DA Lab 2

Analysis of IPL-2019 Dataset

Name: Dheeraj Chaudhary

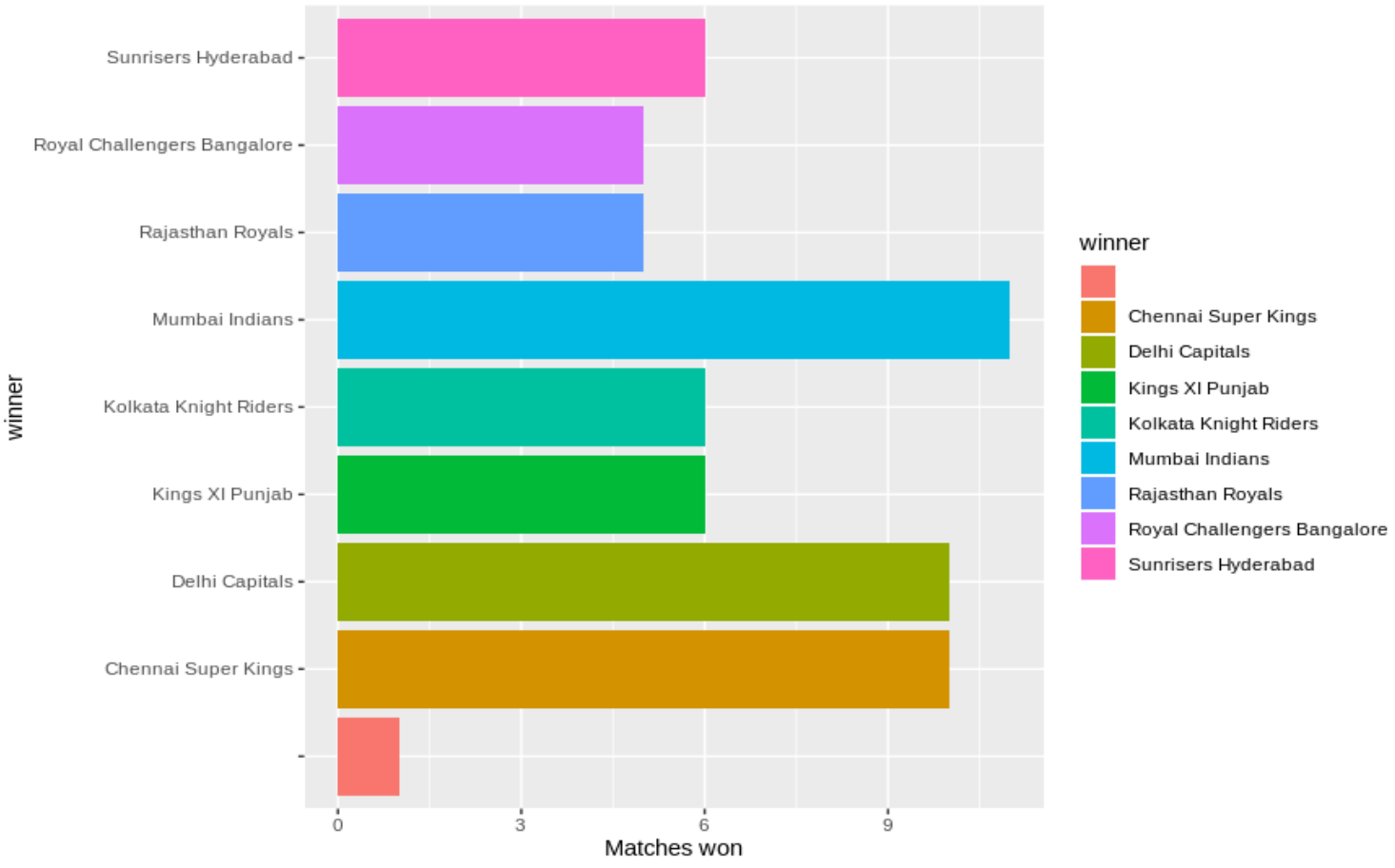
Roll: 17BCS009


```
library(tabulizer)
library(dplyr)
library(ggplot2)
library(reshape2)
library(magrittr)
library(tidyr)
```

```
#########################   READING MATCHES CSV FILE ######################
matches <- read.csv("/home/dheeraj/Desktop/Lecture/6th_sem_Academics/DataScience/Lab2/
matches.csv", stringsAsFactors = FALSE)
data <- read.csv("/home/dheeraj/Desktop/Lecture/6th_sem_Academics/DataScience/Lab2/
deliveries.csv", stringsAsFactors = FALSE)
matches <- matches[,-18]
data$wickets <- as.numeric(ifelse(data$player_dismissed =="" ,"",1))
```

```
######## Number of matches in the dataset   (We can see 60 matches were played in IPL'2019)
summarize(matches,no_of_matches = n())
##### OUTPUT >   no_of_matches  60
```

```
######### Which Team won by maximum runs?    (We can see SRH won y 118 runs)
max_run <- matches[which.max(matches$win_by_runs),]
select(max_run, winner, win_by_runs)
