

Indian Premier League Data Analysis

Dr. Sulove Koirala

5/11/2020

Email: sulovekoirala@gmail.com (mailto:sulovekoirala@gmail.com)

Introduction

We will be going to analyze Indian Premier League (IPL) data. It contains ball by ball data from 2008-2019. We are going to construct question and answer it with the help of the given data. The dataset is downloaded from Kaggle.

Loading the packages

Importing the data and Cleaning it

```
deliveries <- read.csv("C:/Users/sulov/Desktop/PD/ipl/deliveries.csv")
matches <- read.csv("C:/Users/sulov/Desktop/PD/ipl/matches.csv")
```

The deliveries file is somewhat large as it contains ball by ball information. Looking at the structure, it contains characters and integers. Everything looks fine, except the date, which is in character format. We are going to change it in date format.

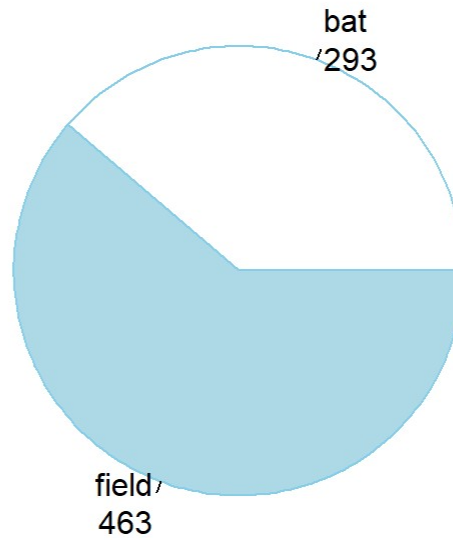
```
# Adjusting to Date format
matches$date = ymd(as.factor(matches$date))
```

Now, let's proceed to questions and answer.

What did the captain choose the most during toss?

```
result <- table(matches$toss_decision)
toss <- paste(names(result), "\n", result, sep="")
pie(result, labels = toss,
     main="Pie Chart of Toss Decision", border = "skyblue")
```

Pie Chart of Toss Decision



Which match was won by a biggest margin in terms of runs/wickets?

Biggest wins by runs

```
runs = head (matches[order(matches$win_by_runs, decreasing = T),], n = 5)
runs = select(runs, winner, win_by_runs, team1, team2, season, date, venue)
kable(runs) %>%
  kable_styling(bootstrap_options = c("striped"))
```

	winner	win_by_runs	team1	team2	season	date	venue
44	Mumbai Indians	146	Mumbai Indians	Delhi Daredevils	2017	2017-05-06	Feroz Shah Kotla
620	Royal Challengers Bangalore	144	Royal Challengers Bangalore	Gujarat Lions	2016	2016-05-14	M Chinnaswamy Stadium
60	Kolkata Knight Riders	140	Kolkata Knight Riders	Royal Challengers Bangalore	2008	2008-04-18	M Chinnaswamy Stadium

	winner	win_by_runs	team1	team2	season	date	venue
557	Royal Challengers Bangalore	138	Royal Challengers Bangalore	Kings XI Punjab	2015	2015-05-06	M Chinnaswamy Stadium
411	Royal Challengers Bangalore	130	Royal Challengers Bangalore	Pune Warriors	2013	2013-04-23	M Chinnaswamy Stadium

