21BDS0340

Abhinav Dinesh Srivatsa

Exploratory Data Analysis Lab

Practice Problem Set – 2

Experiment 3

Code:

```
library(dplyr)
library(tidyr)
library(magrittr)
setwd("/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Assignment 2")
data = read.csv("DS2_Match.csv")
```

Output:

```
> setwd("/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Assignment 2")
> data = read.csv("DS2_Match.csv")
> library(dplyr)
> library(tidyr)
> library(magrittr)
> setwd("/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Assignment 2")
> data = read.csv("DS2_Match.csv")
```

Code:

```
win_type_arranged = data %>% arrange(Win_Type)
View(win_type_arranged)
```

Output:

win_type_arranged

Filter

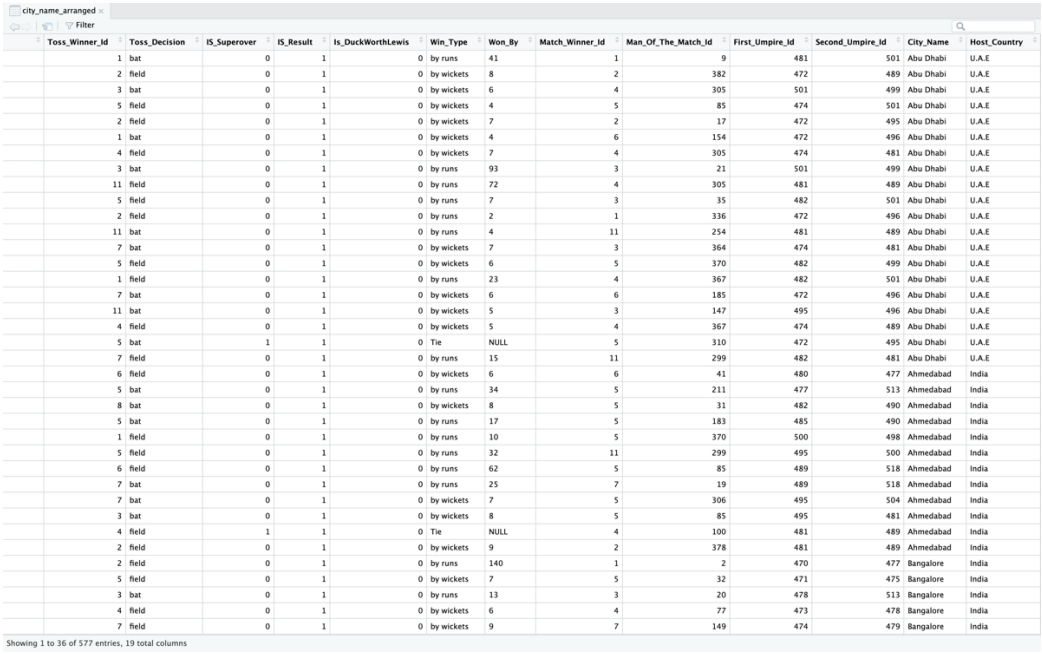
Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id	Man_Of_The_Match_Id	First_Umpire_Id	Second_Umpire_Id	City_Name	Host_Country
6	bat	0	0	0	No Result	NULL	NA	NA	485	518	Delhi	India
5	field	0	0	0	No Result	NULL	NA	NA	494	506	Bangalore	India
2	field	0	0	0	No Result	NULL	NA	NA	482	520	Bangalore	India
1	field	1	1	0	Tie	NULL	5	31	471	481	Cape Town	South Africa
3	field	1	1	0	Tie	NULL	4	233	486	476	Chennai	India
2	bat	1	1	0	Tie	NULL	11	340	495	489	Hyderabad	India
2	field	1	1	0	Tie	NULL	2	8	481	496	Bangalore	India
5	bat	1	1	0	Tie	NULL	5	310	472	495	Abu Dhabi	U.A.E
4	field	1	1	0	Tie	NULL	4	100	481	489	Ahmedabad	India
2	field	0	1	0	by runs	140	1	2	470	477	Bangalore	India
3	bat	0	1	0	by runs	33	3	19	471	487	Chandigarh	India
7	field	0	1	0	by runs	6	3	18	476	512	Chennai	India
7	field	0	1	0	by runs	66	4	26	472	492	Chandigarh	India
3	bat	0	1	0	by runs	13	3	20	478	513	Bangalore	India
2	field	0	1	0	by runs	10	6	118	472	514	Delhi	India
5	bat	0	1	0	by runs	45	5	101	477	512	Jaipur	India
4	bat	0	1	0	by runs	9	4	29	476	514	Chandigarh	India
6	field	0	1	0	by runs	29	7	49	475	477	Mumbai	India
1	bat	0	1	0	by runs	5	1	1	470	475	Kolkata	India
4	field	0	1	0	by runs	18	3	151	479	480	Chennai	India
1	bat	0	1	0	by runs	23	1	1	475	492	Hyderabad	India
1	bat	0	1	0	by runs	23	1	144	470	475	Kolkata	India
5	field	0	1	0	by runs	41	4	100	473	486	Chandigarh	India
8	field	0	1	0	by runs	12	6	136	480	512	Delhi	India
6	bat	0	1	1	by runs	6	4	64	479	477	Delhi	India
2	field	0	1	0	by runs	65	5	74	474	487	Jaipur	India
8	field	0	1	0	by runs	25	7	71	478	476	Hyderabad	India
1	bat	0	1	1	by runs	3	3	122	470	486	Kolkata	India
7	field	0	1	0	by runs	1	4	100	474	512	Mumbai	India
2	bat	0	1	0	by runs	14	2	124	476	514	Chennai	India
5	bat	0	1	0	by runs	10	5	109	476	487	Chennai	India
8	field	0	1	0	by runs	3	2	14	478	487	Bangalore	India
6	field	0	1	0	by runs	105	5	32	474	477	Mumbai	India
3	field	0	1	0	by runs	19	7	133	478	486	Cape Town	South Africa
2	bat	0	1	0	by runs	75	2	6	478	513	Cape Town	South Africa
3	bat	0	1	0	by runs	92	3	121	480	490	Port Elizabeth	South Africa
1	field	0	1	1	by runs	11	1	162	476	515	Durban	South Africa

Showing 1 to 36 of 577 entries, 19 total columns

Code:

```
city_name_arranged = data %>% arrange(City_Name)
View(city_name_arranged)
```

Output:



city_name_arranged

Filter

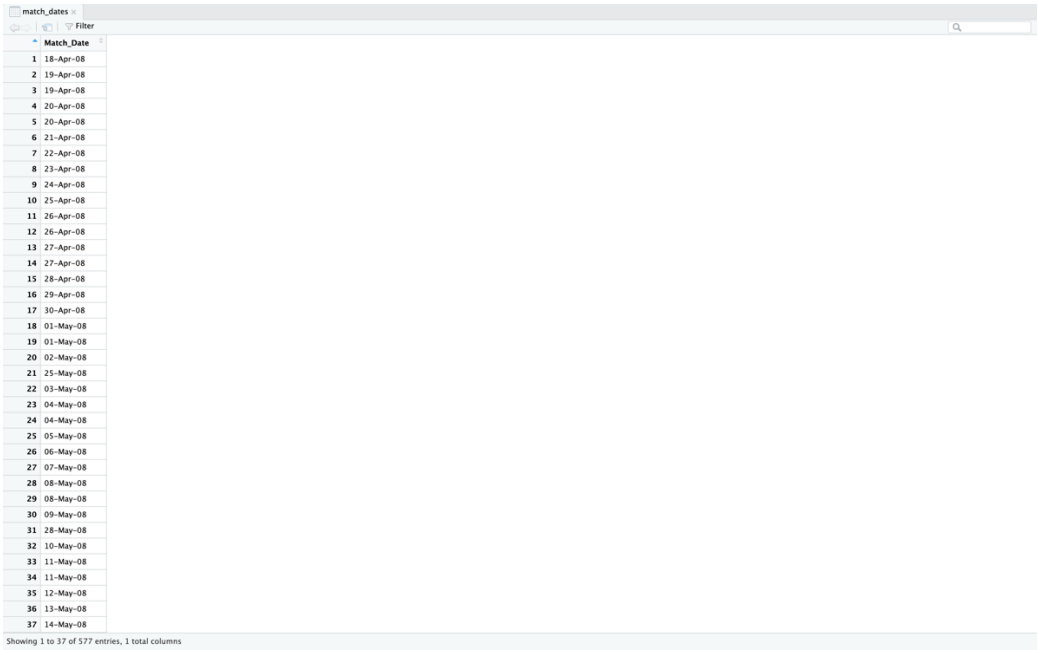
	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id	Man_Of_The_Match_Id	First_Umpire_Id	Second_Umpire_Id	City_Name	Host_Country
	1	bat	0	1	0	by runs	41	1	9	481	501	Abu Dhabi	U.A.E
	2	field	0	1	0	by wickets	8	2	382	472	489	Abu Dhabi	U.A.E
	3	bat	0	1	0	by wickets	6	4	305	501	499	Abu Dhabi	U.A.E
	5	field	0	1	0	by wickets	4	5	85	474	501	Abu Dhabi	U.A.E
	2	field	0	1	0	by wickets	7	2	17	472	495	Abu Dhabi	U.A.E
	1	bat	0	1	0	by wickets	4	6	154	472	496	Abu Dhabi	U.A.E
	4	field	0	1	0	by wickets	7	4	305	474	481	Abu Dhabi	U.A.E
	3	bat	0	1	0	by runs	93	3	21	501	499	Abu Dhabi	U.A.E
	11	field	0	1	0	by runs	72	4	305	481	489	Abu Dhabi	U.A.E
	5	field	0	1	0	by runs	7	3	35	482	501	Abu Dhabi	U.A.E
	2	field	0	1	0	by runs	2	1	336	472	496	Abu Dhabi	U.A.E
	11	bat	0	1	0	by runs	4	11	254	481	489	Abu Dhabi	U.A.E
	7	bat	0	1	0	by wickets	7	3	364	474	481	Abu Dhabi	U.A.E
	5	field	0	1	0	by wickets	6	5	370	482	499	Abu Dhabi	U.A.E
	1	field	0	1	0	by runs	23	4	367	482	501	Abu Dhabi	U.A.E
	7	bat	0	1	0	by wickets	6	6	185	472	496	Abu Dhabi	U.A.E
	11	bat	0	1	0	by wickets	5	3	147	495	496	Abu Dhabi	U.A.E
	4	field	0	1	0	by wickets	5	4	367	474	489	Abu Dhabi	U.A.E
	5	bat	1	1	0	Tie	NULL	5	310	472	495	Abu Dhabi	U.A.E
	7	field	0	1	0	by runs	15	11	299	482	481	Abu Dhabi	U.A.E
	6	field	0	1	0	by wickets	6	6	41	480	477	Ahmedabad	India
	5	bat	0	1	0	by runs	34	5	211	477	513	Ahmedabad	India
	8	bat	0	1	0	by wickets	8	5	31	482	490	Ahmedabad	India
	5	bat	0	1	0	by runs	17	5	183	485	490	Ahmedabad	India
	1	field	0	1	0	by runs	10	5	370	500	498	Ahmedabad	India
	5	field	0	1	0	by runs	32	11	299	495	500	Ahmedabad	India
	6	field	0	1	0	by runs	62	5	85	489	518	Ahmedabad	India
	7	bat	0	1	0	by runs	25	7	19	489	518	Ahmedabad	India
	7	bat	0	1	0	by wickets	7	5	306	495	504	Ahmedabad	India
	3	bat	0	1	0	by wickets	8	5	85	495	481	Ahmedabad	India
	4	field	1	1	0	Tie	NULL	4	100	481	489	Ahmedabad	India
	2	field	0	1	0	by wickets	9	2	378	481	489	Ahmedabad	India
	2	field	0	1	0	by runs	140	1	2	470	477	Bangalore	India
	5	field	0	1	0	by wickets	7	5	32	471	475	Bangalore	India
	3	bat	0	1	0	by runs	13	3	20	478	513	Bangalore	India
	4	field	0	1	0	by wickets	6	4	77	473	478	Bangalore	India
	7	field	0	1	0	by wickets	9	7	149	474	479	Bangalore	India

Showing 1 to 36 of 577 entries, 19 total columns

Code:

```
match_dates = data %>% select(Match_Date)
View(match_dates)
```

Output:



match_dates

Filter

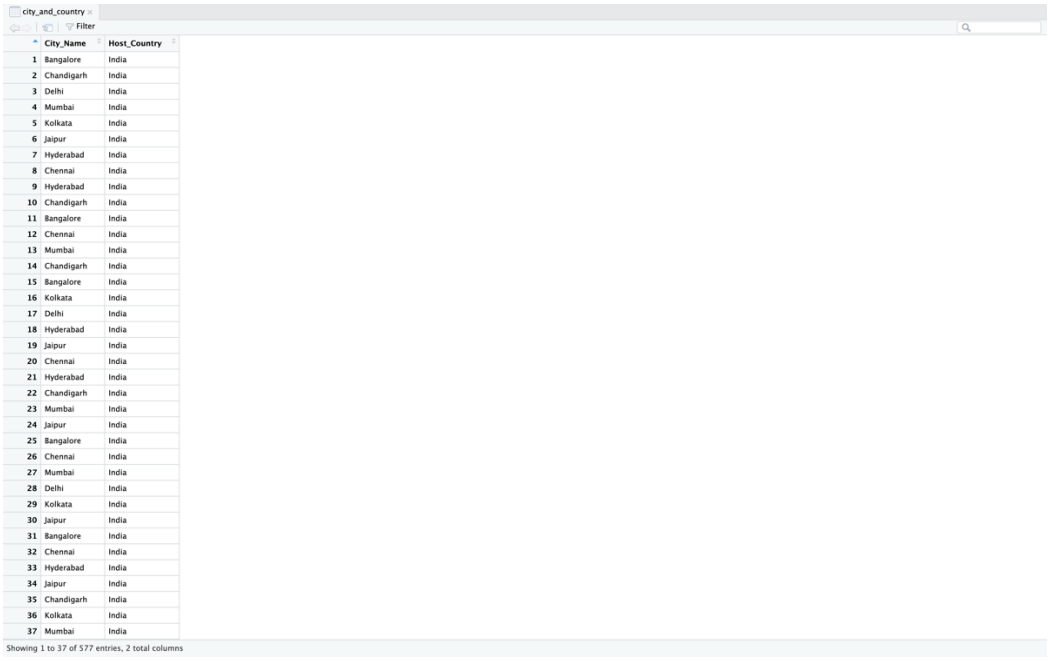
	Match_Date
1	18-Apr-08
2	19-Apr-08
3	19-Apr-08
4	20-Apr-08
5	20-Apr-08
6	21-Apr-08
7	22-Apr-08
8	23-Apr-08
9	24-Apr-08
10	25-Apr-08
11	26-Apr-08
12	26-Apr-08
13	27-Apr-08
14	27-Apr-08
15	28-Apr-08
16	29-Apr-08
17	30-Apr-08
18	01-May-08
19	01-May-08
20	02-May-08
21	25-May-08
22	03-May-08
23	04-May-08
24	04-May-08
25	05-May-08
26	06-May-08
27	07-May-08
28	08-May-08
29	08-May-08
30	09-May-08
31	28-May-08
32	10-May-08
33	11-May-08
34	11-May-08
35	12-May-08
36	13-May-08
37	14-May-08