######  Output >            winner win_by_runs
                     11 Sunrisers Hyderabad         118

######### Which Team won by maximum wickets?   (We ca see SRH won by 9 wickets)
max_run <- matches[which.max(matches$win_by_wickets),]
select(max_run, winner, win_by_wickets)
###### Output > winner win_by_wickets
                     38 Sunrisers Hyderabad              9

############################ Teams and matches won  (We can see MI wo maximum matches)
matches%>%
  group_by(winner)%>%
  summarize(most_win = n())%>%
  ggplot(aes(x = winner,y = most_win,fill = winner))+
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_y_continuous("Matches won")
```



```r
teams <- data %>% select(batting_team)%>%
  distinct()
teams <- rename(teams, team = batting_team)
teams
```

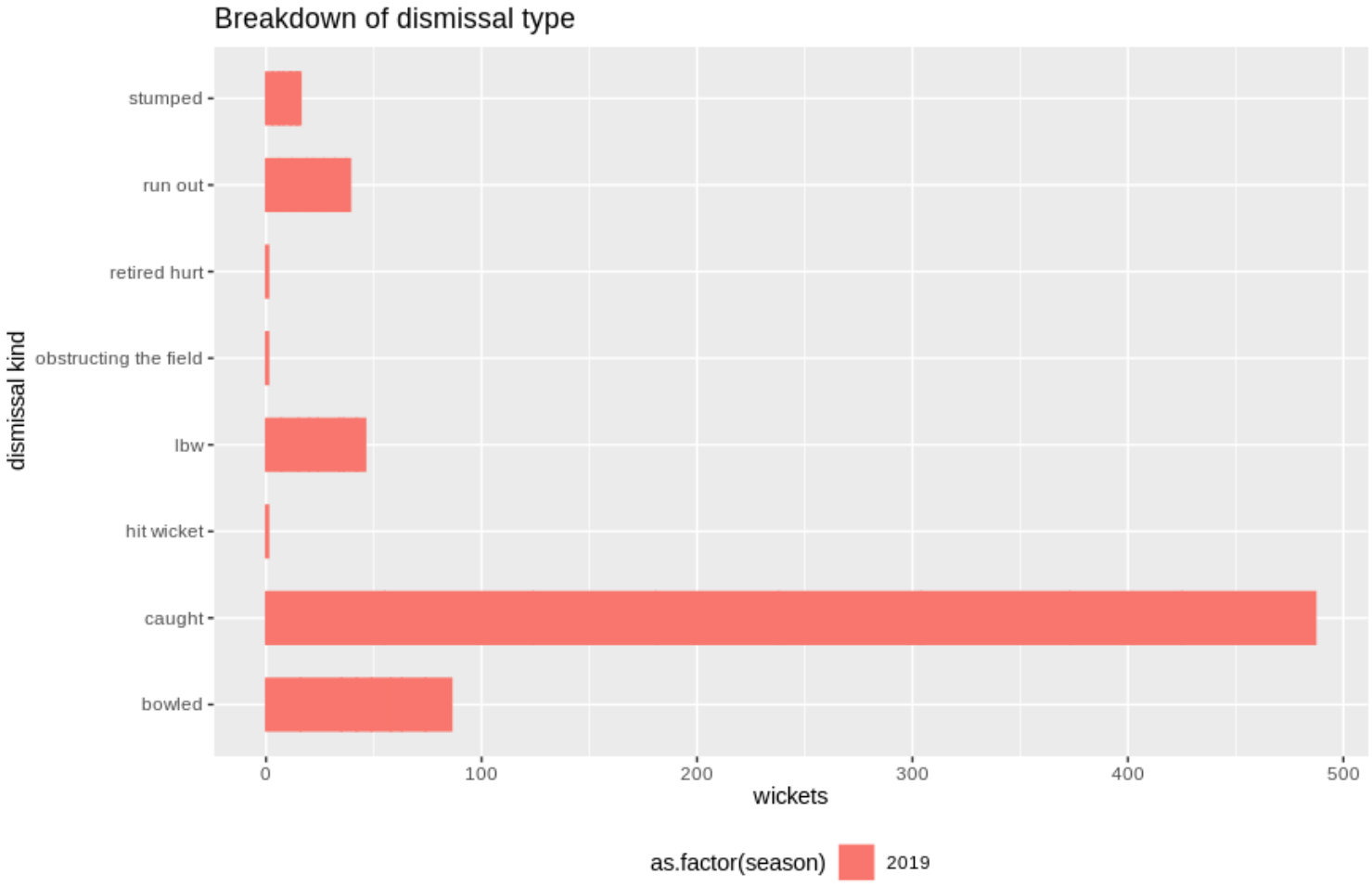
```
###### Output >               team          (following teams played in IPL 2019)

           1 Royal Challengers Bangalore
           2         Chennai Super Kings
           3         Sunrisers Hyderabad
