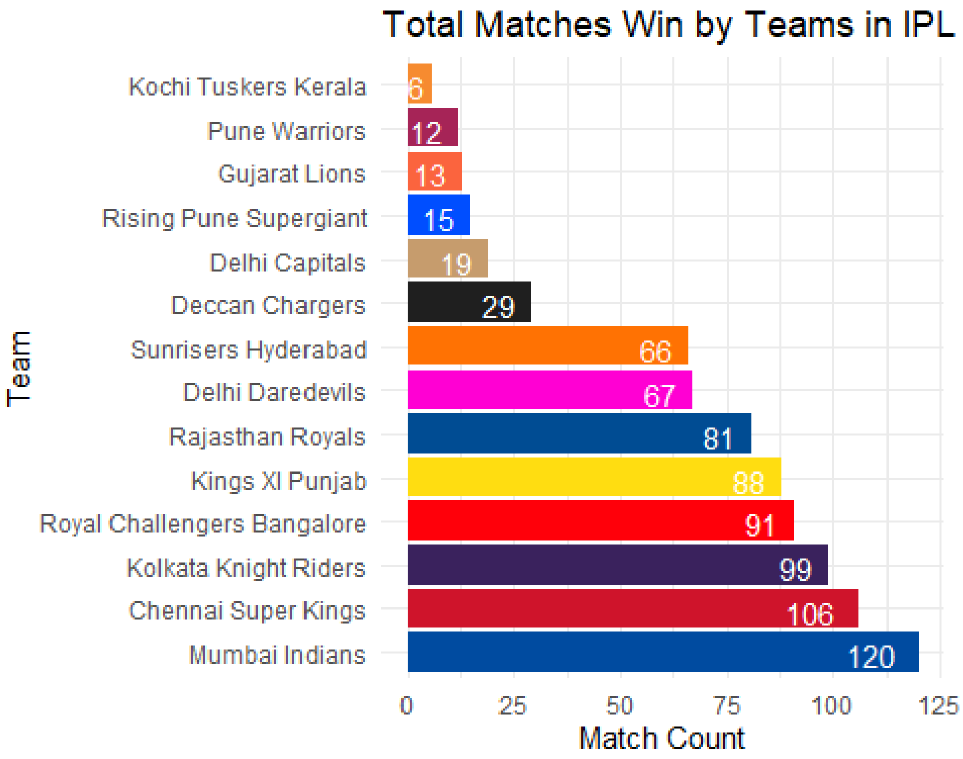
Now we see which team has won how many matches

```
#total match win by a team
match_wins <- count(tbl,winner)

#sorting data in descending order of match wins
match_wins_sorted <- arrange(match_wins,desc(n))

# Remove rows with missing values
match_wins_sorted <- na.omit(match_wins_sorted)
print(match_wins_sorted)
```

```
## # A tibble: 14 × 2
##    winner                          n
##    <chr>                       <int>
##  1 Mumbai Indians                120
##  2 Chennai Super Kings           106
##  3 Kolkata Knight Riders          99
##  4 Royal Challengers Bangalore    91
##  5 Kings XI Punjab                88
##  6 Rajasthan Royals               81
##  7 Delhi Daredevils               67
##  8 Sunrisers Hyderabad            66
##  9 Deccan Chargers                29
## 10 Delhi Capitals                19
## 11 Rising Pune Supergiant        15
## 12 Gujarat Lions                 13
## 13 Pune Warriors                 12
## 14 Kochi Tuskers Kerala           6
```

Now plot graph for it

```
# Create plot for how many match win by a team
ggplot(match_wins_sorted, aes(x = n, y =reorder( winner,-n))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Total Matches Win by Teams in IPL (2008-2020)",
       x = "Match Count", y = "Team") +
  geom_text(aes(label = n), hjust = 1.5, vjust = 0.6, color = "white") +
  theme_minimal()
```



winning percentage

Now we calculate wiining percentage of team by simply dividig won match to total match and then multiply them by 100

