IPL Data Analysis

Srinivas Shekar 01FB15ECS308

By 17 August 2017

Loading Data:

data<-read.csv("deliveries.csv",sep=",")  
result1<-read.csv('matches.csv',sep=',')

Loading required libraries:

library(magrittr)

## Warning: package 'magrittr' was built under R version 3.4.1

library(dplyr)  
library(ggplot2)  
library(gdata)

## Question 1:Analysis on IPL Final 2016

1.Loading required data :

d<-subset(data,match\_id==577)  
data\_srh<-subset(d,batting\_team=='Sunrisers Hyderabad')  
data\_rcb<-subset(d,batting\_team=='Royal Challengers Bangalore')

2.Target set by SRH :

s<-sum(data\_srh$total\_runs)  
print(s)

## [1] 208

3."Beat the Run Rate Boys!!"- V Kohli

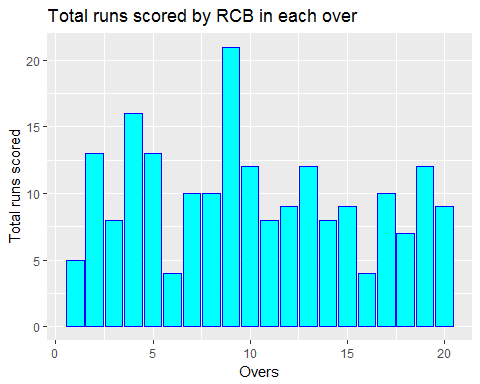
run\_rate<-s/max(data\_srh$over)  
print(run\_rate)

## [1] 10.4

4.RCB's batting analyis :

df\_plot <-data\_rcb %>% group\_by(over)%>% summarise(runs=sum(total\_runs)) %>% arrange(over)  
  
df\_plot %>% ggplot(aes(over,runs))+geom\_histogram(stat="identity",color='blue',fill='cyan')+xlab('Overs')+  
 ylab('Total runs scored')+ggtitle('Total runs scored by RCB in each over')

## Warning: Ignoring unknown parameters: binwidth, bins, pad



## Question 2:Analysis on team's batting performance

1.Hungry for runs???Here is the total runs scored by teams in all seasons

d\_7<-subset(data,(match\_id %in% c(7,27,67,171,414)) & is\_super\_over==0)  
d\_7 %>% group\_by(inning,match\_id,batting\_team) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)

## # A tibble: 10 x 4  
## # Groups: inning, match\_id [10]  
## inning match\_id batting\_team total\_runs  
## <int> <int> <fctr> <int>  
## 1 1 7 Deccan Chargers 142  
## 2 2 7 Delhi Daredevils 143  
## 3 1 27 Rajasthan Royals 103  
## 4 2 27 Mumbai Indians 104  
## 5 1 67 Rajasthan Royals 150  
## 6 2 67 Kolkata Knight Riders 150  
## 7 1 171 Mumbai Indians 133  
## 8 2 171 Kolkata Knight Riders 135  
## 9 1 414 Mumbai Indians 125  
## 10 2 414 Delhi Daredevils 126

2.Record breakers!!

result2<-subset(result1,(result %in% c('normal')) &(dl\_applied==0))  
result2 <- rename.vars(result2, from = "id", to = "match\_id")

##   
## Changing in result2   
## From: id   
## To: match\_id

s1<-subset(data,match\_id %in% result2$match\_id)  
  
q<-s1 %>% group\_by(match\_id,inning,batting\_team) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
  
max\_team<-subset(q,max(q$total\_runs)==total\_runs)  
max\_team

## # A tibble: 1 x 4  
## # Groups: match\_id, inning [1]  
## match\_id inning batting\_team total\_runs  
## <int> <int> <fctr> <int>  
## 1 352 1 Royal Challengers Bangalore 263

3.Oops!! Record breakers or Record ....

min\_team<-subset(q,min(q$total\_runs)==total\_runs)  
min\_team

## # A tibble: 1 x 4  
## # Groups: match\_id, inning [1]  
## match\_id inning batting\_team total\_runs  
## <int> <int> <fctr> <int>  
## 1 60 2 Rajasthan Royals 58

