Dplyr exercise

In [1]:
```
library("dplyr")
```

In [2]:
```
df <- read.csv("deliveries.csv")
```

In [3]:
```
head(df)
```

match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	6	S Dhawan	DA Warner	TS Mills	

Q. Total wickets by Y Chahal

In [4]:
```
distinct(df,bowler)
```

bowler
TS Mills
A Choudhary
YS Chahal
S Aravind
SR Watson
TM Head
STR Binny
A Nehra
B Kumar
BCJ Cutting
Rashid Khan
DJ Hooda
MC Henriques
Bipul Sharma
AB Dinda
DL Chahar
BA Stokes
Imran Tahir
A Zampa
R Bhatia
TG Southee
HH Pandya
MJ McClenaghan
JJ Bumrah
KH Pandya
KA Pollard
TA Boult
PP Chawla
SP Narine
CR Woakes
...
MG Neser
AC Gilchrist
MA Starc

bowler
JDS Neesham
M Vijay
SA Yadav
Shivam Sharma
V Shankar
LMP Simmons
K Santokie
S Gopal
BE Hendricks
JW Hastings
Karanveer Singh
DJ Muthuswami
SA Abbott
J Suchith
RG More
D Wiese
GS Sandhu
Gurkeerat Singh
M Ashwin
C Munro
P Sahu
KJ Abbott
T Shamsi
SM Boland
Sachin Baby
N Rana
KS Williamson

```
In [5]: chahal <- filter(df,bowler=="YS Chahal")
```

In [6]:
```
head(chahal)
```

match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	1	MC Henriques	S Dhawan	YS Chahal	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	2	MC Henriques	S Dhawan	YS Chahal	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	3	S Dhawan	MC Henriques	YS Chahal	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	4	S Dhawan	MC Henriques	YS Chahal	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	5	MC Henriques	S Dhawan	YS Chahal	
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	6	S Dhawan	MC Henriques	YS Chahal	

In [7]:
```
count(chahal,dismissal_kind)
```

dismissal_kind	n
	1147
bowled	9
caught	49
caught and bowled	1
lbw	3
run out	2
stumped	8

Q. Total runs of V Kohli

In [8]:
```
df %>% filter(batsman=="V Kohli") %>% summarize(total_runs=sum(batsman_runs))
```

total_runs
4423

Q. Top 10 batsman (by total runs)

```
df %>% group_by(batsman) %>% summarize(total_runs=sum(batsman_runs)) %>%
arrange(desc(total_runs)) %>% slice(1:10)
```

Q. Highest and Lowest score in an inning

In [10]:

```
x <- df %>% group_by(match_id,inning) %>% summarize(total_runs=sum(total_runs))
```

In [11]:

```
max(x["total_runs"])
```

263

In [12]:

```
min(x["total_runs"])
```

2

Q. V Kohli average runs

```
df %>% filter(batsman=="V Kohli") %>% group_by(match_id) %>%
summarize(runs=sum(batsman_runs)) %>% summarize(avg_runs=mean(runs))
```

Q. What is the probablity of winning a match when a team scores above 200 in first inning ?

In [16]:

```
match = df %>% group_by(match_id,inning) %>% summarize(runs=sum(total_runs))
```

In [18]:

```
inning1 = match %>% filter(inning == 1) %>% filter(runs >= 200)
```

In [19]:

```
inning2 = match %>% filter(inning == 2)
```

In [20]:

```
head(inning1)
```

match_id	inning	runs
1	1	207
9	1	205
20	1	213
32	1	207
36	1	209
41	1	208

In [21]:
```
head(inning2)
```

match_id	inning	runs
1	2	172
2	2	187
3	2	184
4	2	164
5	2	142
6	2	140

In [27]:
```
results = inner_join(inning1,inning2,by="match_id")
```

In [31]:
```
dim(results)
```

```
49  5
```

In [32]:
```
total_matches = 49
```

In [33]:
```
head(results)
```

match_id	inning.x	runs.x	inning.y	runs.y
1	1	207	2	172
9	1	205	2	108
20	1	213	2	192
32	1	207	2	181
36	1	209	2	161
41	1	208	2	214

In [36]:
```
runs_diff = results$runs.x - results$runs.y
```

In [42]:
```
winning_matches = length(runs_diff[runs_diff > 0])
```

In [43]:
```
winning_matches / total_matches
```

```
0.857142857142857
```