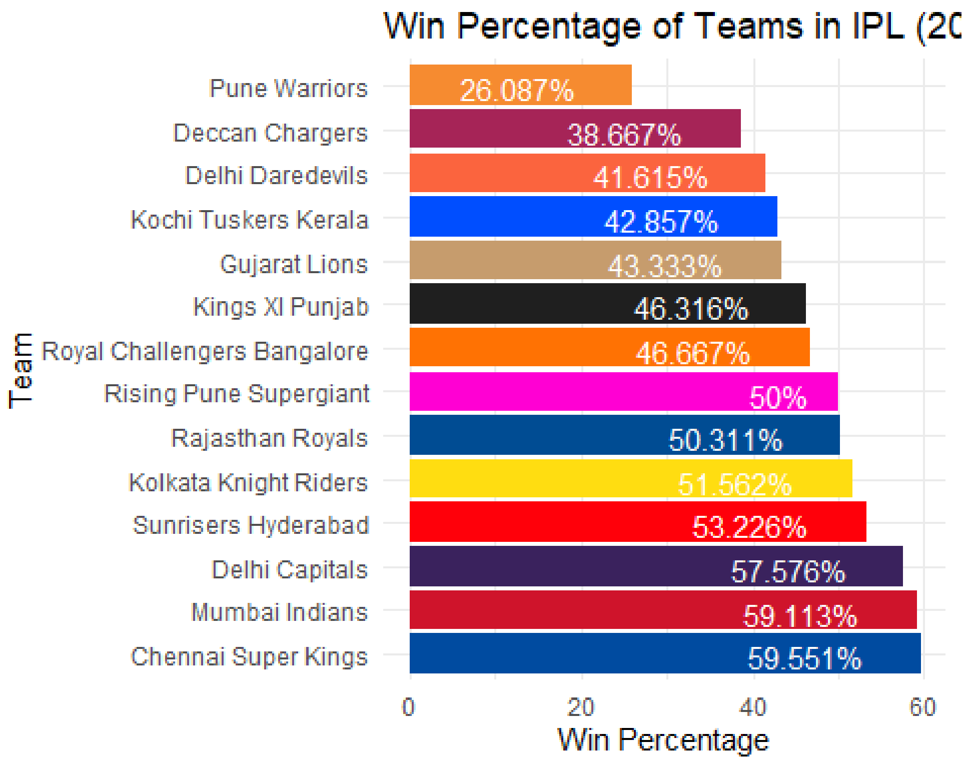
```
# win percentage of every team
win_percentage <-( match_wins$n/total_match$match)*100

## Warning in match_wins$n/total_match$match: longer object length is not a
## multiple of shorter object length
```

```
win_percentage <- round(win_percentage,3)

# creating win percentage data frame
win_percentage_df <- data.frame(`Team`=match_wins$winner,`Win
Percentage`=round(win_percentage,3))

# sorting in decending order
win_percentage_df_sorted <-
arrange(win_percentage_df,desc(win_percentage_df$Win.Percentage))
win_percentage_df_sorted <- na.omit(win_percentage_df_sorted)
print(win_percentage_df_sorted)

##                           Team Win.Percentage
## 1          Chennai Super Kings         59.551
## 2               Mumbai Indians         59.113
## 3               Delhi Capitals         57.576
## 4          Sunrisers Hyderabad         53.226
## 5        Kolkata Knight Riders         51.562
## 6             Rajasthan Royals         50.311
## 7       Rising Pune Supergiant         50.000
## 8  Royal Challengers Bangalore         46.667
## 9              Kings XI Punjab         46.316
## 10               Gujarat Lions         43.333
## 11        Kochi Tuskers Kerala         42.857
## 12            Delhi Daredevils         41.615
## 13             Deccan Chargers         38.667
## 14               Pune Warriors         26.087
```

ploting graph

Now we draw a bar chart for it

```
ggplot(win_percentage_df_sorted, aes(x = Win.Percentage , y =reorder( Team,-
Win.Percentage))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Win Percentage of Teams in IPL (2008-2020)",
       x = "Win Percentage", y = "Team") +
  geom_text(aes(label = paste0(c(Win.Percentage),"%")), hjust = 1.5, vjust =
0.6, color = "white") +
  theme_minimal()
```



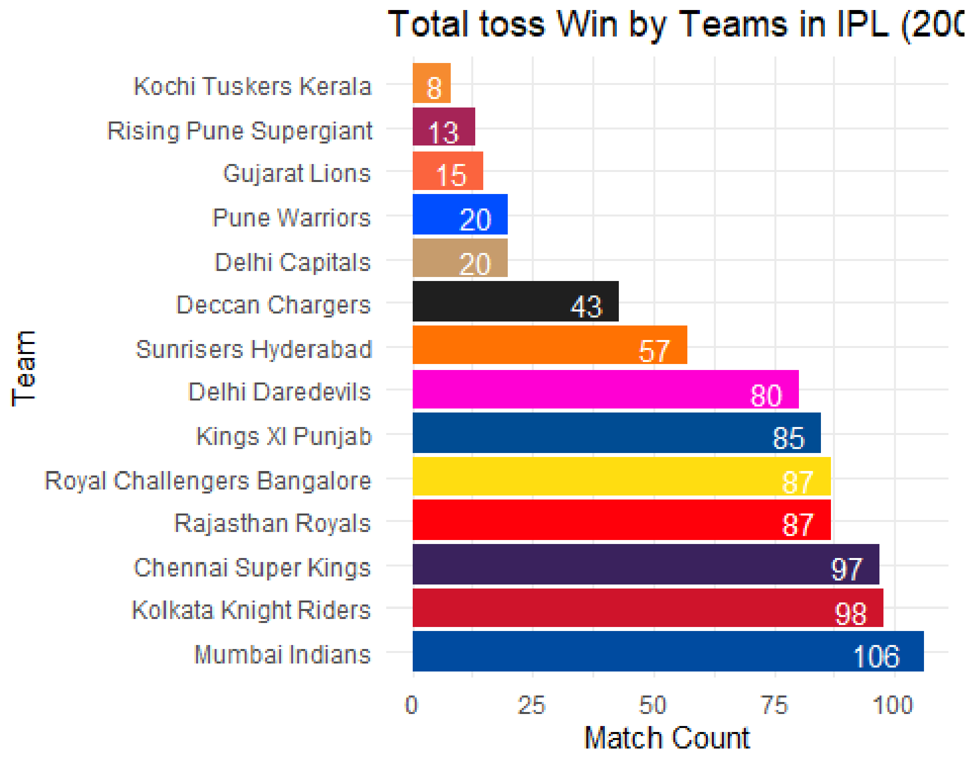
Toss Win

Now we find which team won toss most of time

```
# Count the number of times each team won the toss
toss <- count(tbl, toss_winner)

# Rename the count column to "Toss Wins" and
toss <- rename(toss, Toss_Wins = n,)

#sorteng toss data in descending order
toss_sorted <- arrange(toss,desc(Toss_Wins))
print(toss_sorted)

## # A tibble: 14 × 2
##    toss_winner                 Toss_Wins
##    <chr>                           <int>
##  1 Mumbai Indians                    106
##  2 Kolkata Knight Riders              98
##  3 Chennai Super Kings                97
##  4 Rajasthan Royals                   87
##  5 Royal Challengers Bangalore        87
##  6 Kings XI Punjab                    85
##  7 Delhi Daredevils                   80
##  8 Sunrisers Hyderabad                57
##  9 Deccan Chargers                    43
## 10 Delhi Capitals                     20
## 11 Pune Warriors                      20
## 12 Gujarat Lions                       15
## 13 Rising Pune Supergiant              13
## 14 Kochi Tuskers Kerala                 8
```