## Question 3: IPL Auction:Who is the effective player??Lets look at marquee players

getmode<-function(v){  
 uniqv<-unique(v)  
 uniqv[which.max(tabulate(match(v,uniqv)))]  
}

1.MS Dhoni's IPL career:

msd<-subset(data,batsman=='MS Dhoni')  
msd<-msd %>% group\_by(match\_id,batsman) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
with(msd,sum(total\_runs))

## [1] 3421

with(msd,summary(total\_runs))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 11.00 25.00 26.73 36.25 70.00

with(msd,getmode(total\_runs))

## [1] 43

with(msd,IQR(total\_runs))

## [1] 25.25

with(msd,sd(total\_runs))

## [1] 18.52885

2.V Kohli's IPL career:

kohli<-subset(data,batsman=='V Kohli')  
kohli<-kohli %>% group\_by(match\_id,batsman) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
with(kohli,summary(total\_runs))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 12.00 24.00 32.63 48.50 116.00

with(kohli,sum(total\_runs))

## [1] 4274

with(kohli,getmode(total\_runs))

## [1] 12

with(kohli,IQR(total\_runs))

## [1] 36.5

with(kohli,sd(total\_runs))

## [1] 27.57545

3.ABD Villiers's IPL career:

abd<-subset(data,batsman=='AB de Villiers')  
abd<-abd %>% group\_by(match\_id,batsman) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
with(abd,summary(total\_runs))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 9.00 23.00 31.08 48.00 138.00

with(abd,sum(total\_runs))

## [1] 3388

with(abd,getmode(total\_runs))

## [1] 0

with(abd,IQR(total\_runs))

## [1] 39

with(abd,sd(total\_runs))

## [1] 28.36179

4.Jadeja's IPL career:

jaddu<-subset(data,batsman=='RA Jadeja')  
jaddu<-jaddu %>% group\_by(match\_id,batsman) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
with(jaddu,summary(total\_runs))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 7.00 15.00 16.96 24.75 50.00

with(jaddu,sum(total\_runs))

## [1] 1662

with(jaddu,getmode(total\_runs))

## [1] 0

with(jaddu,IQR(total\_runs))

## [1] 17.75

with(jaddu,sd(total\_runs))

## [1] 12.61395

5.Suresh Raina's IPL career:

raina<-subset(data,batsman=='SK Raina')  
raina<-raina %>% group\_by(match\_id,batsman) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)  
with(raina,summary(total\_runs))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 10.00 26.00 30.01 44.50 105.00

with(raina,sum(total\_runs))

## [1] 4291

with(raina,getmode(total\_runs))

## [1] 0

with(raina,IQR(total\_runs))

## [1] 34.5

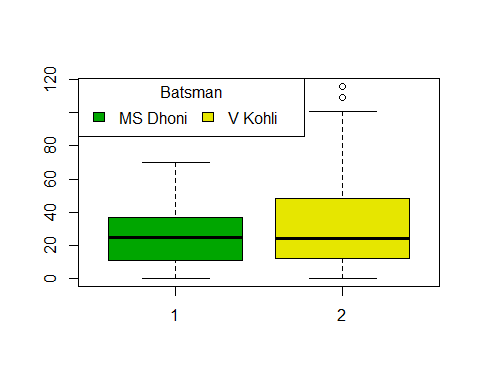
with(raina,sd(total\_runs))

## [1] 24.46455

## Question 4: MSD v/s Kohli

1.Comparitive Boxplots:

boxplot(msd$total\_runs , kohli$total\_runs,col = terrain.colors(4))  
legend("topleft",title="Batsman",c("MS Dhoni","V Kohli"),fill=terrain.colors(4),horiz=TRUE)



2.90th Percentile

quantile(msd$total\_runs,prob=0.9)

## 90%   
## 55.3

quantile(kohli$total\_runs,prob=0.9)

## 90%   
## 74

* 1. Top performances by Kohli

kohli<-kohli %>% arrange(total\_runs)  
tail(kohli)