Biggest wins by wickets in hand

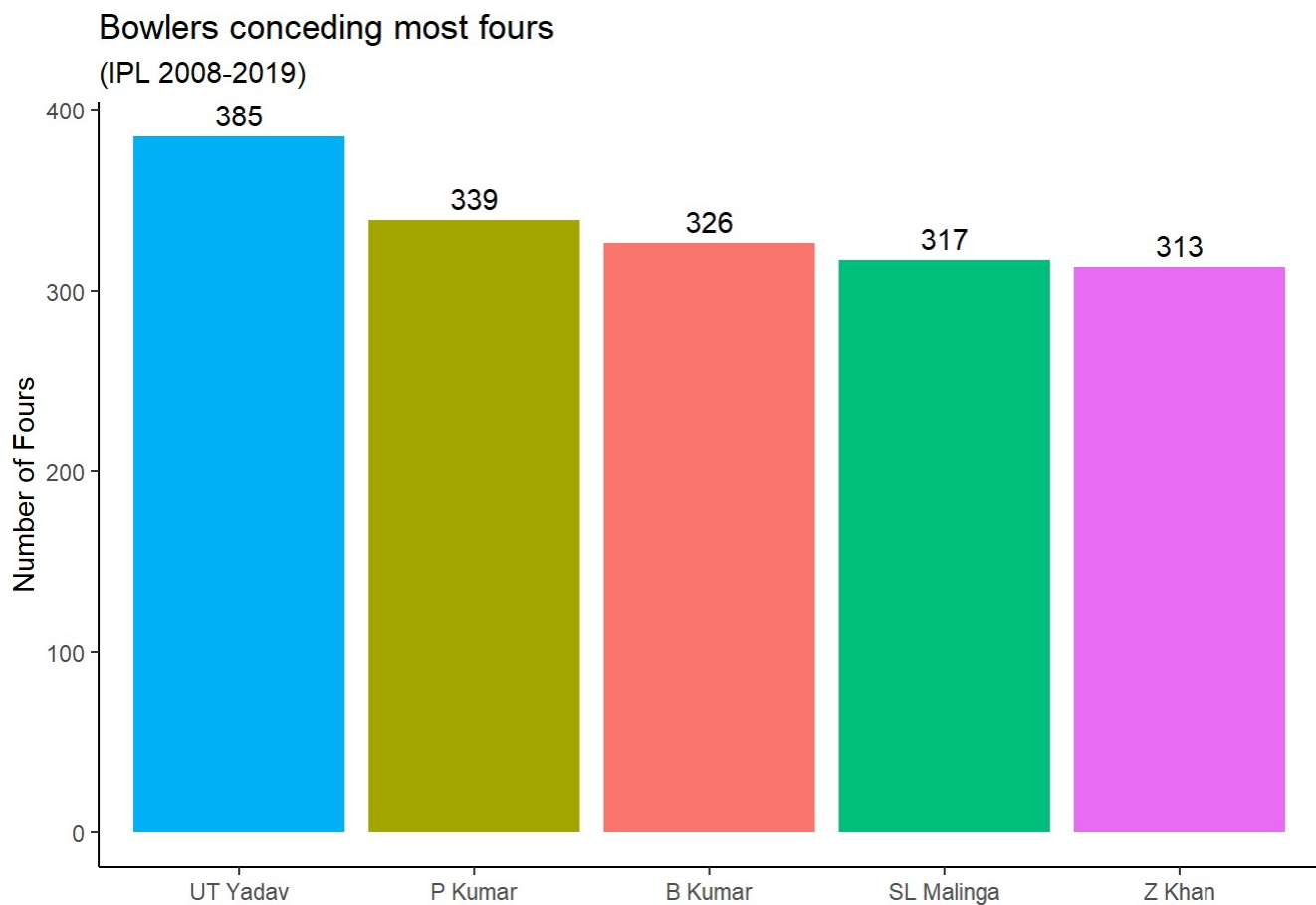
```
wickets = filter(matches, win_by_wickets == 10)
wickets = select(wickets, winner, win_by_wickets, team1, team2, season, date, venue)
kable(wickets) %>%
  kable_styling(bootstrap_options = c("striped"))
```