Ploting graph

```
ggplot(toss_sorted, aes(x = Toss_Wins, y =reorder( toss_winner,-Toss_Wins)))
+
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Total toss Win by Teams in IPL (2008-2020)",
       x = "Match Count", y = "Team") +
  geom_text(aes(label = Toss_Wins), hjust = 1.5, vjust = 0.6, color =
"white") +
  theme_minimal()
```



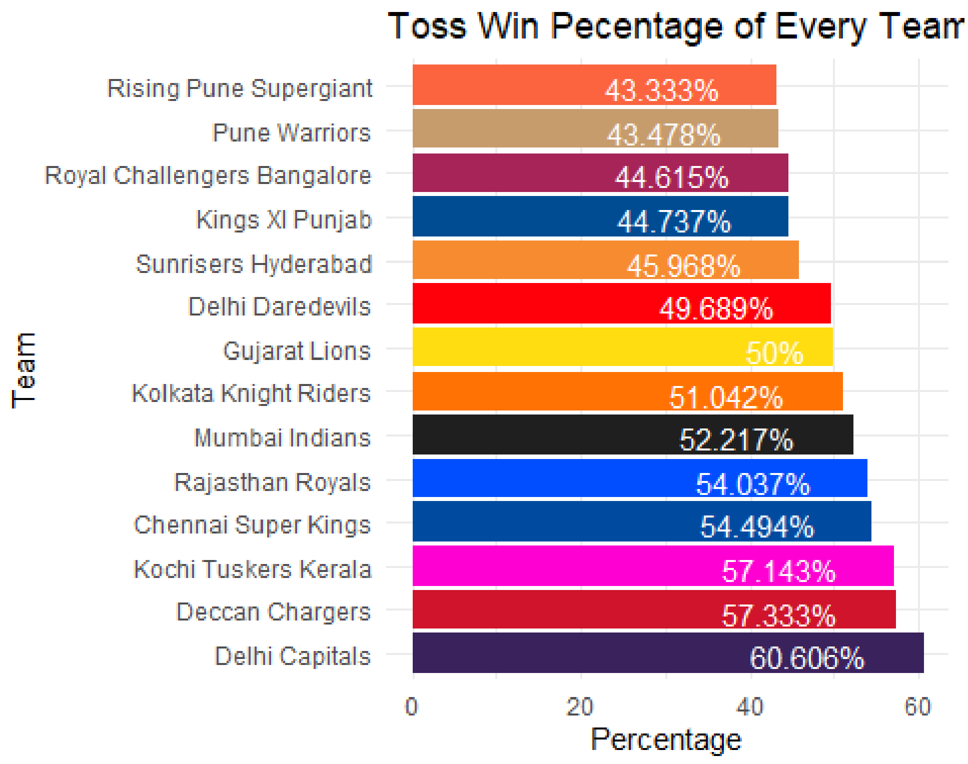
## Toss Win Percentage

```
# calculate toss wins percentage
toss_percentage <- (toss$Toss_Wins/total_match$match)*100
toss_percentage <- round(toss_percentage,3)

#creating data frame for toss percentage
toss_percentage_df <- data.frame(`Team`=total_match$Team,toss_percentage)

#sorting data
toss_percentage_sorted <- arrange(toss_percentage_df,desc(toss_percentage))
print(toss_percentage_sorted)
```

```
##                             Team toss_percentage
## 1                Delhi Capitals          60.606
## 2               Deccan Chargers          57.333
## 3           Kochi Tuskers Kerala          57.143
## 4            Chennai Super Kings          54.494
## 5              Rajasthan Royals          54.037
## 6                Mumbai Indians          52.217
## 7          Kolkata Knight Riders          51.042
## 8                 Gujarat Lions          50.000
## 9              Delhi Daredevils          49.689
## 10          Sunrisers Hyderabad          45.968
## 11              Kings XI Punjab          44.737
## 12 Royal Challengers Bangalore          44.615
## 13               Pune Warriors          43.478
## 14       Rising Pune Supergiant          43.333
```

##Ploting Graph

```
# toss win percentage
ggplot(toss_percentage_df, aes(x = toss_percentage, y =reorder( Team,-
toss_percentage))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Toss Win Pecentage of Every Teams in IPL (2008-2020)",
       x = "Percentage", y = "Team") +
  geom_text(aes(label = paste0(c(toss_percentage),"%")), hjust = 1.5, vjust =
0.6, color = "white") +
  theme_minimal()
```