Showing 1 to 37 of 577 entries, 1 total columns

Code:

```
city_and_country = data %>% select(City_Name, Host_Country)
View(city_and_country)
```

Output:



city_and_country

Filter

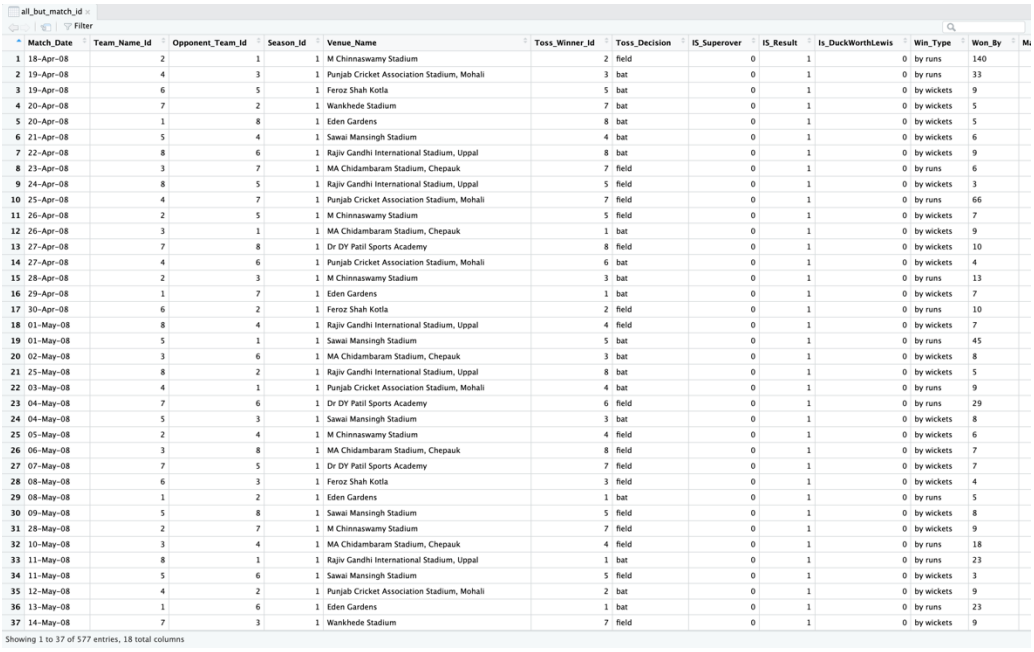
	City_Name	Host_Country
1	Bangalore	India
2	Chandigarh	India
3	Delhi	India
4	Mumbai	India
5	Kolkata	India
6	Jaipur	India
7	Hyderabad	India
8	Chennai	India
9	Hyderabad	India
10	Chandigarh	India
11	Bangalore	India
12	Chennai	India
13	Mumbai	India
14	Chandigarh	India
15	Bangalore	India
16	Kolkata	India
17	Delhi	India
18	Hyderabad	India
19	Jaipur	India
20	Chennai	India
21	Hyderabad	India
22	Chandigarh	India
23	Mumbai	India
24	Jaipur	India
25	Bangalore	India
26	Chennai	India
27	Mumbai	India
28	Delhi	India
29	Kolkata	India
30	Jaipur	India
31	Bangalore	India
32	Chennai	India
33	Hyderabad	India
34	Jaipur	India
35	Chandigarh	India
36	Kolkata	India
37	Mumbai	India

Showing 1 to 37 of 577 entries, 2 total columns

Code:

```
all_but_match_id = data %>% select(-Match_Id)
View(all_but_match_id)
```

Output:



all_but_match_id

Filter

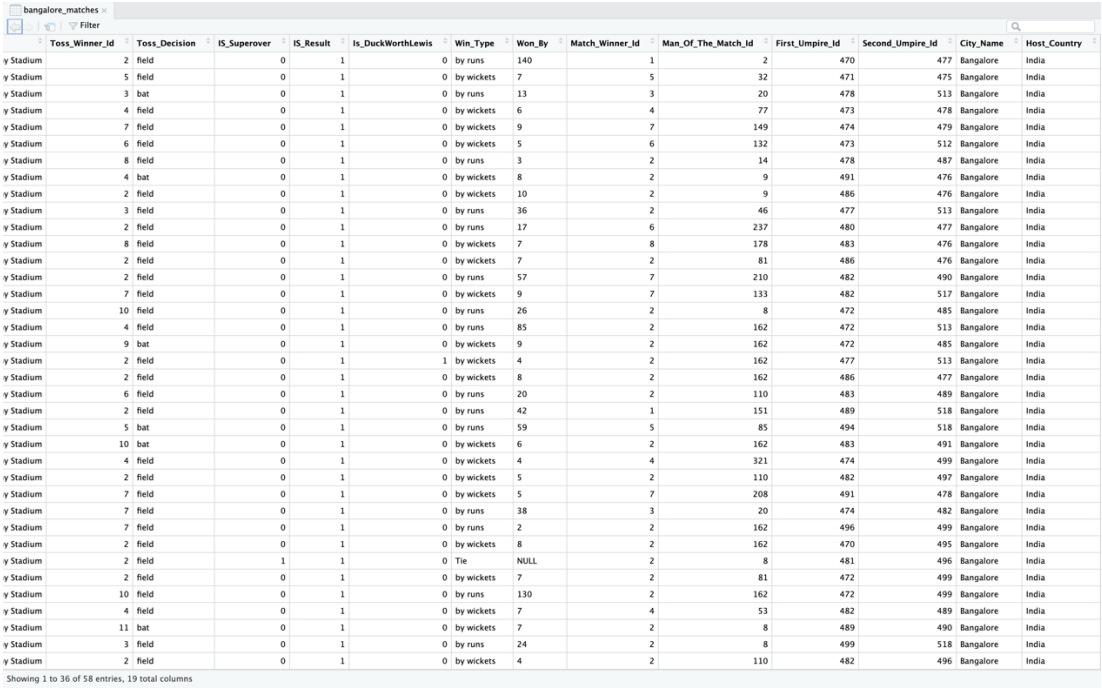
	Match_Date	Team_Name_Id	Opponent_Team_Id	Season_Id	Venue_Name	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Ma
1	18-Apr-08	2	1	1	M Chinnaswamy Stadium	2	field	0	1	0	by runs	140	
2	19-Apr-08	4	3	1	Punjab Cricket Association Stadium, Mohali	3	bat	0	1	0	by runs	33	
3	19-Apr-08	6	5	1	Feroz Shah Kotla	5	bat	0	1	0	by wickets	9	
4	20-Apr-08	7	2	1	Wankhede Stadium	7	bat	0	1	0	by wickets	5	
5	20-Apr-08	1	8	1	Eden Gardens	8	bat	0	1	0	by wickets	5	
6	21-Apr-08	5	4	1	Sawai Mansingh Stadium	4	bat	0	1	0	by wickets	6	
7	22-Apr-08	8	6	1	Rajiv Gandhi International Stadium, Uppal	8	bat	0	1	0	by wickets	9	
8	23-Apr-08	3	7	1	MA Chidambaram Stadium, Chepauk	7	field	0	1	0	by runs	6	
9	24-Apr-08	8	5	1	Rajiv Gandhi International Stadium, Uppal	5	field	0	1	0	by wickets	3	
10	25-Apr-08	4	7	1	Punjab Cricket Association Stadium, Mohali	7	field	0	1	0	by runs	66	
11	26-Apr-08	2	5	1	M Chinnaswamy Stadium	5	field	0	1	0	by wickets	7	
12	26-Apr-08	3	1	1	MA Chidambaram Stadium, Chepauk	1	bat	0	1	0	by wickets	9	
13	27-Apr-08	7	8	1	Dr DY Patil Sports Academy	8	field	0	1	0	by wickets	10	
14	27-Apr-08	4	6	1	Punjab Cricket Association Stadium, Mohali	6	bat	0	1	0	by wickets	4	
15	28-Apr-08	2	3	1	M Chinnaswamy Stadium	3	bat	0	1	0	by runs	13	
16	29-Apr-08	1	7	1	Eden Gardens	1	bat	0	1	0	by wickets	7	
17	30-Apr-08	6	2	1	Feroz Shah Kotla	2	field	0	1	0	by runs	10	
18	01-May-08	8	4	1	Rajiv Gandhi International Stadium, Uppal	4	field	0	1	0	by wickets	7	
19	01-May-08	5	1	1	Sawai Mansingh Stadium	5	bat	0	1	0	by runs	45	
20	02-May-08	3	6	1	MA Chidambaram Stadium, Chepauk	3	bat	0	1	0	by wickets	8	
21	25-May-08	8	2	1	Rajiv Gandhi International Stadium, Uppal	8	bat	0	1	0	by wickets	5	
22	03-May-08	4	1	1	Punjab Cricket Association Stadium, Mohali	4	bat	0	1	0	by runs	9	
23	04-May-08	7	6	1	Dr DY Patil Sports Academy	6	field	0	1	0	by runs	29	
24	04-May-08	5	3	1	Sawai Mansingh Stadium	3	bat	0	1	0	by wickets	8	
25	05-May-08	2	4	1	M Chinnaswamy Stadium	4	field	0	1	0	by wickets	6	
26	06-May-08	3	8	1	MA Chidambaram Stadium, Chepauk	8	field	0	1	0	by wickets	7	
27	07-May-08	7	5	1	Dr DY Patil Sports Academy	7	field	0	1	0	by wickets	7	
28	08-May-08	6	3	1	Feroz Shah Kotla	3	field	0	1	0	by wickets	4	
29	08-May-08	1	2	1	Eden Gardens	1	bat	0	1	0	by runs	5	
30	09-May-08	5	8	1	Sawai Mansingh Stadium	5	field	0	1	0	by wickets	8	
31	28-May-08	2	7	1	M Chinnaswamy Stadium	7	field	0	1	0	by wickets	9	
32	10-May-08	3	4	1	MA Chidambaram Stadium, Chepauk	4	field	0	1	0	by runs	18	
33	11-May-08	8	1	1	Rajiv Gandhi International Stadium, Uppal	1	bat	0	1	0	by runs	23	
34	11-May-08	5	6	1	Sawai Mansingh Stadium	5	field	0	1	0	by wickets	3	
35	12-May-08	4	2	1	Punjab Cricket Association Stadium, Mohali	2	bat	0	1	0	by wickets	9	
36	13-May-08	1	6	1	Eden Gardens	1	bat	0	1	0	by runs	23	
37	14-May-08	7	3	1	Wankhede Stadium	7	field	0	1	0	by wickets	9	

