

# Data Analytics Lab Assignment-2

Kapil Jhade(17BCS013)

→ Ranking and Analysis of 10 ten IPL-2019 players(bowlers/batsmens) using weighted model

*(Below is the weitage for Ranking top 10 batsmens and bowlers based on some weightage)*

- **Weightage taken for different parameters for batsman:**

```
w0=-2 #duck  
w1=1 #six  
w2=2 #fours  
w3=3 #Strike rate  
w4=10 #Average  
w5=8 #no_of_50s  
w6=16 #no_of_100s
```

- **Weightage taken for different parameters for bowlers:**

```
wb0=4 #madains  
wb1=8 #5wkts_in_a_match  
wb2=4 #4wkts_in_a_match  
wb3=5 #wkts/match  
wb4=4 #economy
```

*(Below is the code for Ranking top 10 batsmens and bowlers based on some weightage)*

```
library(igraph)  
library(ggplot2)  
bat_data=read.table("/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA LAB/lab2/espn  
batsman.csv", sep = ",", header = TRUE)  
#ipl_data  
w0=-2  
w1=1  
w2=2  
w3=3  
w4=10  
w5=8  
w6=16
```

```
bat_data['rank_score1']=w6*bat_data$X100 + w5*bat_data$X50 + w4*bat_data$Ave + w3*bat_data$SR +
w2*bat_data$X4s + w1*bat_data$X6s + w0*bat_data$X0
#Rank scaling done below
bat_data['rank_score1']=bat_data['rank_score1']/14
R1 <- bat_data[order(bat_data$rank_score1,decreasing=T),]
R1
R1[1:10,]
write.csv(R1[1:10,],"/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA
LAB/lab2/top_10_batsman.csv",)
```

```
BOWLER_DATA=read.table("/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA LAB/lab2/espn
bowler.csv", sep = ",", header = TRUE)
#BOWLER_DATA
BOWLER_DATA['WKTS/MAT']=BOWLER_DATA['Wkts']/BOWLER_DATA['Mat']
wb0=4 #8
wb1=8 #16
wb2=4 #8
wb3=5 #10
wb4=4 #8
```

```
BOWLER_DATA['rank_score2']=wb0*BOWLER_DATA$Mdns + wb1*BOWLER_DATA$X5 +
wb2*BOWLER_DATA$X4 + wb3*BOWLER_DATA$'WKTS/MAT'- wb4*BOWLER_DATA$Econ
BOWLER_DATA['rank_score2']=BOWLER_DATA['rank_score2'] + 100
R2 <- BOWLER_DATA[order(BOWLER_DATA$rank_score2,decreasing=T),]
R2
R2[1:10,]
write.csv(R2[1:10,],"/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA
LAB/lab2/top_10_bowler.csv")
```

**(Below is the rankings obtained for top 10 batsmens and bowlers as per weightages )**

### ★ Top 10 Batsmen ranking

```
> bat_analysis
```

	X	Player	Mat	Inns	NO	Runs	HS	Ave	BF	SR	X100	X50	X0	X4s	X6s	rank_score1
1	1	DA Warner	12	12	2	692	100*	69.20	481	143.86	1	8	0	57	21	95.61286
2	13	MS Dhoni	15	12	7	416	84*	83.20	309	134.62	0	3	0	22	23	94.77571
3	5	AD Russell	14	13	4	510	80*	56.66	249	204.81	0	4	1	31	52	94.64500
4	10	JM Bairstow	10	10	2	445	114	55.62	283	157.24	1	2	1	48	18	83.70857
5	2	KL Rahul	14	14	3	593	100*	53.90	438	135.38	1	6	0	49	25	80.86714
6	16	HH Pandya	16	15	6	402	91	44.66	210	191.42	0	1	1	28	29	79.41857
7	11	AB de Villiers	13	13	3	442	82*	44.20	287	154.00	0	5	0	31	26	73.71429
8	6	CH Gayle	13	13	1	490	99*	40.83	319	153.60	0	4	0	45	34	73.22143
9	7	RR Pant	16	16	3	488	78*	37.53	300	162.66	0	3	0	37	27	70.59143
10	37	MP Stoinis	10	10	6	211	46*	52.75	156	135.25	0	0	2	14	10	69.0892