           4       Kolkata Knight Riders
           5              Delhi Capitals
           6              Mumbai Indians
           7            Kings XI Punjab
           8           Rajasthan Royals
```

```r
s_team <- c("RCB","CSK","SRH","KKR","DC","MI","KXIP","RR")
s_team
```

```
#####  OUTPUT > [1] "RCB"  "CSK"  "SRH"  "KKR"  "DC"   "MI"   "KXIP" "RR"
```

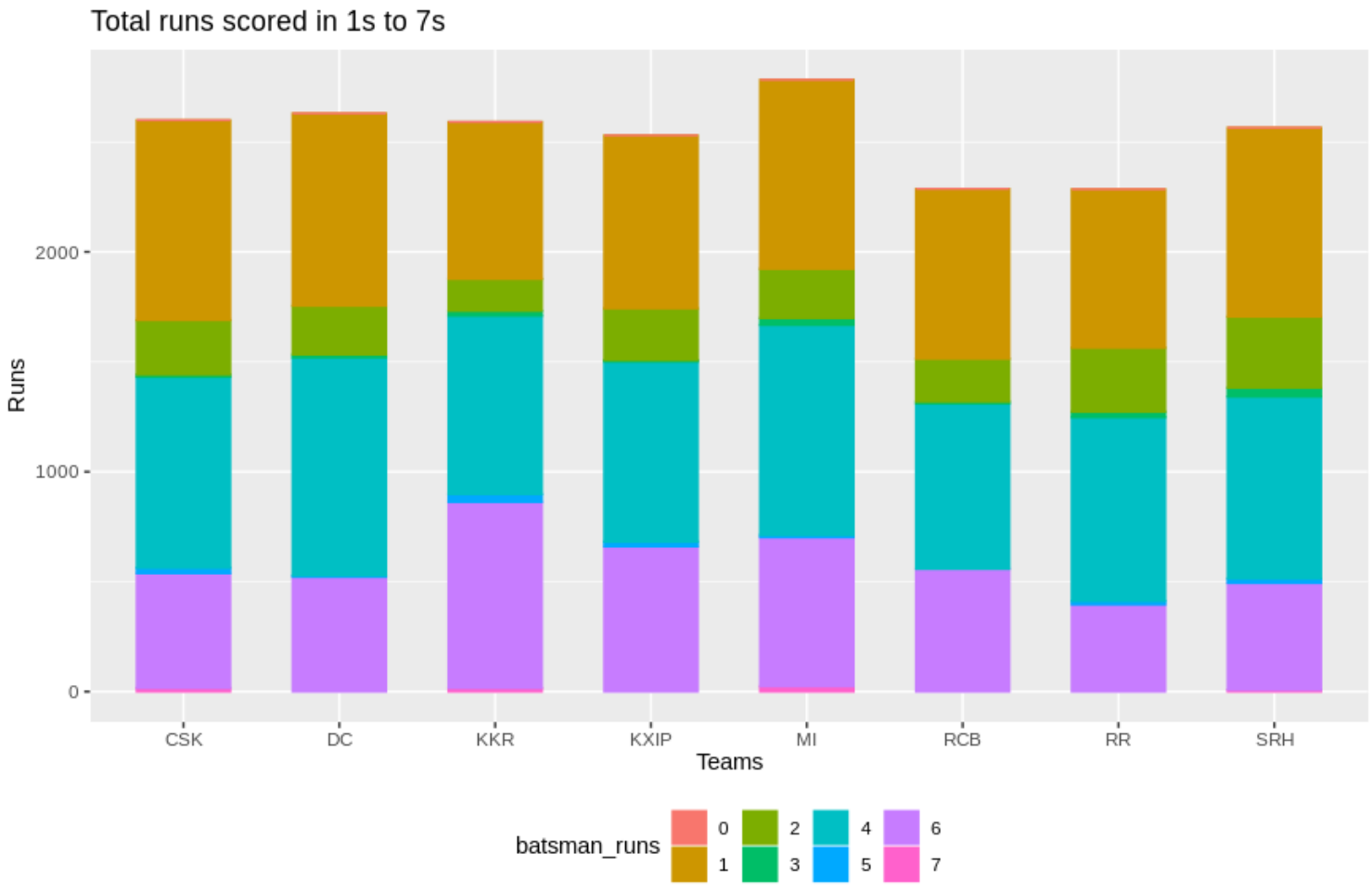
```r
teams <- cbind(teams, s_team)
player_of_match <- matches%>% select(id,player_of_match,season) %>%
  distinct()
player_of_match <- rename(player_of_match, player=player_of_match)
matches$city <- as.character(matches$city)
matches$city[matches$city==""] <- "Dubai"
venue_city <- matches %>%
  select(city)%>%
  distinct()
```

```
############### Dissmissal type and number of dismissal#############################
dismissal <- data%>%
  left_join(matches, by=c("match_id"="id"))%>%
  left_join(teams,by=c("batting_team"="team"))%>%
  filter(dismissal_kind!="")%>%
  group_by(season,dismissal_kind,s_team)%>%
  summarize(wickets =n())