Showing 1 to 37 of 577 entries, 18 total columns

Code:

```
bangalore_matches = data %>% filter(City_Name == "Bangalore")
View(bangalore_matches)
```

Output:



bangalore_matches

Filter

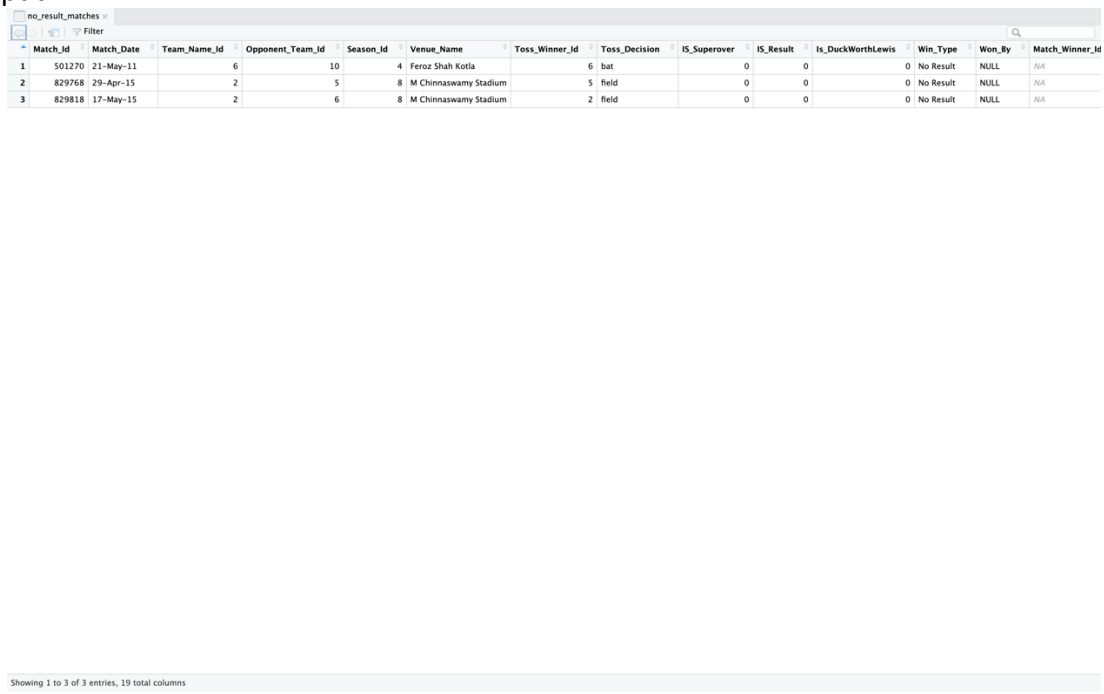
	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id	Man_Of_The_Match_Id	First_Umpire_Id	Second_Umpire_Id	City_Name	Host_Country
y Stadium	2	field	0	1	0	by runs	140	1	2	470	477	Bangalore	India
y Stadium	5	field	0	1	0	by wickets	7	5	32	471	475	Bangalore	India
y Stadium	3	bat	0	1	0	by runs	13	3	20	478	513	Bangalore	India
y Stadium	4	field	0	1	0	by wickets	6	4	77	473	478	Bangalore	India
y Stadium	7	field	0	1	0	by wickets	9	7	149	474	479	Bangalore	India
y Stadium	6	field	0	1	0	by wickets	5	6	132	473	512	Bangalore	India
y Stadium	8	field	0	1	0	by runs	3	2	14	478	487	Bangalore	India
y Stadium	4	bat	0	1	0	by wickets	8	2	9	491	476	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	10	2	9	486	476	Bangalore	India
y Stadium	3	field	0	1	0	by runs	36	2	46	477	513	Bangalore	India
y Stadium	2	field	0	1	0	by runs	17	6	237	480	477	Bangalore	India
y Stadium	8	field	0	1	0	by wickets	7	8	178	483	476	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	7	2	81	486	476	Bangalore	India
y Stadium	2	field	0	1	0	by runs	57	7	210	482	490	Bangalore	India
y Stadium	7	field	0	1	0	by wickets	9	7	133	482	517	Bangalore	India
y Stadium	10	field	0	1	0	by runs	26	2	8	472	485	Bangalore	India
y Stadium	4	field	0	1	0	by runs	85	2	162	472	513	Bangalore	India
y Stadium	9	bat	0	1	0	by wickets	9	2	162	472	485	Bangalore	India
y Stadium	2	field	0	1	1	by wickets	4	2	162	477	513	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	8	2	162	486	477	Bangalore	India
y Stadium	6	field	0	1	0	by runs	20	2	110	483	489	Bangalore	India
y Stadium	2	field	0	1	0	by runs	42	1	151	489	518	Bangalore	India
y Stadium	5	bat	0	1	0	by runs	59	5	85	494	518	Bangalore	India
y Stadium	10	bat	0	1	0	by wickets	6	2	162	483	491	Bangalore	India
y Stadium	4	field	0	1	0	by wickets	4	4	321	474	499	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	5	2	110	482	497	Bangalore	India
y Stadium	7	field	0	1	0	by wickets	5	7	208	491	478	Bangalore	India
y Stadium	7	field	0	1	0	by runs	38	3	20	474	482	Bangalore	India
y Stadium	7	field	0	1	0	by runs	2	2	162	496	499	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	8	2	162	470	495	Bangalore	India
y Stadium	2	field	1	1	0	Tie	NULL	2	8	481	496	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	7	2	81	472	499	Bangalore	India
y Stadium	10	field	0	1	0	by runs	130	2	162	472	499	Bangalore	India
y Stadium	4	field	0	1	0	by wickets	7	4	53	482	489	Bangalore	India
y Stadium	11	bat	0	1	0	by wickets	7	2	8	489	490	Bangalore	India
y Stadium	3	field	0	1	0	by runs	24	2	8	499	518	Bangalore	India
y Stadium	2	field	0	1	0	by wickets	4	2	110	482	496	Bangalore	India