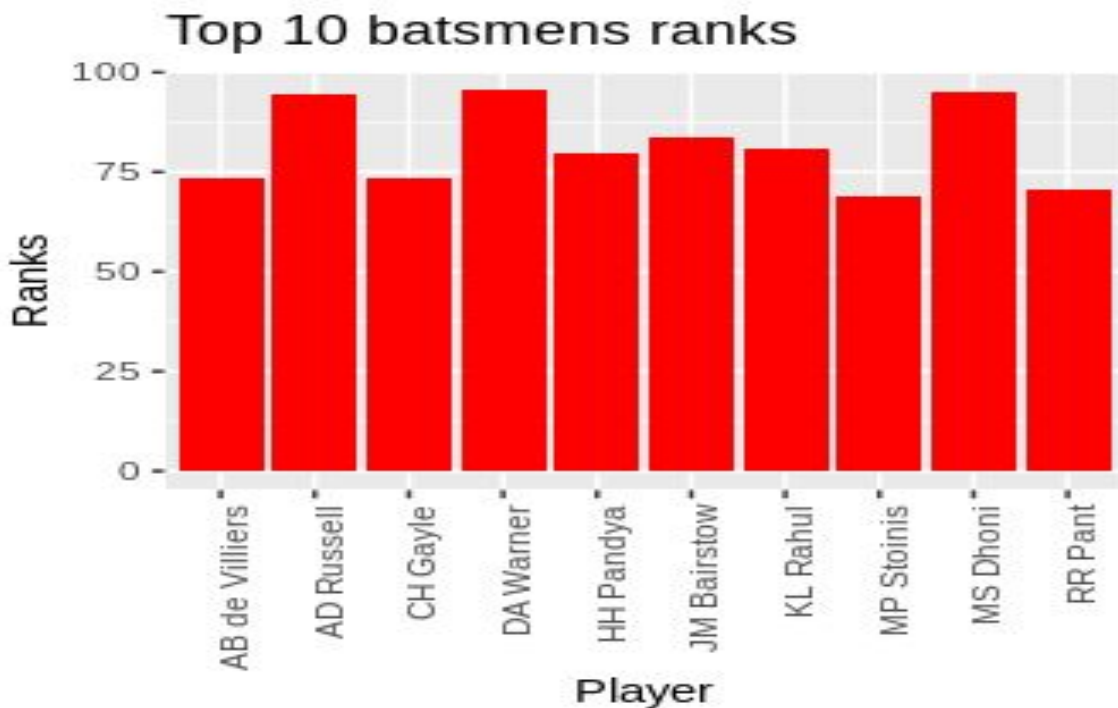
## ★ Top 10 Bowler ranking

```
> R2[1:10,]
```

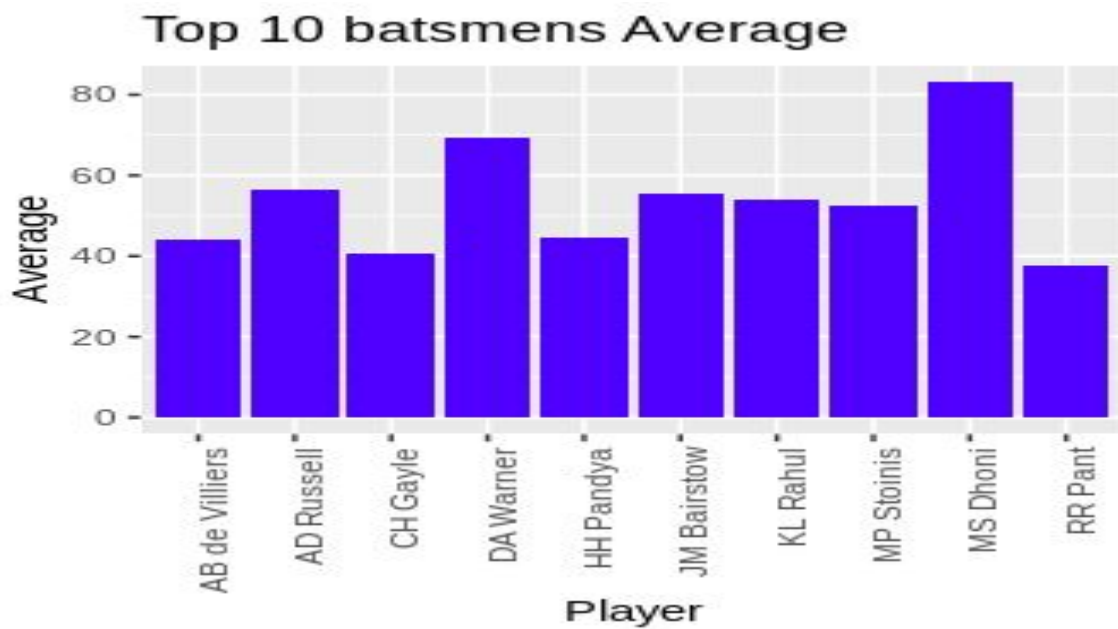
	Player	Mat	Inns	Overs	Mdns	Runs	Wkts	BBI	Ave	Econ	SR	X4	X5	WKTS/MAT	rank_score2
1	Imran Tahir	17	17	64.2	1	431	26	4/12	16.57	6.69	14.8	2	0	1.529412	92.88706
2	K Rabada	12	12	47.0	0	368	25	4/21	14.72	7.82	11.2	2	0	2.083333	87.13667
23	JC Archer	11	11	43.0	2	291	11	3/15	26.45	6.76	23.4	0	0	1.000000	85.96000
3	DL Chahar	17	17	64.3	2	482	22	3/20	21.90	7.47	17.5	0	0	1.294118	84.59059
9	Rashid Khan	15	15	60.0	1	377	17	3/21	22.17	6.28	21.1	0	0	1.133333	84.54667
6	JJ Bumrah	16	16	61.4	1	409	19	3/20	21.52	6.63	19.4	0	0	1.187500	83.41750
12	RA Jadeja	16	16	54.0	1	343	15	3/9	22.86	6.35	21.6	0	0	0.937500	83.28750
8	YS Chahal	14	14	49.2	1	386	18	4/38	21.44	7.82	16.4	1	0	1.285714	83.14857
10	Harbhajan Singh	11	11	44.0	1	312	16	3/20	19.50	7.09	16.5	0	0	1.454545	82.91273
32	Mohammad Nabi	8	8	29.1	0	194	8	4/11	24.25	6.65	21.8	1	0	1.000000	82.40000