	winner	win_by_wickets	team1	team2	season	date	venue
	Kolkata Knight Riders	10	Gujarat Lions	Kolkata Knight Riders	2017	2017-04-07	Saurashtra Cricket Association Stadium
	Kings XI Punjab	10	Delhi Daredevils	Kings XI Punjab	2017	2017-04-30	Punjab Cricket Association IS Bindra Stadium, Mohali
	Deccan Chargers	10	Mumbai Indians	Deccan Chargers	2008	2008-04-27	Dr DY Patil Sports Academy
	Delhi Daredevils	10	Kings XI Punjab	Delhi Daredevils	2009	2009-04-19	Newlands
	Royal Challengers Bangalore	10	Rajasthan Royals	Royal Challengers Bangalore	2010	2010-03-18	M Chinnaswamy Stadium
	Rajasthan Royals	10	Mumbai Indians	Rajasthan Royals	2011	2011-05-20	Wankhede Stadium
	Mumbai Indians	10	Rajasthan Royals	Mumbai Indians	2012	2012-05-20	Sawai Mansingh Stadium
	Chennai Super Kings	10	Kings XI Punjab	Chennai Super Kings	2013	2013-04-10	Punjab Cricket Association Stadium, Mohali

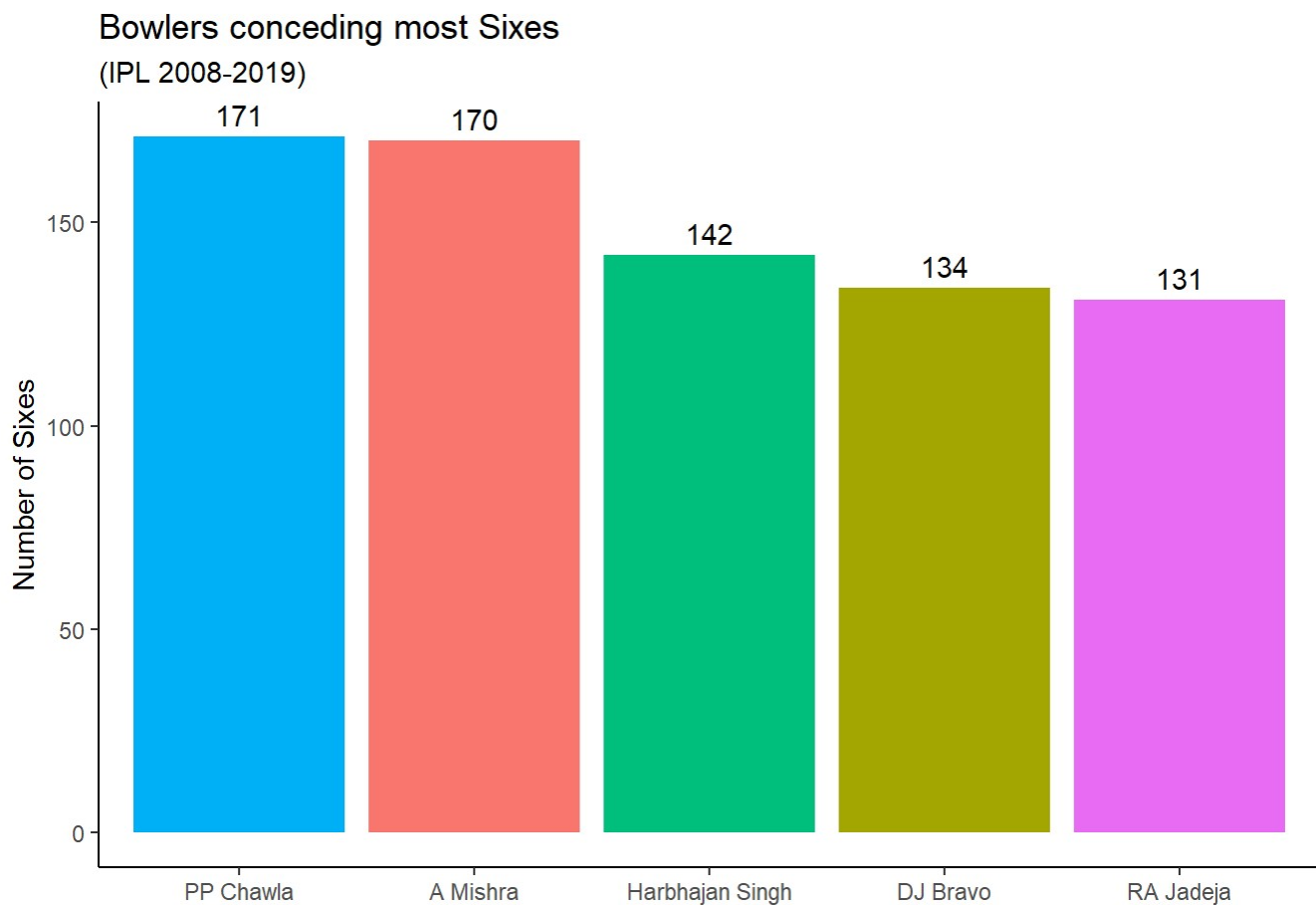
winner	win_by_wickets	team1	team2	season	date	venue
Royal Challengers Bangalore	10	Delhi Daredevils	Royal Challengers Bangalore	2015	2015-04-26	Feroz Shah Kotla
Sunrisers Hyderabad	10	Gujarat Lions	Sunrisers Hyderabad	2016	2016-04-21	Saurashtra Cricket Association Stadium
Royal Challengers Bangalore	10	Kings XI Punjab	Royal Challengers Bangalore	2018	2014-05-18	Holkar Cricket Stadium