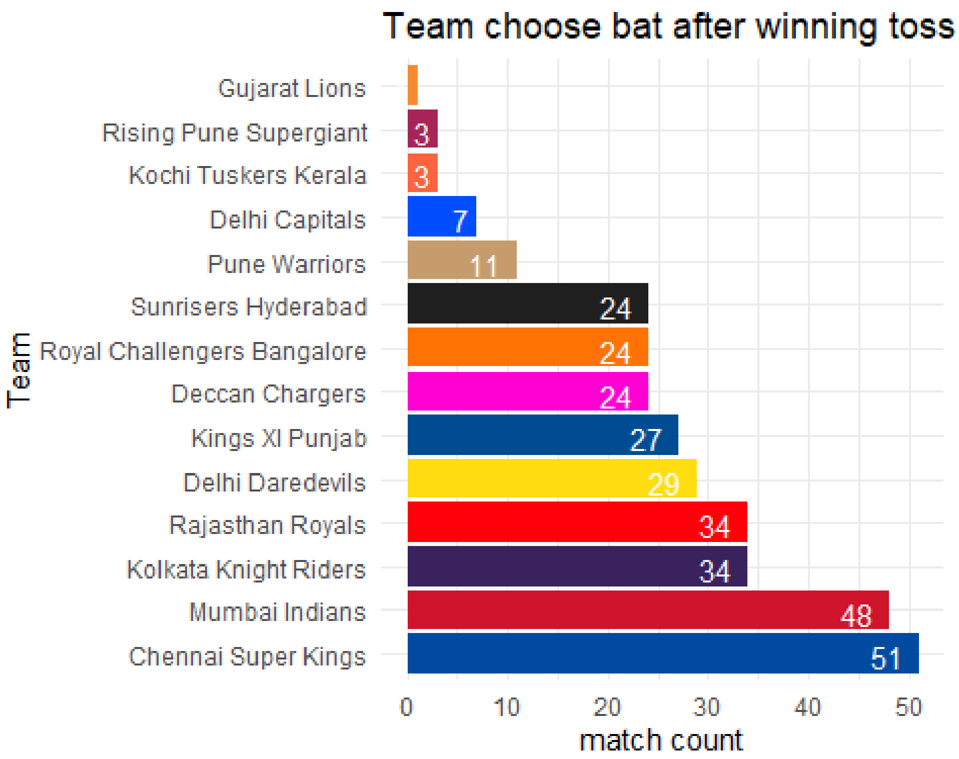
## Toss Decision

Now we see how many time team choose bat after winning toss

```
# finding team that choose batting after winning toss
toss_bat <- tbl%>% filter(toss_decision =="bat")


#group them
tb<- group_by(toss_bat,toss_winner)

# counting team choose bat no. of times after winning toss
bat_first<- count(tb)

#sorting in descending order
bat_first_sorted<- arrange(bat_first,desc(n))
print(bat_first_sorted)

## # A tibble: 14 × 2
## # Groups:   toss_winner [14]
##    toss_winner                     n
##    <chr>                       <int>
##  1 Chennai Super Kings            51
##  2 Mumbai Indians                 48
##  3 Kolkata Knight Riders          34
##  4 Rajasthan Royals               34
##  5 Delhi Daredevils               29
##  6 Kings XI Punjab                27
##  7 Deccan Chargers                24
##  8 Royal Challengers Bangalore    24
##  9 Sunrisers Hyderabad          24
## 10 Pune Warriors                11
## 11 Delhi Capitals                7
## 12 Kochi Tuskers Kerala          3
## 13 Rising Pune Supergiant        3
## 14 Gujarat Lions                 1
```

#plot

```
ggplot(bat_first_sorted, aes(x = n, y =reorder( toss_winner,-n))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team choose bat after winning toss in IPL (2008-2020)",
       x = "match count", y = "Team") +
  geom_text(aes(label = n), hjust = 1.5, vjust = 0.6, color = "white") +
  theme_minimal()
```