## ★ Bar plot for rankings(as per model), Average, and strike rate of top 10 batsmens

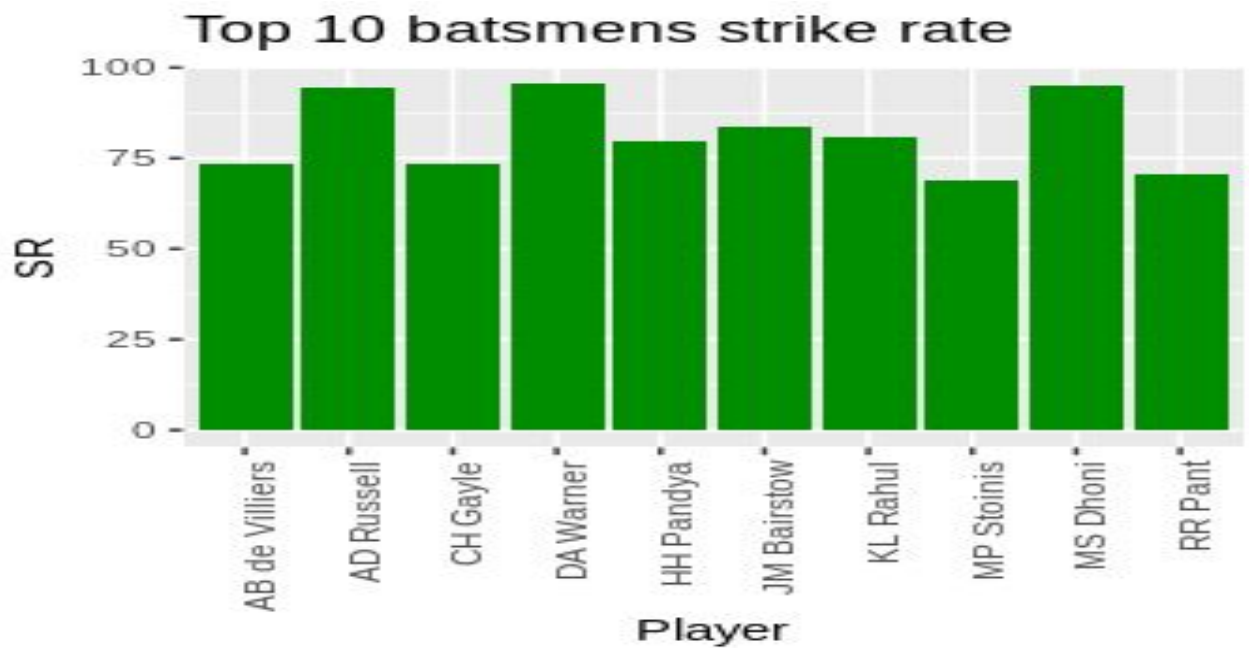
```
> ggplot(bat_analysis,aes(x =Player,y=rank_score1))+ylab("Ranks") + geom_bar(stat = "identity",position = "dodge",fill="#FF0000") + theme(axis.text.x = element_text(angle = 90, hjust = 1))+ggtitle("Top 10 batsmens ranks")
```



```
> ggplot(bat_analysis,aes(x =Player,y=Ave))+ylab("Average") + geom_bar(stat =
"identity",position = "dodge",fill="#0000FF") + theme(axis.text.x =
element_text(angle = 90, hjust = 1))+ggtitle("Top 10 batsmens Average")
```