ggplot(dismissal,aes(x=dismissal_kind,y=wickets,colour=as.factor(season),
fill=as.factor(season)))+
  geom_bar(position = "stack", show.legend = TRUE, width =.6,stat="identity")+
  theme(legend.position="bottom")+
  coord_flip()+
  theme(legend.direction = "horizontal") +
  scale_y_continuous(name="wickets")+
  scale_x_discrete(name="dismissal kind")+
  ggtitle("Breakdown of dismissal type ")
```



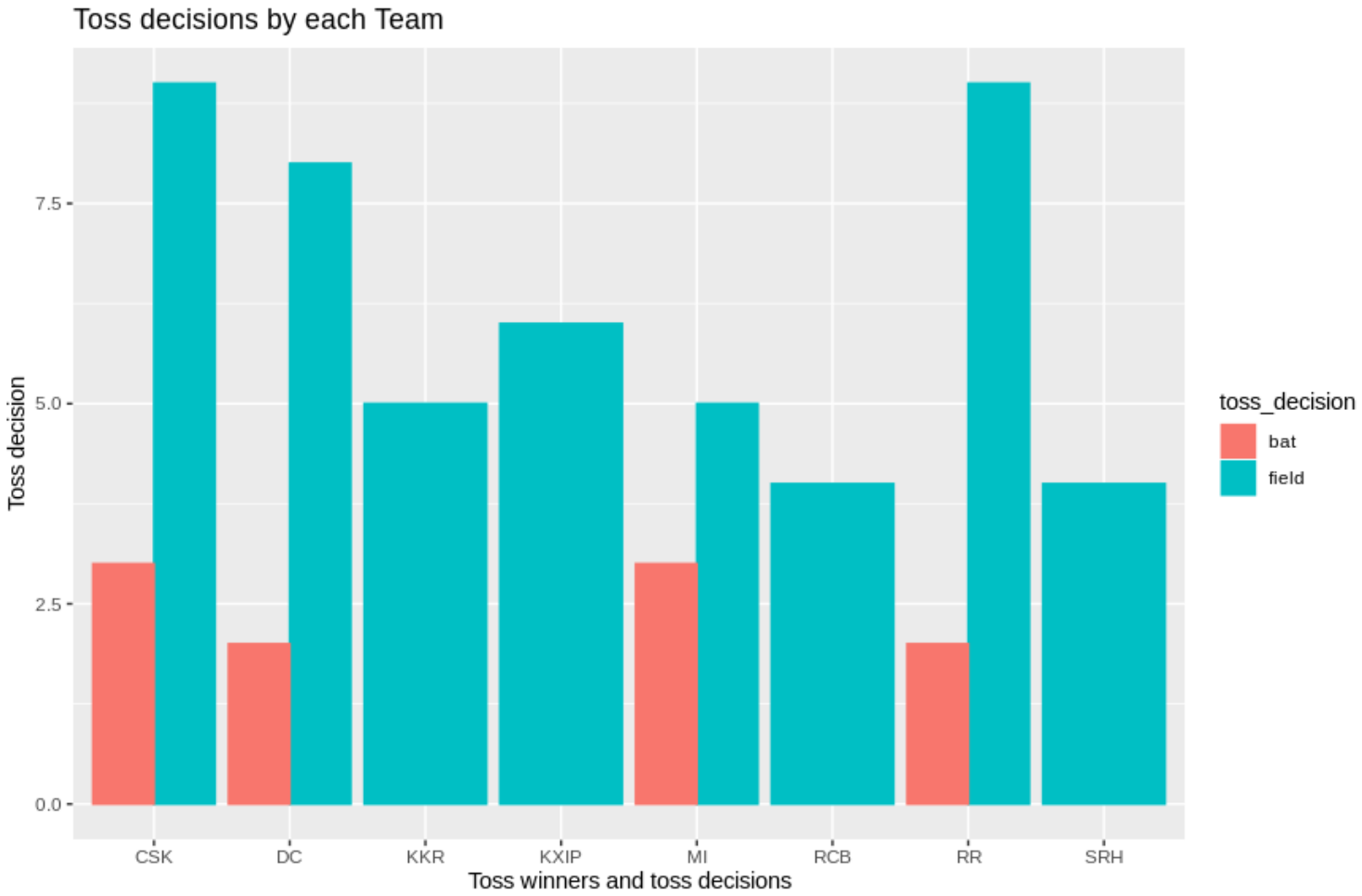
(We ca see in above plot that maximum dismissal was happened due to caught)

```
######################## Run scored in 1s to 7s  ####################
runs_cat <- data %>%
  left_join(matches,by=c("match_id"="id"))%>%
  left_join(teams,by=c("batting_team"="team"))%>%
  group_by(s_team,batsman_runs)%>%
  summarize(no=n(),runs=sum(total_runs))

runs_cat$batsman_runs <- as.factor(runs_cat$batsman_runs)

ggplot(runs_cat,aes(x=s_team,y=runs,colour=batsman_runs,fill=batsman_runs))+
  geom_bar(position = "stack", show.legend = TRUE, width =.6,stat="identity")+