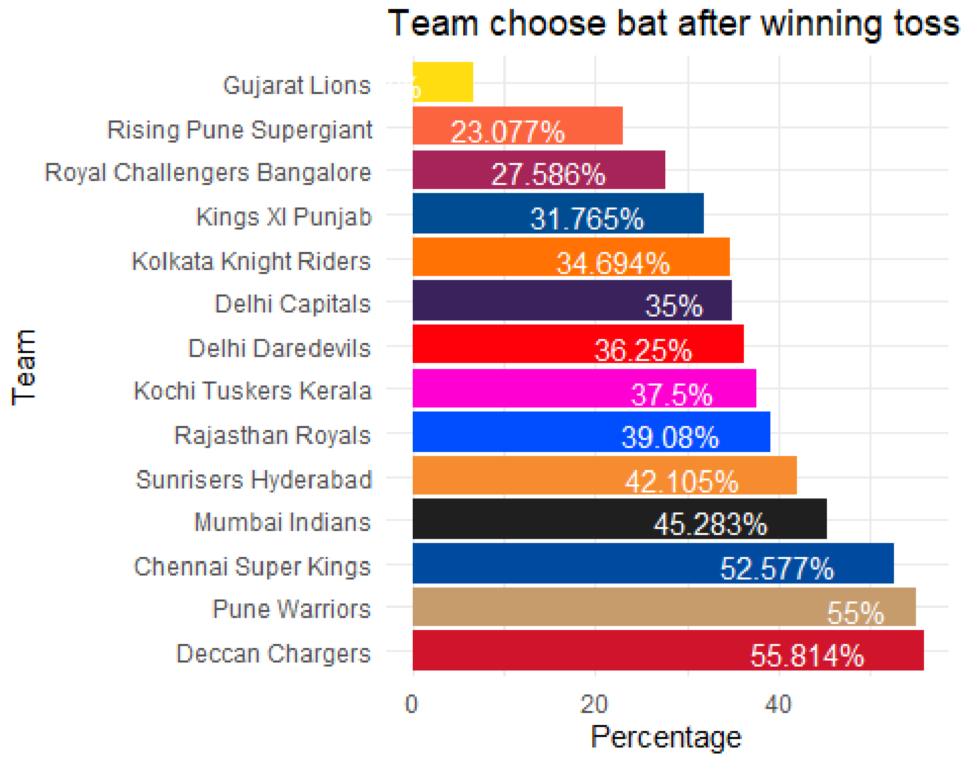


now we check percentage of time team choose bat afte winning toss

```
#bat first percentage
toss_bat_percentage <-
data.frame(Team=bat_first$toss_winner,bat_percentage=(round(bat_first$n/toss$
Toss_Wins*100,3)))
print(toss_bat_percentage)

##                           Team bat_percentage
## 1          Chennai Super Kings         52.577
## 2              Deccan Chargers         55.814
## 3               Delhi Capitals         35.000
## 4              Delhi Daredevils         36.250
## 5                 Gujarat Lions          6.667
## 6               Kings XI Punjab         31.765
## 7          Kochi Tuskers Kerala         37.500
## 8         Kolkata Knight Riders         34.694
## 9                Mumbai Indians         45.283
## 10                Pune Warriors         55.000
## 11             Rajasthan Royals         39.080
## 12       Rising Pune Supergiant         23.077
## 13 Royal Challengers Bangalore          27.586
## 14          Sunrisers Hyderabad         42.105
```

graph

```
ggplot(toss_bat_percentage, aes(x = bat_percentage, y =reorder( Team,-bat_percentage))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team choose bat after winning toss in IPL (2008-2020)",
       x = "Percentage", y = "Team") +
  geom_text(aes(label = paste0(c(bat_percentage),"%")), hjust = 1.5, vjust = 0.6, color = "white") +
  theme_minimal()
```



Now we see how many time team choose fielding after winning toss

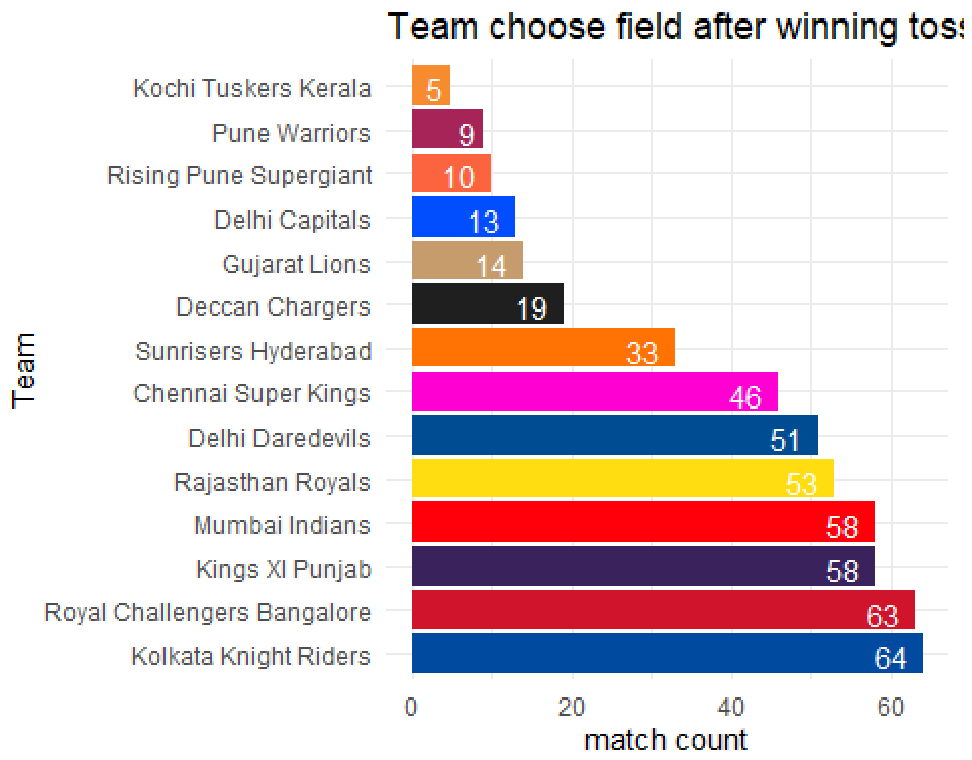
```
# finding team that choose fielding after winning toss
toss_field <- tbl%>% filter(toss_decision =="field")

#group them
```

```
tf<- group_by(toss_field,toss_winner)

# counting team choose fielding no. of times after winning toss
field_first<- count(tf)

#sorting in descending order
field_first_sorted<- arrange(field_first,desc(n))
print(field_first_sorted)
```

```
## # A tibble: 14 × 2
## # Groups:   toss_winner [14]
##    toss_winner                     n
##    <chr>                       <int>
##  1 Kolkata Knight Riders          64
##  2 Royal Challengers Bangalore    63
##  3 Kings XI Punjab                58
##  4 Mumbai Indians                 58
##  5 Rajasthan Royals               53
##  6 Delhi Daredevils               51
##  7 Chennai Super Kings            46
##  8 Sunrisers Hyderabad            33
##  9 Deccan Chargers                19
## 10 Gujarat Lions                  14
## 11 Delhi Capitals                 13
## 12 Rising Pune Supergiant         10
## 13 Pune Warriors                   9
## 14 Kochi Tuskers Kerala            5
```

Graph

```
# toss winner choose fielding
ggplot(field_first_sorted, aes(x = n, y =reorder( toss_winner,-n))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team choose field after winning toss in IPL (2008-2020)",
       x = "match count", y = "Team") +
  geom_text(aes(label = n), hjust = 1.5, vjust = 0.6, color = "white") +
  theme_minimal()
```