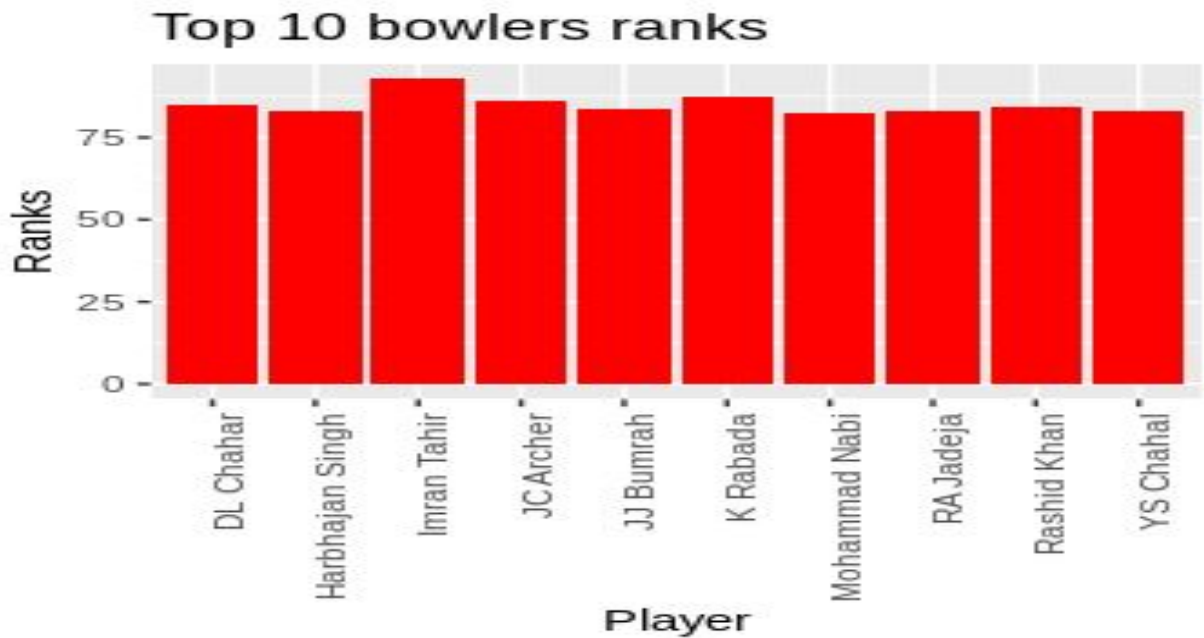


```
> ggplot(bat_analysis,aes(x =Player,y=rank_score1))+ylab("SR") + geom_bar(stat =
"identity",position = "dodge",fill="#008000") + theme(axis.text.x =
element_text(angle = 90, hjust = 1))+ggtitle("Top 10 batsmens strike rate")
```