  theme(legend.position="bottom")+
  theme(legend.direction = "horizontal") +
  scale_y_continuous(name="Runs")+
  scale_x_discrete(name="Teams")+
  ggtitle("Total runs scored in 1s to 7s")
```



(We ca see in above plot that most of the runs were scored in 1st, 3rd and 6th ball)

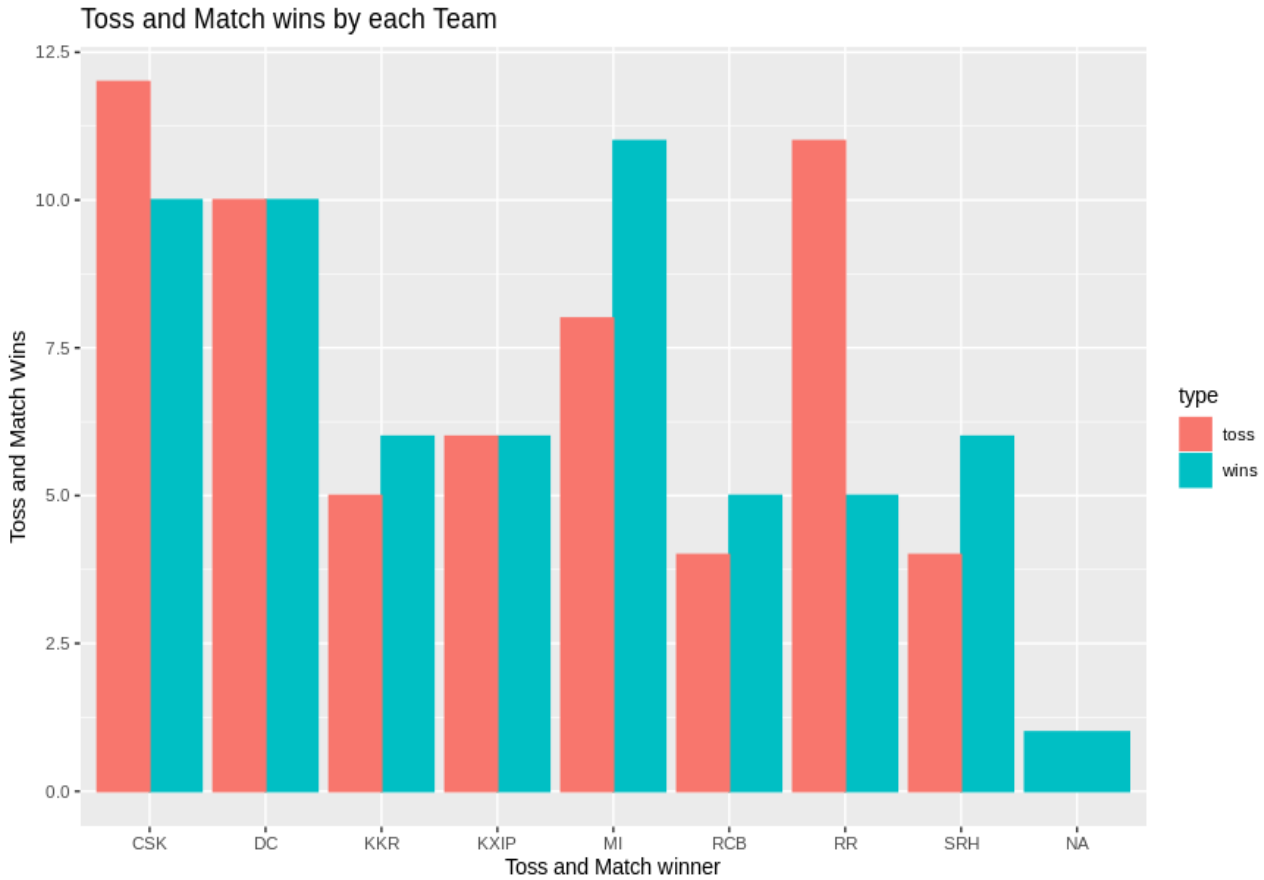
```
##################### toss decision of toss winner ######################
wins_1 <- matches%>%
  left_join(teams,by=c("toss_winner"="team") )%>%
  select(s_team,toss_winner,toss_decision)%>%
  group_by(s_team,toss_decision)%>%
  summarize(wins=n())



ggplot(wins_1,aes(x=s_team,y=wins,colour=toss_decision,fill=toss_decision))+
  geom_bar(position = "dodge",stat = "identity")+
  theme(legend.position="right")+
  scale_y_continuous(name="Toss decision")+
  scale_x_discrete(name="Toss winners and toss decisions")+
  ggtitle("Toss decisions by each Team")
```



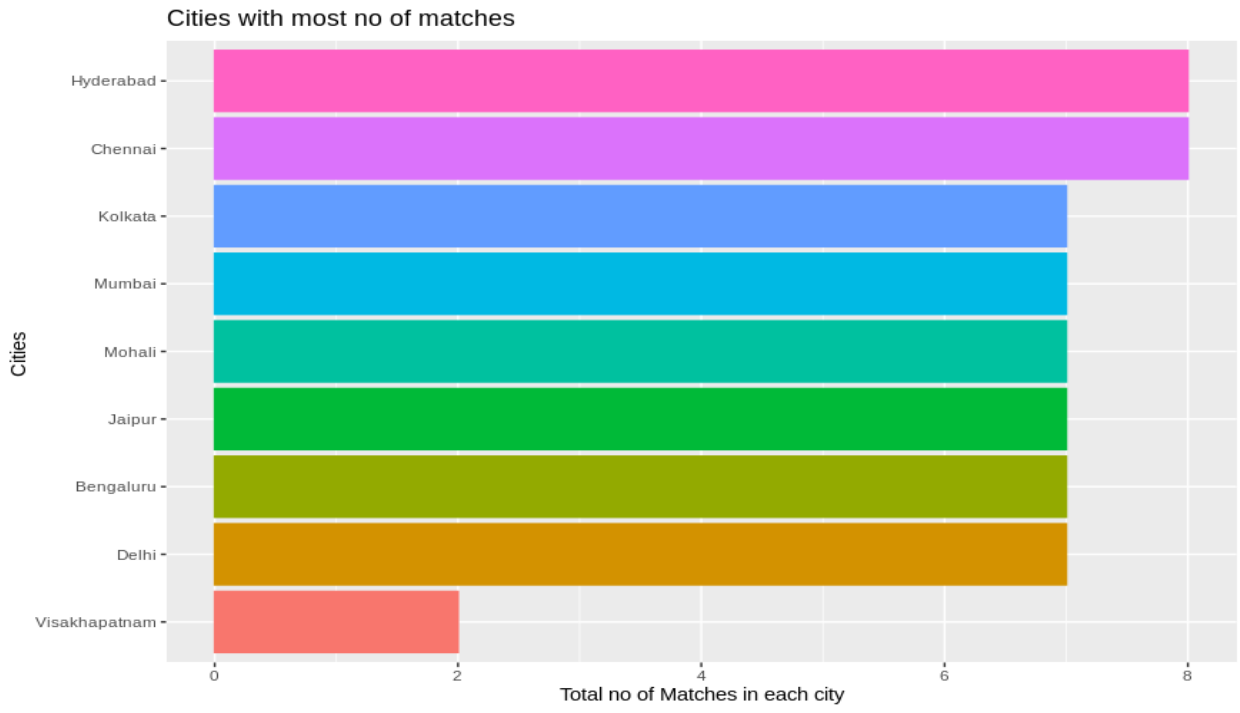

```
(We can see that CSK and RR choosen fielding after winning the toss and KKR, KXIP, SRH never