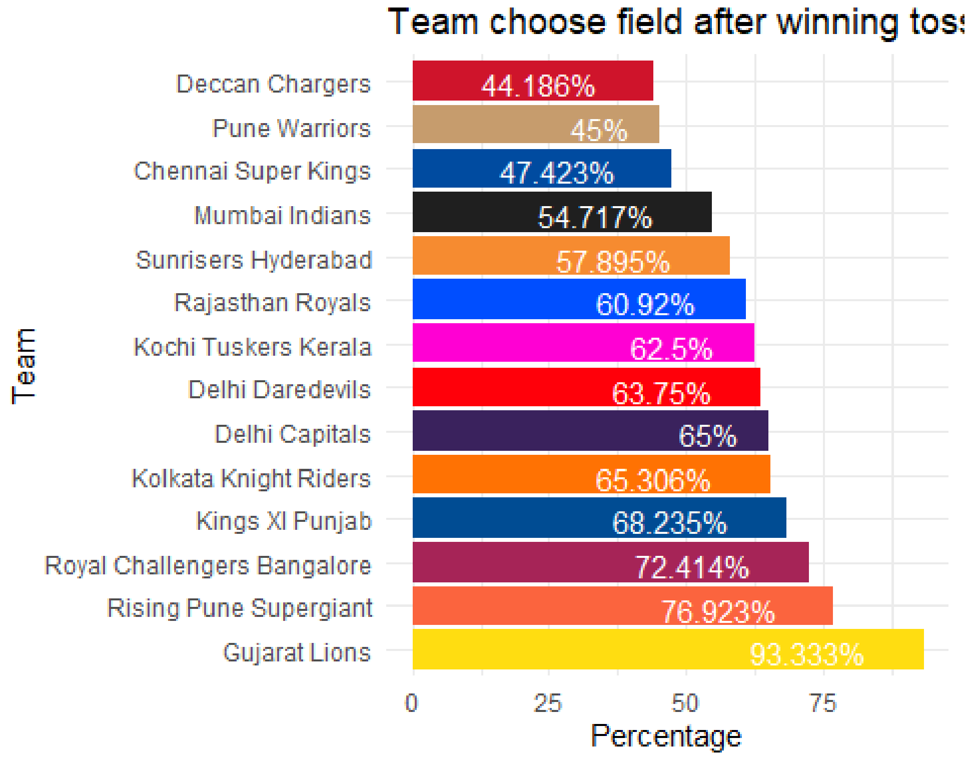
now we check percentage of time team choose field afte winning toss

```
#field  first percentage
toss_field_percentage <-
data.frame(Team=field_first$toss_winner,field_percentage=(round(field_first$n
/toss$Toss_Wins*100,3)))
print(toss_field_percentage)
```

```
##                           Team field_percentage
## 1          Chennai Super Kings           47.423
## 2              Deccan Chargers           44.186
## 3              Delhi Capitals           65.000
## 4             Delhi Daredevils           63.750
## 5                Gujarat Lions           93.333
## 6              Kings XI Punjab           68.235
## 7         Kochi Tuskers Kerala           62.500
## 8        Kolkata Knight Riders           65.306
## 9               Mumbai Indians           54.717
## 10               Pune Warriors           45.000
## 11            Rajasthan Royals           60.920
## 12      Rising Pune Supergiant           76.923
## 13 Royal Challengers Bangalore           72.414
## 14         Sunrisers Hyderabad           57.895
```

graph

```
# Toss winner choose fielding in percent
ggplot(toss_field_percentage, aes(x = field_percentage, y =reorder( Team,-
field_percentage))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team choose field after winning toss in IPL (2008-2020)",
       x = "Percentage", y = "Team") +
  geom_text(aes(label = paste0(c(field_percentage),"%")), hjust = 1.5, vjust
= 0.6, color = "white") +
  theme_minimal()
```