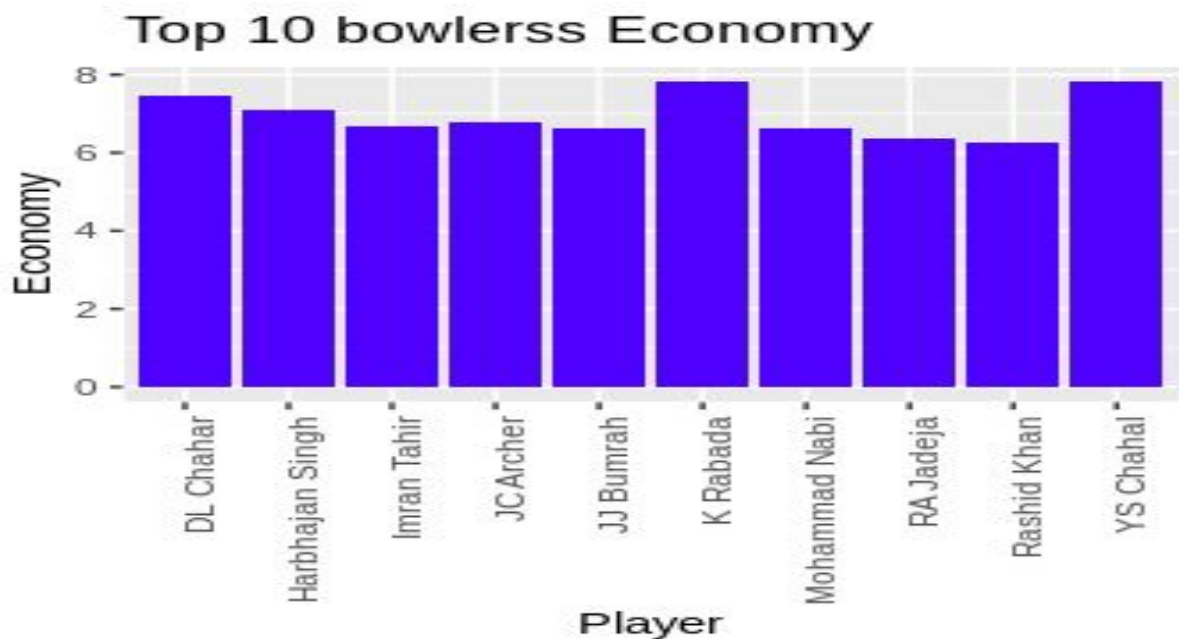


★ **Bar plot for rankings(as per model), Average, and strike rate of top 10 batsmens**

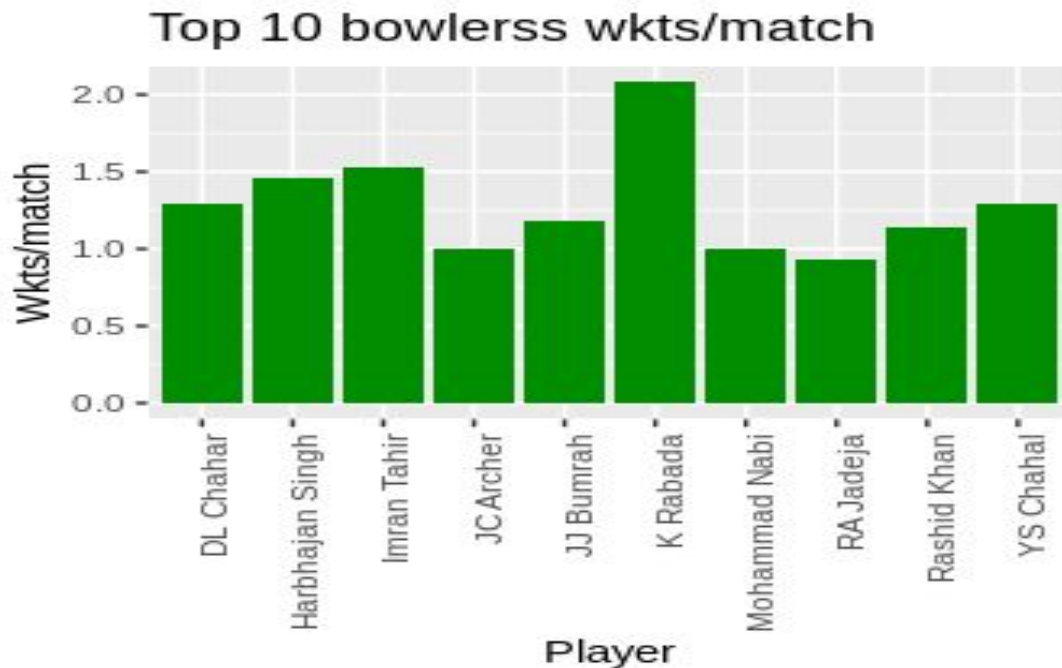
```
> ggplot(bol_analysis,aes(x =Player,y=rank_score2))+ylab("Ranks") + geom_bar(stat = "identity",position = "dodge",fill="#FF0000") + theme(axis.text.x = element_text(angle = 90, hjust = 1))+ggtitle("Top 10 bowlers ranks")
```



```
> ggplot(bol_analysis,aes(x =Player,y=Econ))+ylab("Economy") + geom_bar(stat = "identity",position = "dodge",fill="#0000FF") + theme(axis.text.x = element_text(angle = 90, hjust = 1))+ggtitle("Top 10 bowlersss Economy")
```



```
> ggplot(bol_analysis,aes(x =Player,y=WKTS.MAT))+ylab("Wkts/match") +
geom_bar(stat = "identity",position = "dodge",fill="#008000") + theme(axis.text.x
= element_text(angle = 90, hjust = 1))+ggtitle("Top 10 bowlerss wkts/match")
```



→ **Coefficient of variation(consistency plot) and Correlation coefficient(heatmap plot) of 10 IPL-2019 players(batsmens)**