Showing 1 to 36 of 58 entries, 19 total columns

Code:

```
no_result_matches = data %>% filter(Win_Type == "No Result")
View(no_result_matches)
```

Output:



no_result_matches

Filter

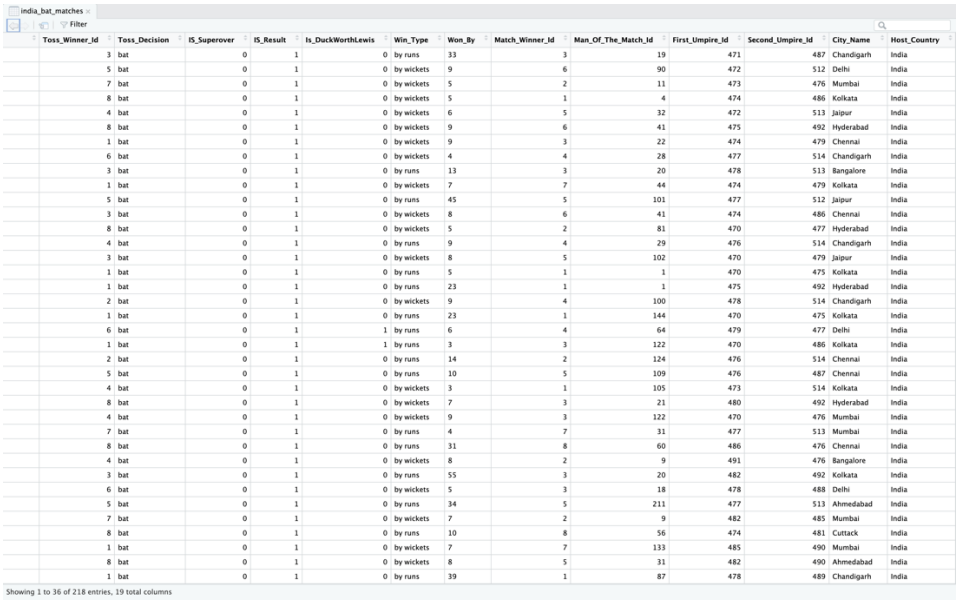
	Match_Id	Match_Date	Team_Name_Id	Opponent_Team_Id	Season_Id	Venue_Name	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id
1	501270	21-May-11	6	10	4	Feroz Shah Kotla	6	bat	0	0	0	No Result	NULL	NA
2	829768	29-Apr-15	2	5	8	M Chinnaswamy Stadium	5	field	0	0	0	No Result	NULL	NA
3	829818	17-May-15	2	6	8	M Chinnaswamy Stadium	2	field	0	0	0	No Result	NULL	NA

Showing 1 to 3 of 3 entries, 19 total columns

Code:

```
india_bat_matches = data %>% filter(Host_Country == "India", Toss_Decision == "bat")
View(india_bat_matches)
```

Output:



india_bat_matches

Filter

	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id	Man_Of_The_Match_Id	First_Umpire_Id	Second_Umpire_Id	City_Name	Host_Country
	3	bat	0	1	0	by runs	33	3	19	471	487	Chandigarh	India
	5	bat	0	1	0	by wickets	9	6	90	472	512	Delhi	India
	7	bat	0	1	0	by wickets	5	2	11	473	476	Mumbai	India
	8	bat	0	1	0	by wickets	5	1	4	474	486	Kolkata	India
	4	bat	0	1	0	by wickets	6	5	32	472	513	Jaipur	India
	8	bat	0	1	0	by wickets	9	6	41	475	492	Hyderabad	India
	1	bat	0	1	0	by wickets	9	3	22	474	479	Chennai	India
	6	bat	0	1	0	by wickets	4	4	28	477	514	Chandigarh	India
	3	bat	0	1	0	by runs	13	3	20	478	513	Bangalore	India
	1	bat	0	1	0	by wickets	7	7	44	474	479	Kolkata	India
	5	bat	0	1	0	by runs	45	5	101	477	512	Jaipur	India
	3	bat	0	1	0	by wickets	8	6	41	474	486	Chennai	India
	8	bat	0	1	0	by wickets	5	2	81	470	477	Hyderabad	India
	4	bat	0	1	0	by runs	9	4	29	476	514	Chandigarh	India
	3	bat	0	1	0	by wickets	8	5	102	470	479	Jaipur	India
	1	bat	0	1	0	by runs	5	1	1	470	475	Kolkata	India
	1	bat	0	1	0	by runs	23	1	1	475	492	Hyderabad	India
	2	bat	0	1	0	by wickets	9	4	100	478	514	Chandigarh	India
	1	bat	0	1	0	by runs	23	1	144	470	475	Kolkata	India
	6	bat	0	1	1	by runs	6	4	64	479	477	Delhi	India
	1	bat	0	1	1	by runs	3	3	122	470	486	Kolkata	India
	2	bat	0	1	0	by runs	14	2	124	476	514	Chennai	India
	5	bat	0	1	0	by runs	10	5	109	476	487	Chennai	India
	4	bat	0	1	0	by wickets	3	1	105	473	514	Kolkata	India
	8	bat	0	1	0	by wickets	7	3	21	480	492	Hyderabad	India
	4	bat	0	1	0	by wickets	9	3	122	470	476	Mumbai	India
	7	bat	0	1	0	by runs	4	7	31	477	513	Mumbai	India
	8	bat	0	1	0	by runs	31	8	60	486	476	Chennai	India
	4	bat	0	1	0	by wickets	8	2	9	491	476	Bangalore	India
	3	bat	0	1	0	by runs	55	3	20	482	492	Kolkata	India
	6	bat	0	1	0	by wickets	5	3	18	478	488	Delhi	India
	5	bat	0	1	0	by runs	34	5	211	477	513	Ahmedabad	India
	7	bat	0	1	0	by wickets	7	2	9	482	485	Mumbai	India
	8	bat	0	1	0	by runs	10	8	56	474	481	Cuttack	India
	1	bat	0	1	0	by wickets	7	7	133	485	490	Mumbai	India
	8	bat	0	1	0	by wickets	8	5	31	482	490	Ahmedabad	India
	1	bat	0	1	0	by runs	39	1	87	478	489	Chandigarh	India