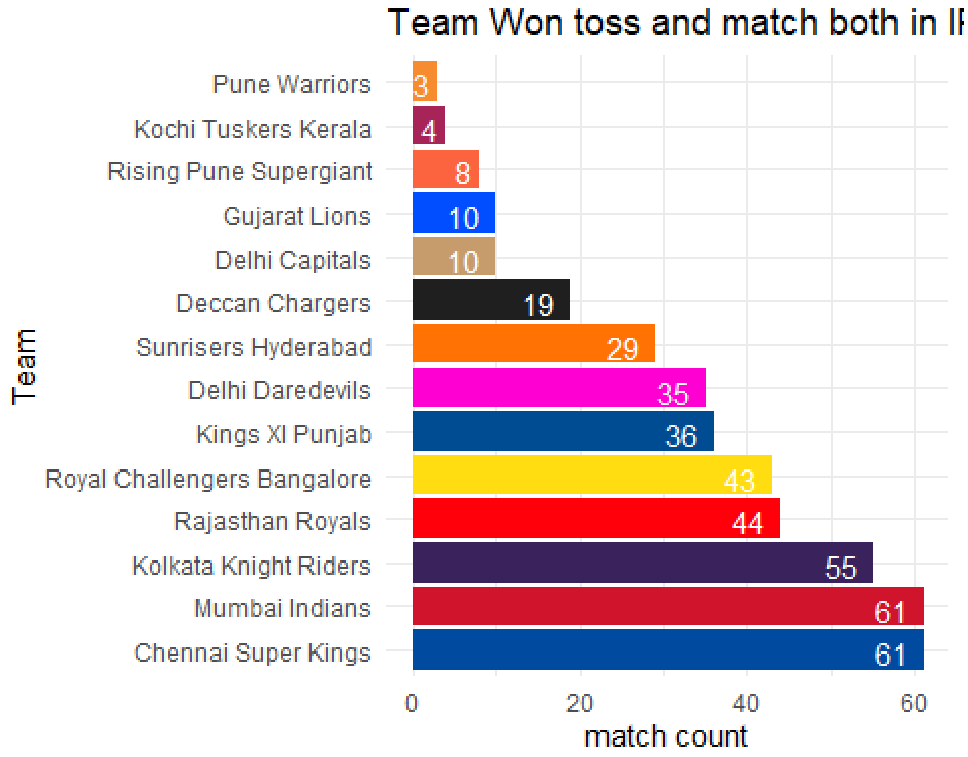


## Team toss and win NOw we check which team won the match after winning toss

```
#finding team that win match and toss both
t_m_w <- tbl %>%filter(toss_winner == winner)

#grouping them
toss_and_match_winner <-group_by(t_m_w,winner)

#counting no of time team win toss and match both
toss_and_match_winner <- count(toss_and_match_winner)

#sorting in descending order
toss_and_match_winner_sorted=arrange(toss_and_match_winner,desc(n))
print(toss_and_match_winner_sorted)

## # A tibble: 14 × 2
## # Groups:   winner [14]
##    winner                          n
##    <chr>                       <int>
##  1 Chennai Super Kings            61
##  2 Mumbai Indians                 61
##  3 Kolkata Knight Riders          55
##  4 Rajasthan Royals               44
##  5 Royal Challengers Bangalore    43
##  6 Kings XI Punjab                36
##  7 Delhi Daredevils               35
##  8 Sunrisers Hyderabad            29
##  9 Deccan Chargers                19
## 10 Delhi Capitals                 10
## 11 Gujarat Lions                  10
## 12 Rising Pune Supergiant          8
## 13 Kochi Tuskers Kerala            4
## 14 Pune Warriors                   3
```

Graph

```
# toss and match both win by team
ggplot(toss_and_match_winner_sorted, aes(x = n, y =reorder( winner,-n))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team Won toss and match both in IPL (2008-2020)",
       x = "match count", y = "Team") +
  geom_text(aes(label = n), hjust = 1.5, vjust = 0.6, color = "white") +
  theme_minimal()
```



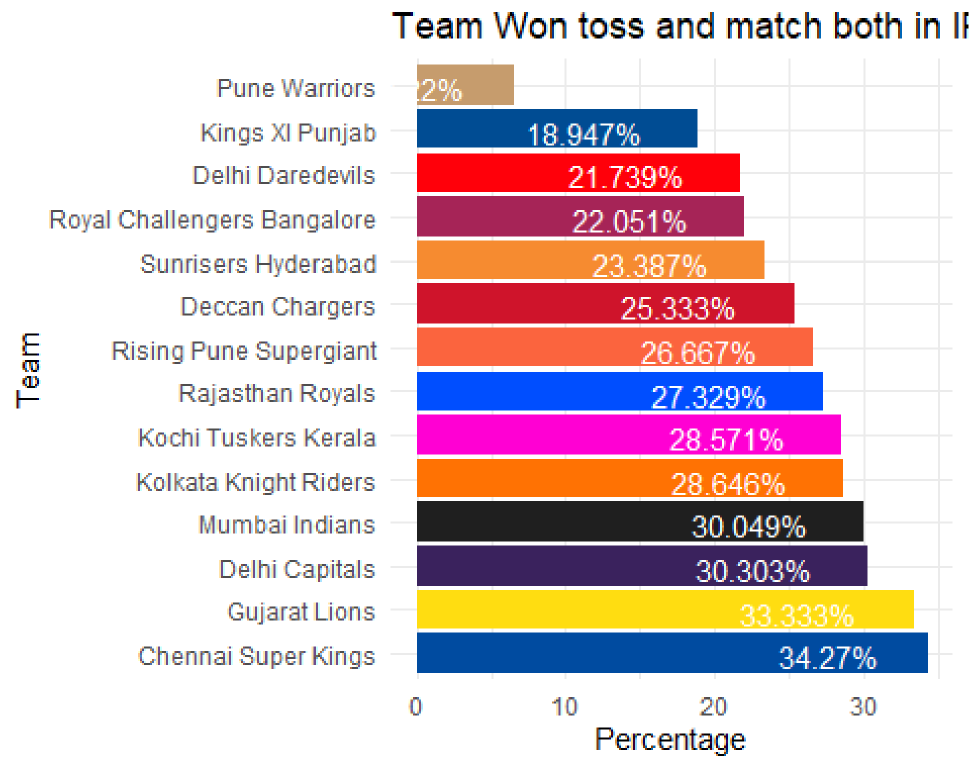
now we check percentage of time team won toss and match both

```
#check percentage how many time team won match and toss both
t_a_m_percentage <-
data.frame(Team=toss_and_match_winner$winner,Tm_percentage=(round(toss_and_ma
tch_winner$n/total_match$match*100,3)))
print(t_a_m_percentage)
```

```
##                                Team Tm_percentage
## 1           Chennai Super Kings        34.270
## 2               Deccan Chargers        25.333
## 3                Delhi Capitals        30.303
## 4              Delhi Daredevils        21.739
## 5                 Gujarat Lions        33.333
## 6               Kings XI Punjab        18.947
## 7          Kochi Tuskers Kerala        28.571
## 8         Kolkata Knight Riders        28.646
## 9                Mumbai Indians        30.049
## 10                Pune Warriors         6.522
## 11             Rajasthan Royals        27.329
## 12       Rising Pune Supergiant        26.667
## 13 Royal Challengers Bangalore        22.051
## 14          Sunrisers Hyderabad        23.387
```