*(Below is the code for finding match wise score of top 10 players(Batsmens) from deliveries.csv)*

```
library(igraph)
library(ggplot2)
d = read.csv("/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA LAB/lab2/deliveries.csv")
d
players_runs = d[,c('match_id','batsman','batsman_runs')]
players_runs

players = unique(players_runs$batsman)
match_id = unique(d$match_id)

batman = c('DA Warner','KL Rahul', 'MS Dhoni', 'J Bairstow', 'MP Stoinis', 'AD Russell', 'CH Gayle', 'HH Pandya', 'AB
de Villiers','RR Pant')
```

```

#View(d)
s = c()
rs = c()
for(match in match_id){
  df = d[which(d$match_id == match),c('batsman','batsman_runs')]
  player_name = c()
  runs = c()
  for(p in unique(df$batsman)){

    if(p %in% batman){
      player_name = c(player_name,p)
      runs =c(runs , sum(df[which(df$batsman==p),'batsman_runs']))}
    print(match)
  }
  s = c(s,player_name )
  rs = c(rs,runs)
}
dfo = data.frame(s,rs)
dfo
total_run = c()
j = 1
for (pr in batman)
{

  run = c()
  for(i in rownames(dfo)){
    if(dfo[i,1]== pr){
      run=c(run,dfo[i,2])
    }

  }
  total_run[[j]] = run
  j = j+1
}
batman
total_run
dfd = data.frame(matrix(ncol = 10,nrow=10))
dfd
colnames(dfd) <- batman
dfd
for(i in rownames(dfd)){
  i= as.numeric(i)
  for(j in 1:10){
    dfd[i,j] = total_run[[j]][i]
  }}

```

```
dfd
check = dfd
check
write.csv(check, file = "/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA
LAB/lab2/top10player_cv.csv", row.names =
c('match1','match2','match3','match4','match5','match6','match7','match8','match9','match10'))
```

```
data_for_covarience=read.table("/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA
LAB/lab2/top10player_cv.csv", sep = ",", header = TRUE)
```

**(Below is the summary for top 10 batsmen scores for different matches)**

```
> summary(data_for_covarience)
```

X	DA.Warner	KL.Rahul	MS.Dhoni	J.Bairstow
MP.Stoinis	AD.Russell			
match1 :1	Min. : 14.00	Min. : 1.00	Min. :10.00	Min. : 0.0
Min. : 0.00	Min. :10.0			
match10:1	1st Qu.: 51.00	1st Qu.: 12.75	1st Qu.:20.50	1st Qu.: 22.5
1st Qu.:14.25	1st Qu.:23.5			
match2 :1	Median : 64.50	Median : 35.50	Median :38.50	Median : 46.0
Median :23.50	Median :49.5			
match3 :1	Mean : 59.60	Mean : 40.50	Mean :42.40	Mean : 46.8
Mean :21.60	Mean :43.1			
match4 :1	3rd Qu.: 73.25	3rd Qu.: 67.00	3rd Qu.:56.25	3rd Qu.: 60.5
3rd Qu.:30.25	3rd Qu.:52.0			
match5 :1	Max. :101.00	Max. :104.00	Max. :89.00	Max. :115.0
Max. :48.00	Max. :72.0			
(Other):4				
CH.Gayle	HH.Pandya	AB.de.Villiers	RR.Pant	
Min. : 6.00	Min. : 0.00	Min. : 1.00	Min. : 5.00	
1st Qu.: 21.75	1st Qu.:25.00	1st Qu.:14.25	1st Qu.: 9.75	
Median : 37.50	Median :28.00	Median :42.50	Median :21.00	
Mean : 47.20	Mean :25.90	Mean :42.60	Mean :27.80	
3rd Qu.: 68.00	3rd Qu.:32.75	3rd Qu.:71.00	3rd Qu.:39.25	
Max. :105.00	Max. :40.00	Max. :86.00	Max. :80.00	



**(coefficient of variance of top 10 batsmen scores)**

```
> sd(data_for_covariance$DA.Warner)/mean(data_for_covariance$DA.Warner)*100
[1] 46.82113
> sd(data_for_covariance$KL.Rahul)/mean(data_for_covariance$KL.Rahul)*100
[1] 87.03485
> sd(data_for_covariance$MS.Dhoni)/mean(data_for_covariance$MS.Dhoni)*100
[1] 63.29666
> sd(data_for_covariance$J.Bairstow)/mean(data_for_covariance$J.Bairstow)*100
[1] 75.10649
> sd(data_for_covariance$MP.Stoinis)/mean(data_for_covariance$MP.Stoinis)*100
[1] 69.67383
> sd(data_for_covariance$AD.Russell)/mean(data_for_covariance$AD.Russell)*100
[1] 51.79381
> sd(data_for_covariance$CH.Gayle)/mean(data_for_covariance$CH.Gayle)*100
[1] 71.1828
> sd(data_for_covariance$HH.Pandya)/mean(data_for_covariance$HH.Pandya)*100
[1] 44.26424
> sd(data_for_covariance$AB.de.Villiers)/mean(data_for_covariance$AB.de.Villiers)*100
[1] 75.50111
> sd(data_for_covariance$RR.Pant)/mean(data_for_covariance$RR.Pant)*100
[1] 84.4077
```