Which bowler has gone from most sixes and fours?

```
fours = select(filter(deliveries, batsman_runs == 4 | batsman_runs == 5), bowler, batsman_runs)
fours = aggregate(fours$batsman_runs~fours$bowler, FUN = sum)
colnames(fours) = c("Bowler", "Number")
fours$Number = round(fours$Number/4, digits = 0)
fours = fours[order(fours$Number, decreasing = T),]
ggplot(head(fours, n = 5))+
  aes(reorder(Bowler, -Number), Number, fill = Bowler)+
  geom_bar(stat = "identity")+
  geom_text(aes(Bowler, Number, label = Number), vjust = -0.5)+
  theme_classic()+
  labs(x = "", y = "Number of Fours", title = "Bowlers conceding most fours", subtitle = "(IPL 2008-2019)")+
  guides(fill = FALSE)
```



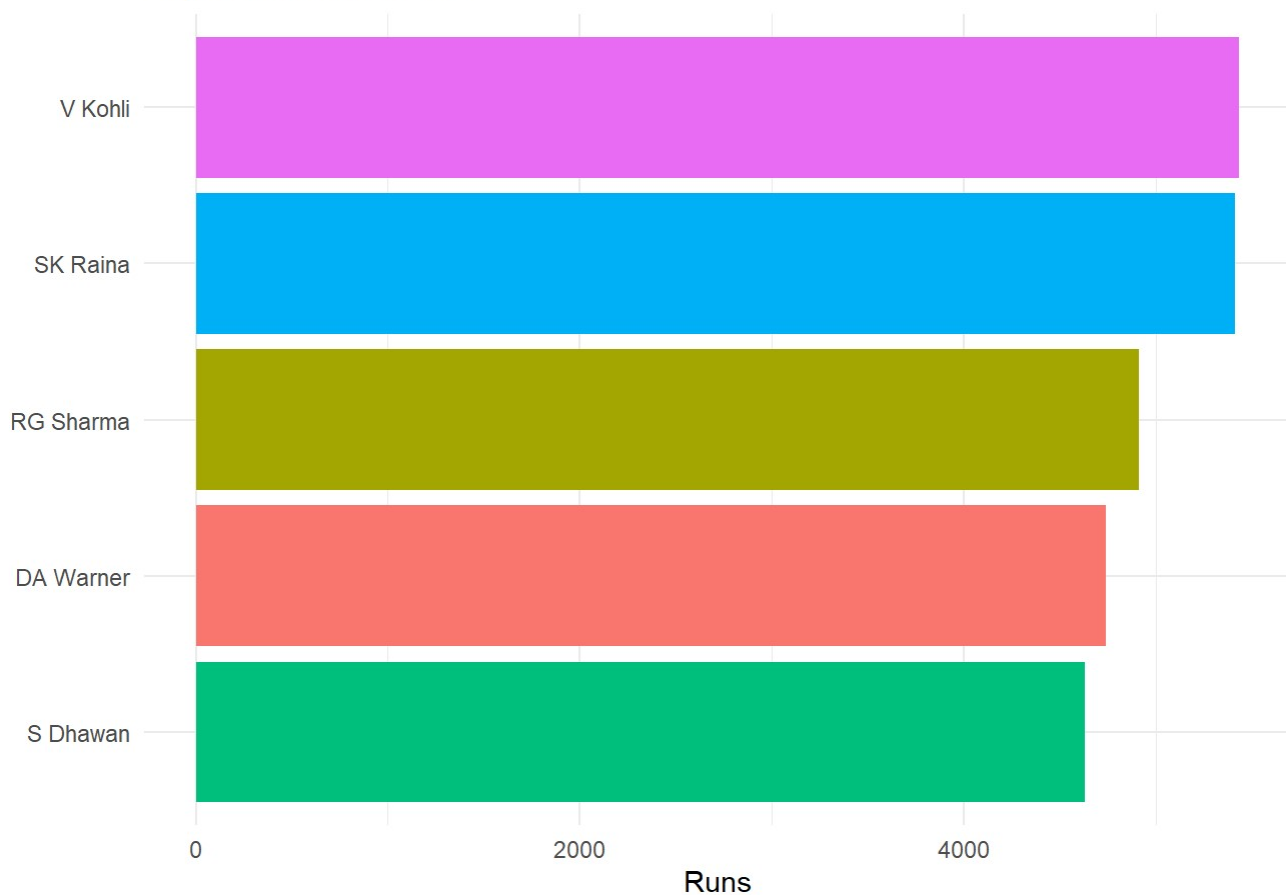
```
six = select(filter(deliveries, batsman_runs == 6 | batsman_runs == 57), bowler, batsman_runs)
six = aggregate(six$batsman_runs~six$bowler, FUN = sum)
colnames(six) = c ("Bowler", "Number")
six$Number = round(six$Number/6, digits = 0)
six = six[order(six$Number, decreasing = T),]
ggplot(head(six, n = 5))+
  aes(reorder(Bowler, -Number), Number, fill = Bowler)+
  geom_bar(stat = "identity")+
  geom_text(aes(Bowler, Number, label = Number), vjust = -0.5)+
  theme_classic()+
  labs(x = "", y = "Number of Sixes", title = "Bowlers conceding most Sixes", subtitle = "(IPL 2008-2019)") +
  guides(fill = FALSE)
```



Which batsman has score most runs in IPL ?

```
deliveries%>%
  group_by(batsman)%>%
  summarize(runs = sum(batsman_runs))%>%
  top_n(n = 5, wt = runs)%>%
  ggplot(aes(x = reorder(batsman, runs), y = runs, fill = batsman))+
  geom_bar(stat = "identity", show.legend = FALSE)+
  coord_flip()+
  labs(x = "", y = "Runs", title = "Top 5 batsman in IPL")+
  theme_minimal()
```