Graph

```
# toss and match both win by team percentage
ggplot(t_a_m_percentage, aes(x = Tm_percentage, y =reorder( Team,-
Tm_percentage))) +
  geom_bar(stat = "identity", fill = team_colors) +
  scale_fill_manual(values = team_colors) +
  labs(title = "Team Won toss and match both in IPL (2008-2020)",
       x = "Percentage", y = "Team") +
  geom_text(aes(label = paste0(c(Tm_percentage),"%")), hjust = 1.5, vjust =
0.6, color = "white") +
  theme_minimal()
```



## Player of the match Now we check which player have won player of the maytch award for maximum number of time

```
#count the no. of time potm award won by player
most_potm <- count(tbl, player_of_match)

# sorting the most potm award in descending order
most_potm_sorted <- arrange(most_potm, desc(n))
print(most_potm_sorted)

## # A tibble: 234 × 2
##    player_of_match     n
##    <chr>           <int>
##  1 AB de Villiers     23
##  2 CH Gayle           22
##  3 RG Sharma          18
##  4 DA Warner          17
##  5 MS Dhoni           17
##  6 SR Watson          16
##  7 YK Pathan          16
##  8 SK Raina           14
##  9 G Gambhir          13
## 10 V Kohli            13
## # … with 224 more rows
```

Graph

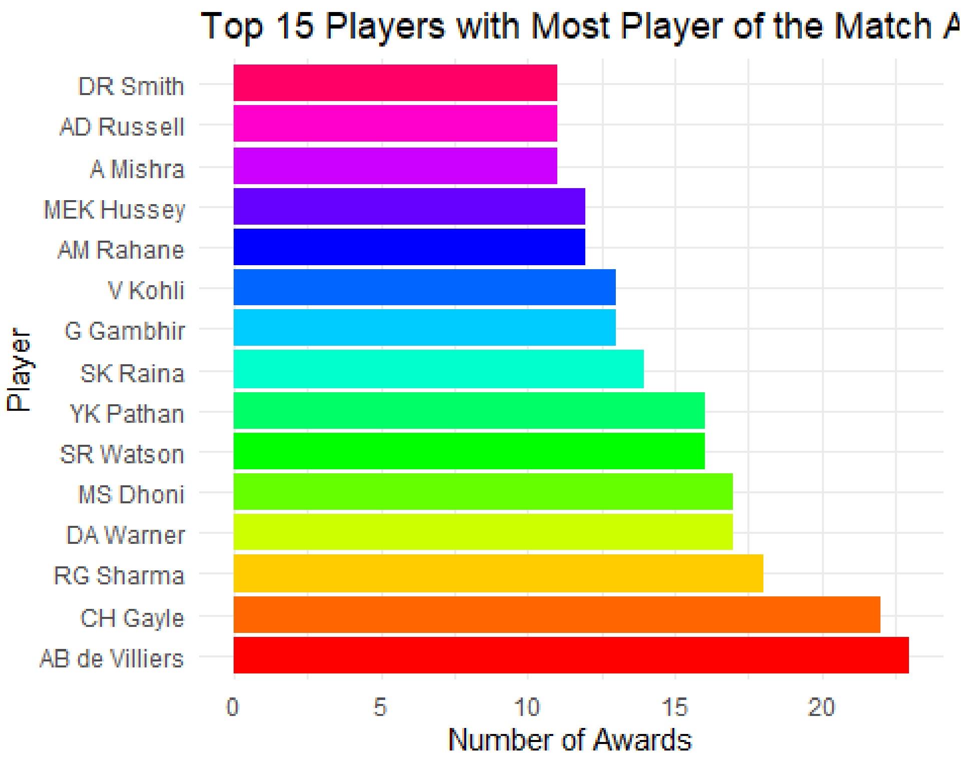
we plot graph for top 15 player for that first we do

```
#selectenging top 15 player only
potm <- head(most_potm_sorted,15)
print(potm)
```

```
## # A tibble: 15 × 2
##    player_of_match     n
##    <chr>           <int>
##  1 AB de Villiers     23
##  2 CH Gayle           22
##  3 RG Sharma          18
##  4 DA Warner          17
##  5 MS Dhoni           17
##  6 SR Watson          16
##  7 YK Pathan          16
##  8 SK Raina           14
##  9 G Gambhir          13
## 10 V Kohli            13
## 11 AM Rahane          12
## 12 MEK Hussey         12
## 13 A Mishra           11
## 14 AD Russell         11
## 15 DR Smith           11
```

now we plot graph for these 15 player

```
# Create a color palette for the players
player_colors <- rainbow(nrow(potm))
#potm
ggplot(potm, aes(x = reorder(`player_of_match`, -n), y = n)) +
  geom_col(fill = player_colors) +
  coord_flip() +
  theme_minimal() +
  labs(title = "Top 15 Players with Most Player of the Match Awards",
       x = "Player", y = "Number of Awards")
```


Top 15 Players with Most Player of the Match A
Player
DR Smith
AD Russell
A Mishra
MEK Hussey
AM Rahane
V Kohli
G Gambhir
SK Raina
YK Pathan
SR Watson
MS Dhoni
DA Warner
RG Sharma
CH Gayle
AB de Villiers
0
5
10
15
20
Number of Awards