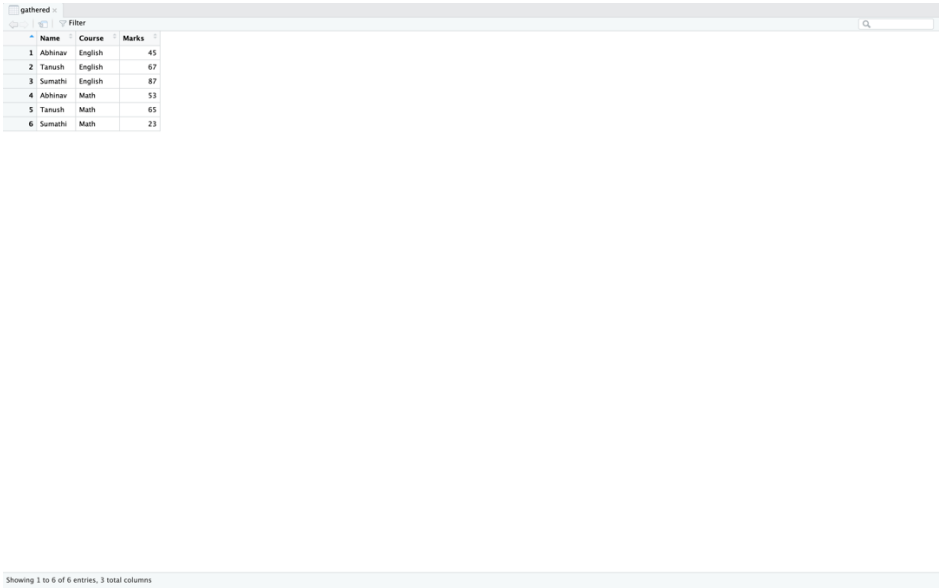
Showing 1 to 36 of 218 entries, 19 total columns

Code:

```
custom_data = data.frame(
  Name = c("Abhinav", "Tanush", "Sumathi"),
  English = c(45, 67, 87),
  Math = c(53, 65, 23)
)

gathered = custom_data %>% gather("Course", "Marks", 2:3)
View(gathered)
```

Output:



gathered

Filter

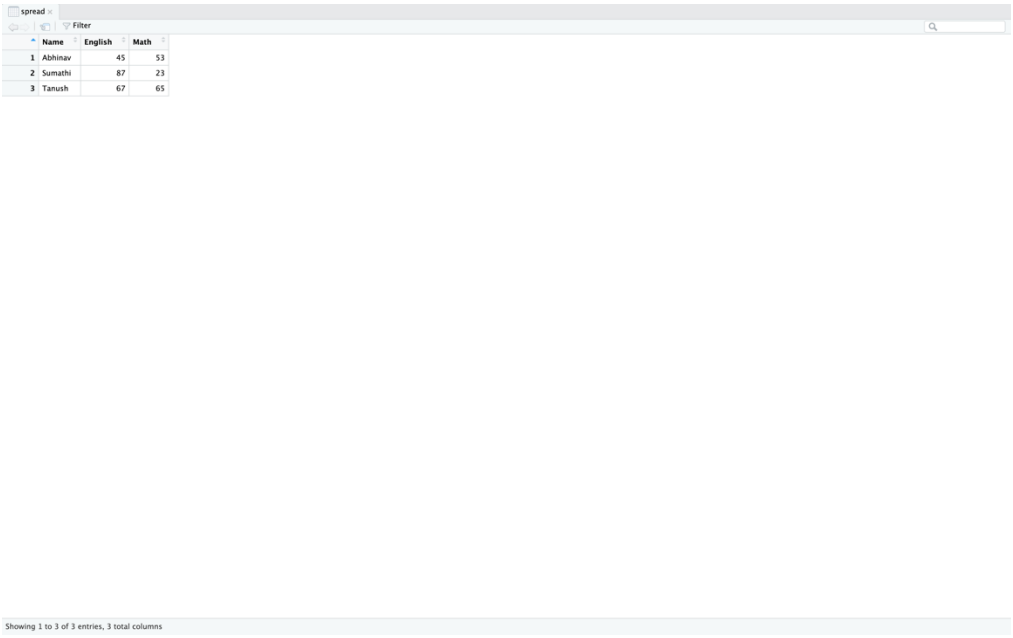
	Name	Course	Marks
1	Abhinav	English	45
2	Tanush	English	67
3	Sumathi	English	87
4	Abhinav	Math	53
5	Tanush	Math	65
6	Sumathi	Math	23

Showing 1 to 6 of 6 entries, 3 total columns

Code:

```
spread = gathered %>% spread("Course", "Marks")
View(spread)
```

Output:



spread

Filter

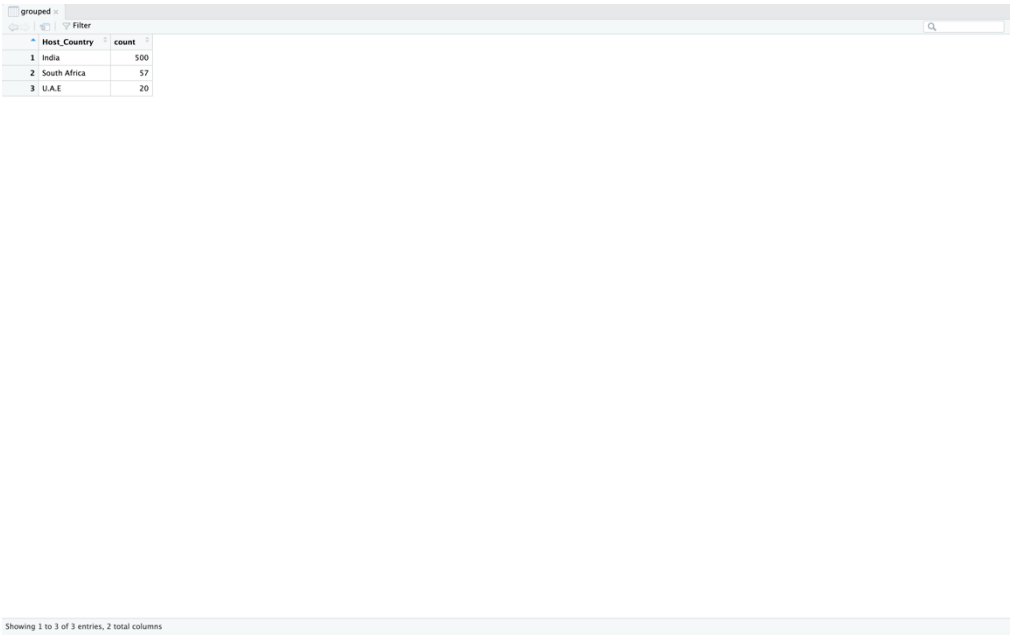
	Name	English	Math
1	Abhinav	45	53
2	Sumathi	87	23
3	Tanush	67	65

Showing 1 to 3 of 3 entries, 3 total columns

Code:

```
grouped = data %>% group_by(Host_Country) %>%
  summarise(count = n())
View(grouped)
```