                          batted first after winning the toss)
############################## Toss and match win  ###################################
toss <- matches%>%
  left_join(teams,by=c("toss_winner"="team") )%>%
```

```
  select(s_team,toss_winner)%>%
  group_by(s_team)%>%
  summarize(wins=n())
toss$type <- "toss"
wins <-matches%>%
  left_join(teams,by=c("winner"="team") )%>%
  select(s_team,winner)%>%
  group_by(s_team)%>%
  summarize(wins=n())
wins$type <- "wins"
toss_w <- rbind(toss,wins)
toss_w <- toss_w %>%
  group_by(s_team, type)%>%
  summarize(wins=sum(wins))
ggplot(toss_w,aes(x=s_team,y=wins,colour=type,fill=type))+
  geom_bar(position = "dodge",stat = "identity")+
  theme(legend.position="right")+
  scale_y_continuous(name="Toss and Match Wins")+
  scale_x_discrete(name="Toss and Match winner")+
  ggtitle("Toss and Match wins by each Team")
```

(We can see in the below plot that DC and XXIP won every match when they won the toss)



```
######################### city with most number of match  ###################
venue_c <- data%>%
  left_join(matches,by=c("match_id"="id"))%>%
  select(match_id,city,total_runs,wickets)%>%
  group_by(city)%>%
  summarize(runs=sum(total_runs),wickets=sum(wickets,na.rm=TRUE))
city_mat <- matches %>%
  group_by(city)%>%
  summarize(matches=n())
venue_c <- venue_c %>%
  left_join(city_mat, by=c("city"="city"))%>%
  mutate(Avg_runs=runs/matches)%>%
  mutate(Avg_wkt =wickets/matches)%>%
  arrange(city)
venue_all <- venue_c%>%
  left_join(venue_city, by=c("city"="city"))%>%
  arrange(Avg_runs)
venue_all$city <- factor(venue_all$city, levels = venue_all$city[order(venue_all$matches)])
ggplot(venue_all,aes(x=city,y=matches,colour=city,fill=city))+
  geom_bar(position = "dodge",stat = "identity")+
  theme(legend.position="none")+ coord_flip()+
  scale_y_continuous(name="Total no of Matches in each city")+
  scale_x_discrete(name="Cities ")+
 ggtitle("Cities with most no of matches")
```



(We can see in the above plot that most of the maximum of 9 matches were played in Chennain and Hyderabad)