## # A tibble: 6 x 3  
## # Groups: match\_id [6]  
## match\_id batsman total\_runs  
## <int> <fctr> <int>  
## 1 373 V Kohli 97  
## 2 379 V Kohli 101  
## 3 536 V Kohli 101  
## 4 552 V Kohli 109  
## 5 561 V Kohli 109  
## 6 567 V Kohli 116

1. Top performances by MSD

msd<-msd %>% arrange(total\_runs)  
tail(msd)

## # A tibble: 6 x 3  
## # Groups: match\_id [6]  
## match\_id batsman total\_runs  
## <int> <fctr> <int>  
## 1 15 MS Dhoni 66  
## 2 123 MS Dhoni 67  
## 3 398 MS Dhoni 67  
## 4 570 MS Dhoni 68  
## 5 355 MS Dhoni 69  
## 6 243 MS Dhoni 70

4.(a) Worst performances by Kohli

head(kohli,8)

## # A tibble: 8 x 3  
## # Groups: match\_id [8]  
## match\_id batsman total\_runs  
## <int> <fctr> <int>  
## 1 31 V Kohli 0  
## 2 403 V Kohli 0  
## 3 429 V Kohli 0  
## 4 574 V Kohli 0  
## 5 17 V Kohli 1  
## 6 358 V Kohli 1  
## 7 422 V Kohli 1  
## 8 487 V Kohli 1

1. Worst performances by MSD

head(msd,8)

## # A tibble: 8 x 3  
## # Groups: match\_id [8]  
## match\_id batsman total\_runs  
## <int> <fctr> <int>  
## 1 147 MS Dhoni 0  
## 2 165 MS Dhoni 0  
## 3 514 MS Dhoni 0  
## 4 24 MS Dhoni 1  
## 5 284 MS Dhoni 1  
## 6 468 MS Dhoni 1  
## 7 527 MS Dhoni 1  
## 8 2 MS Dhoni 2

## Question 5: Is Suresh Raina a key player in winning matches for Gujarat Lions??

1.Loading Gujarat Lion's Data...

gl<-subset(data,batting\_team=='Gujarat Lions')  
gl<-gl %>% group\_by(match\_id) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)

2.Loading Raina's data...

raina\_gl<-subset(data,batting\_team=='Gujarat Lions' & batsman=='SK Raina')  
raina\_gl<-raina\_gl %>% group\_by(match\_id) %>% summarise(total\_runs=sum(total\_runs)) %>% arrange(match\_id)

3.Looking at Suresh Raina's impact in the win percentage of the team

raina\_vs\_gl<-merge(raina\_gl,gl,by='match\_id')  
raina\_vs\_gl$impact<-round((raina\_vs\_gl[2]/raina\_vs\_gl[3])\*100,2)  
raina\_vs\_gl

## match\_id total\_runs.x total\_runs.y total\_runs.x  
## 1 520 22 162 13.58  
## 2 523 26 164 15.85  
## 3 526 28 147 19.05  
## 4 532 76 135 56.30  
## 5 536 30 182 16.48  
## 6 540 2 172 1.16  
## 7 542 34 196 17.35  
## 8 545 18 131 13.74  
## 9 548 26 149 17.45  
## 10 551 20 126 15.87  
## 11 555 14 164 8.54  
## 12 568 61 125 48.80  
## 13 571 59 173 34.10  
## 14 574 1 158 0.63  
## 15 576 1 162 0.62

win<-subset(result1,winner=="Gujarat Lions")  
win1<-subset(raina\_vs\_gl,match\_id %in% win$id)  
  
colnames(win1)<-c("Match id","Raina's score","Team Score","Impact of Raina on the team")  
win1

## Match id Raina's score Team Score total\_runs.x  
## 1 520 22 162 13.58  
## 2 523 26 164 15.85  
## 3 526 28 147 19.05  
## 5 536 30 182 16.48  
## 6 540 2 172 1.16  
## 7 542 34 196 17.35  
## 11 555 14 164 8.54  
## 12 568 61 125 48.80  
## 13 571 59 173 34.10