Output:



grouped

Filter

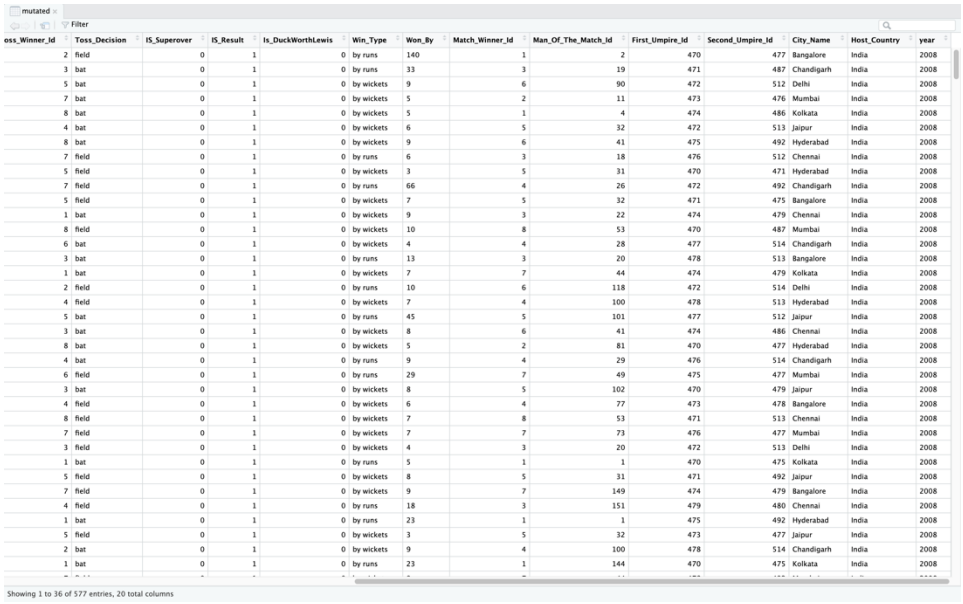
	Host_Country	count
1	India	500
2	South Africa	57
3	U.A.E	20

Showing 1 to 3 of 3 entries, 2 total columns

Code:

```
mutated = data %>% mutate(year=format(as.Date(data$Match_Date, format="%d-%b-%y"), "%Y"))
View(mutated)
```

Output:



mutated

Filter

oss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type	Won_By	Match_Winner_Id	Man_Of_The_Match_Id	First_Umpire_Id	Second_Umpire_Id	City_Name	Host_Country	year
2	field	0	1	0	by runs	140	1	2	470	477	Bangalore	India	2008
3	bat	0	1	0	by runs	33	3	19	471	487	Chandigarh	India	2008
5	bat	0	1	0	by wickets	9	6	90	472	512	Delhi	India	2008
7	bat	0	1	0	by wickets	5	2	11	473	476	Mumbai	India	2008
8	bat	0	1	0	by wickets	5	1	4	474	486	Kolkata	India	2008
4	bat	0	1	0	by wickets	6	5	32	472	513	Jaipur	India	2008
8	bat	0	1	0	by wickets	9	6	41	475	492	Hyderabad	India	2008
7	field	0	1	0	by runs	6	3	18	476	512	Chennai	India	2008
5	field	0	1	0	by wickets	3	5	31	470	471	Hyderabad	India	2008
7	field	0	1	0	by runs	66	4	26	472	492	Chandigarh	India	2008
5	field	0	1	0	by wickets	7	5	32	471	475	Bangalore	India	2008
1	bat	0	1	0	by wickets	9	3	22	474	479	Chennai	India	2008
8	field	0	1	0	by wickets	10	8	53	470	487	Mumbai	India	2008
6	bat	0	1	0	by wickets	4	4	28	477	514	Chandigarh	India	2008
3	bat	0	1	0	by runs	13	3	20	478	513	Bangalore	India	2008
1	bat	0	1	0	by wickets	7	7	44	474	479	Kolkata	India	2008
2	field	0	1	0	by runs	10	6	118	472	514	Delhi	India	2008
4	field	0	1	0	by wickets	7	4	100	478	513	Hyderabad	India	2008
5	bat	0	1	0	by runs	45	5	101	477	512	Jaipur	India	2008
3	bat	0	1	0	by wickets	8	6	41	474	486	Chennai	India	2008
8	bat	0	1	0	by wickets	5	2	81	470	477	Hyderabad	India	2008
4	bat	0	1	0	by runs	9	4	29	476	514	Chandigarh	India	2008
6	field	0	1	0	by runs	29	7	49	475	477	Mumbai	India	2008
3	bat	0	1	0	by wickets	8	5	102	470	479	Jaipur	India	2008
4	field	0	1	0	by wickets	6	4	77	473	478	Bangalore	India	2008
8	field	0	1	0	by wickets	7	8	53	471	513	Chennai	India	2008
7	field	0	1	0	by wickets	7	7	73	476	477	Mumbai	India	2008
3	field	0	1	0	by wickets	4	3	20	472	513	Delhi	India	2008
1	bat	0	1	0	by runs	5	1	1	470	475	Kolkata	India	2008
5	field	0	1	0	by wickets	8	5	31	471	492	Jaipur	India	2008
7	field	0	1	0	by wickets	9	7	149	474	479	Bangalore	India	2008
4	field	0	1	0	by runs	18	3	151	479	480	Chennai	India	2008
1	bat	0	1	0	by runs	23	1	1	475	492	Hyderabad	India	2008
5	field	0	1	0	by wickets	3	5	32	473	477	Jaipur	India	2008
2	bat	0	1	0	by wickets	9	4	100	478	514	Chandigarh	India	2008
1	bat	0	1	0	by runs	23	1	144	470	475	Kolkata	India	2008

Showing 1 to 36 of 577 entries, 20 total columns

Experiment 4

Code:

```
library(dplyr)
library(ggplot2)
library(lubridate)
setwd("/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Assignment 2")
data = read.csv("./DS2_Match.csv")
```

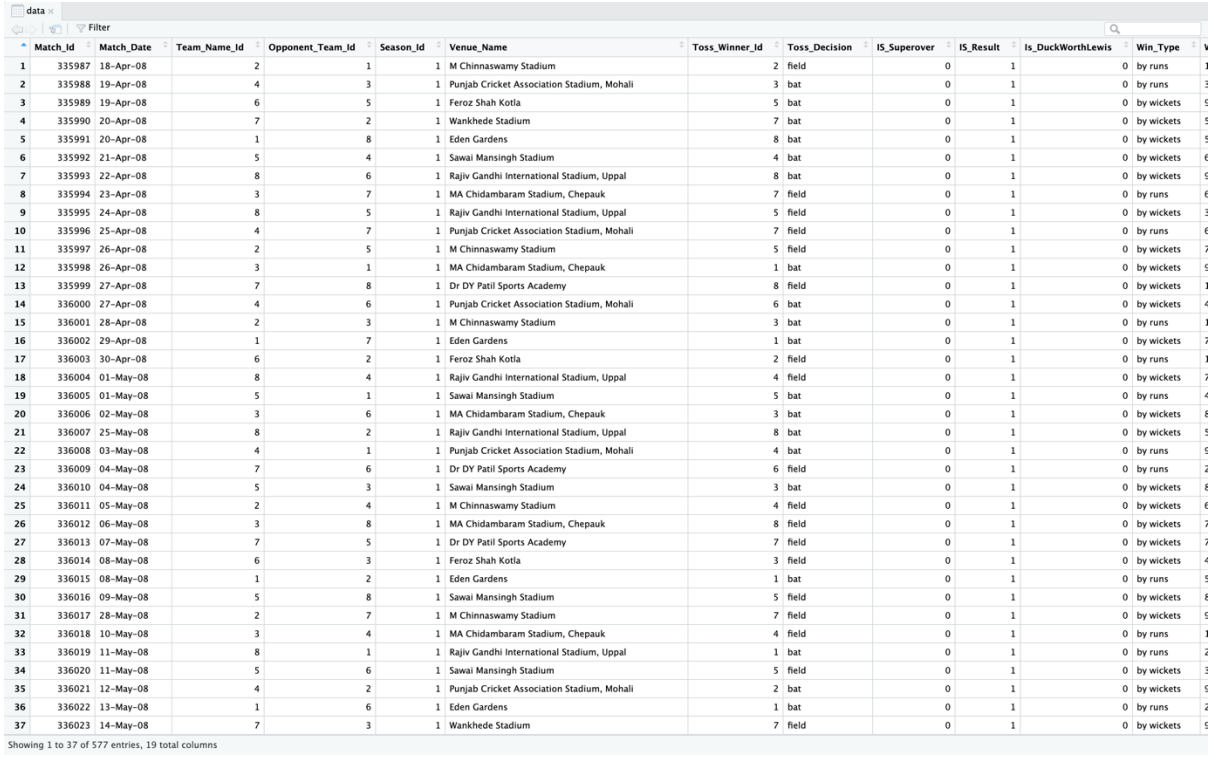
Output:

```
> library(dplyr)
> library(ggplot2)
> library(lubridate)
> setwd("/Users/abhi/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis Lab/Assignment 2")
> data = read.csv("./DS2_Match.csv")
```