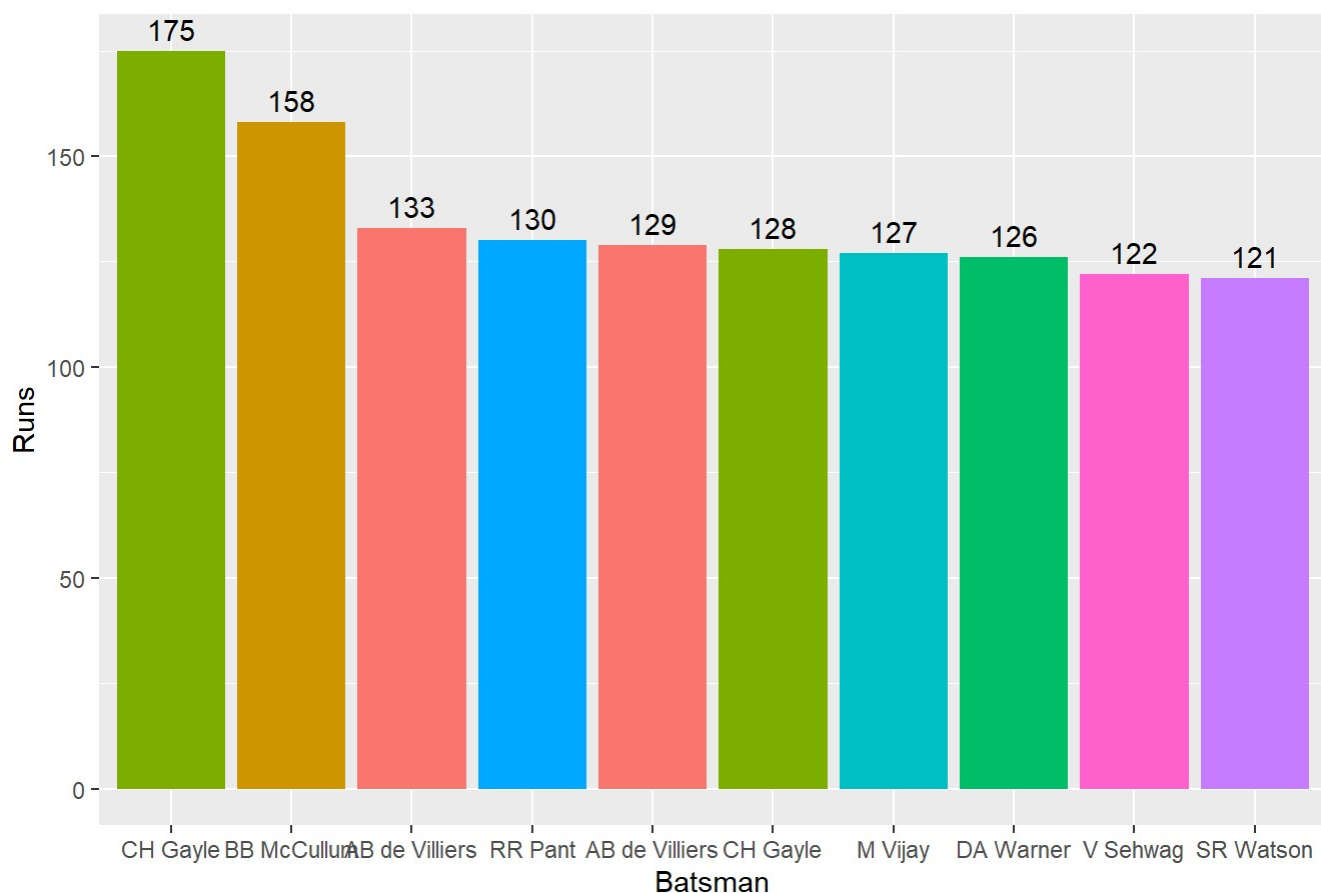
Top 5 batsman in IPL



What is the highest score in the IPL?

```
score <- deliveries%>%
  group_by(match_id,batsman)%>%
  summarize(personal = sum(batsman_runs))%>%
  arrange(desc(personal))%>%
  ungroup()%>%
  top_n(10, personal)%>%
  mutate(id = as.factor(row_number()))
score%>%
  ggplot(aes(x= id, y = personal,fill = batsman))+
  geom_bar(stat= "identity",show.legend = FALSE)+
  scale_x_discrete(labels = score$batsman, name = "Batsman")+
  labs(y = "Runs", title ="Highest individual Score in a Match")+
  geom_text(aes(label = personal), vjust = -0.5)
```

Highest individual Score in a Match



Who has completed most catches?

```
catch = filter(deliveries, deliveries$dismissal_kind == "caught")
catch = table(catch$fielder)
catch = as.data.frame(catch)
catch = catch [order(catch$Freq, decreasing = T),]
kable(head(catch, n = 5)) %>%
  kable_styling(bootstrap_options = c("striped"))
```

	Var1	Freq
210	KD Karthik	109
419	SK Raina	99
285	MS Dhoni	98
19	AB de Villiers	93
370	RV Uthappa	84

Bibliography

<https://www.kaggle.com/nowke9> (<https://www.kaggle.com/nowke9>)