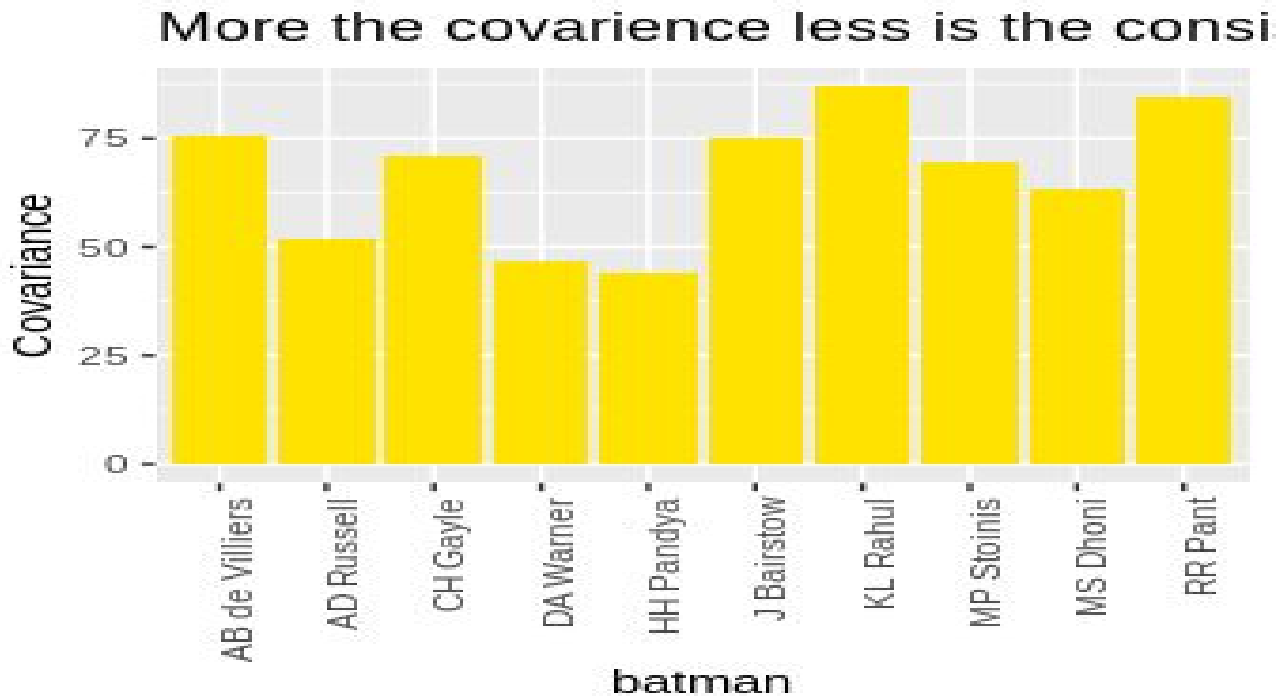
**(Below is the code to create covariance matrix for top 10 batsmen scores)**

```
cov=c(sd(data_for_covariance$DA.Warner)/mean(data_for_covariance$DA.Warner)*100,
      sd(data_for_covariance$KL.Rahul)/mean(data_for_covariance$KL.Rahul)*100,
      sd(data_for_covariance$MS.Dhoni)/mean(data_for_covariance$MS.Dhoni)*100,
      sd(data_for_covariance$J.Bairstow)/mean(data_for_covariance$J.Bairstow)*100,
      sd(data_for_covariance$MP.Stoinis)/mean(data_for_covariance$MP.Stoinis)*100,
      sd(data_for_covariance$AD.Russell)/mean(data_for_covariance$AD.Russell)*100,
      sd(data_for_covariance$CH.Gayle)/mean(data_for_covariance$CH.Gayle)*100,
      sd(data_for_covariance$HH.Pandya)/mean(data_for_covariance$HH.Pandya)*100,
      sd(data_for_covariance$AB.de.Villiers)/mean(data_for_covariance$AB.de.Villiers)*100,
      sd(data_for_covariance$RR.Pant)/mean(data_for_covariance$RR.Pant)*100
)
cov
batman
gg = data.frame(batman,cov)
gg
ggplot(gg,aes(x =batman,y=cov))+ylab("Covariance") + geom_bar(stat = "identity",position = "dodge",fill =
"#FFD700") + theme(axis.text.x = element_text(angle = 90, hjust = 1))+ggtitle("More the covariance less is
the consistency")
```

### ★ Bar plot for coefficient of variance of top 10 batsmens scores

```
> ggplot(gg,aes(x =batman,y=cov))+ylab("Covariance") + geom_bar(stat = "identity",position =  
"dodge",fill = "#FFD700") + theme(axis.text.x = element_text(angle = 90, hjust =  
1))+ggtitle("More the covariance less is the consistency")
```

**(Consistency plot for top 10 batsmans)**

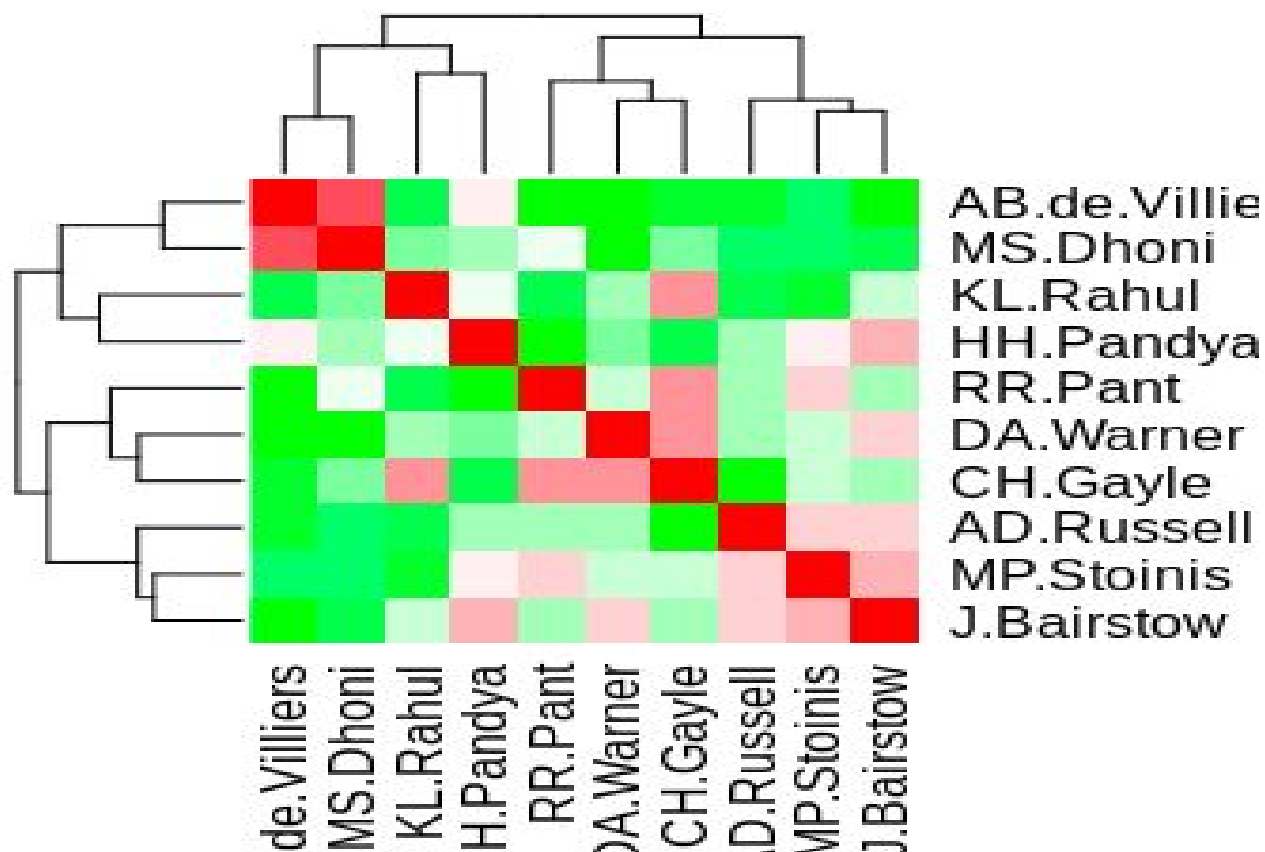


**(Below is the code to create correlation matrix for top 10 batsmen scores and heatmap of that data)**

```
#FINDING Correlation coefficient for all the player combinations and heatmap for it  
data_for_covariance=read.table("/home/kapil/desktop/study material notes/6TH SEM/Data Analysis/DA  
LAB/lab2/top10player_cv.csv", sep = ",", header = TRUE)  
data_for_covariance  
#cor(data_for_covariance$DA.Warner,data_for_covariance$DA.Warner,method="spearman")  
k=cor(data_for_covariance[sapply(data_for_covariance, is.numeric)])  
k  
palette = colorRampPalette(c("green","white","red")) (20)  
heatmap(k,col=palette,symm = TRUE)
```

## ★ Heatmap for correlation coefficient of top 10 batsmens scores

```
> palette = colorRampPalette(c("green","white","red")) (20)
> heatmap(k,col=palette,symm = TRUE)
```



\*\*\*\*\*//\*\*\*\*\*