Code:

```
# viewing data
View(data)
```

Output:



data

Filter

	Match_Id	Match_Date	Team_Name_Id	Opponent_Team_Id	Season_Id	Venue_Name	Toss_Winner_Id	Toss_Decision	IS_Superover	IS_Result	Is_DuckWorthLewis	Win_Type
1	335987	18-Apr-08	2	1	1	M Chinnaswamy Stadium	2	field	0	1	0	by runs
2	335988	19-Apr-08	4	3	1	Punjab Cricket Association Stadium, Mohali	3	bat	0	1	0	by runs
3	335989	19-Apr-08	6	5	1	Feroz Shah Kotla	5	bat	0	1	0	by wickets
4	335990	20-Apr-08	7	2	1	Wankhede Stadium	7	bat	0	1	0	by wickets
5	335991	20-Apr-08	1	8	1	Eden Gardens	8	bat	0	1	0	by wickets
6	335992	21-Apr-08	5	4	1	Sawai Mansingh Stadium	4	bat	0	1	0	by wickets
7	335993	22-Apr-08	8	6	1	Rajiv Gandhi International Stadium, Uppal	8	bat	0	1	0	by wickets
8	335994	23-Apr-08	3	7	1	MA Chidambaram Stadium, Chepauk	7	field	0	1	0	by runs
9	335995	24-Apr-08	8	5	1	Rajiv Gandhi International Stadium, Uppal	5	field	0	1	0	by wickets
10	335996	25-Apr-08	4	7	1	Punjab Cricket Association Stadium, Mohali	7	field	0	1	0	by runs
11	335997	26-Apr-08	2	5	1	M Chinnaswamy Stadium	5	field	0	1	0	by wickets
12	335998	26-Apr-08	3	1	1	MA Chidambaram Stadium, Chepauk	1	bat	0	1	0	by wickets
13	335999	27-Apr-08	7	8	1	Dr DY Patil Sports Academy	8	field	0	1	0	by wickets
14	336000	27-Apr-08	4	6	1	Punjab Cricket Association Stadium, Mohali	6	bat	0	1	0	by wickets
15	336001	28-Apr-08	2	3	1	M Chinnaswamy Stadium	3	bat	0	1	0	by runs
16	336002	29-Apr-08	1	7	1	Eden Gardens	1	bat	0	1	0	by wickets
17	336003	30-Apr-08	6	2	1	Feroz Shah Kotla	2	field	0	1	0	by runs
18	336004	01-May-08	8	4	1	Rajiv Gandhi International Stadium, Uppal	4	field	0	1	0	by wickets
19	336005	01-May-08	5	1	1	Sawai Mansingh Stadium	5	bat	0	1	0	by runs
20	336006	02-May-08	3	6	1	MA Chidambaram Stadium, Chepauk	3	bat	0	1	0	by wickets
21	336007	25-May-08	8	2	1	Rajiv Gandhi International Stadium, Uppal	8	bat	0	1	0	by wickets
22	336008	03-May-08	4	1	1	Punjab Cricket Association Stadium, Mohali	4	bat	0	1	0	by runs
23	336009	04-May-08	7	6	1	Dr DY Patil Sports Academy	6	field	0	1	0	by runs
24	336010	04-May-08	5	3	1	Sawai Mansingh Stadium	3	bat	0	1	0	by wickets
25	336011	05-May-08	2	4	1	M Chinnaswamy Stadium	4	field	0	1	0	by wickets
26	336012	06-May-08	3	8	1	MA Chidambaram Stadium, Chepauk	8	field	0	1	0	by wickets
27	336013	07-May-08	7	5	1	Dr DY Patil Sports Academy	7	field	0	1	0	by wickets
28	336014	08-May-08	6	3	1	Feroz Shah Kotla	3	field	0	1	0	by wickets
29	336015	08-May-08	1	2	1	Eden Gardens	1	bat	0	1	0	by runs
30	336016	09-May-08	5	8	1	Sawai Mansingh Stadium	5	field	0	1	0	by wickets
31	336017	28-May-08	2	7	1	M Chinnaswamy Stadium	7	field	0	1	0	by wickets
32	336018	10-May-08	3	4	1	MA Chidambaram Stadium, Chepauk	4	field	0	1	0	by runs
33	336019	11-May-08	8	1	1	Rajiv Gandhi International Stadium, Uppal	1	bat	0	1	0	by runs
34	336020	11-May-08	5	6	1	Sawai Mansingh Stadium	5	field	0	1	0	by wickets
35	336021	12-May-08	4	2	1	Punjab Cricket Association Stadium, Mohali	2	bat	0	1	0	by wickets
36	336022	13-May-08	1	6	1	Eden Gardens	1	bat	0	1	0	by runs
37	336023	14-May-08	7	3	1	Wankhede Stadium	7	field	0	1	0	by wickets

Showing 1 to 37 of 577 entries, 19 total columns

Code:

```
# dimentions and names of columns
dim(data)
names(data)
```

Output:

```
> dim(data)
[1] 577  19
> names(data)
 [1] "Match_Id"            "Match_Date"          "Team_Name_Id"
"Opponent_Team_Id"
 [5] "Season_Id"           "Venue_Name"          "Toss_Winner_Id"
"Toss_Decision"
 [9] "IS_Superover"        "IS_Result"           "Is_DuckWorthLewis"   "Win_Type"
[13] "Won_By"              "Match_Winner_Id"     "Man_Of_The_Match_Id"
"First_Umpire_Id"
[17] "Second_Umpire_Id"    "City_Name"           "Host_Country"
```

Code:

```
# sorting data by win type
head(sort(data$Win_Type))
```

Output:

```
> head(sort(data$Win_Type))
[1] "by runs" "by runs" "by runs" "by runs" "by runs" "by runs"
```

Code:

```
# summary of data
summary(data)
```

Output:

```
> summary(data)
    Match_Id       Match_Date         Team_Name_Id    Opponent_Team_Id   Season_Id
 Min.   :335987   Length:577         Min.   : 1.000   Min.   : 1.000   Min.
:1.000
 1st Qu.:419140   Class :character   1st Qu.: 3.000   1st Qu.: 3.000   1st
Qu.:3.000
 Median :548353   Mode  :character   Median : 5.000   Median : 5.000   Median
:5.000
 Mean   :591636                      Mean   : 5.102   Mean   : 5.211   Mean
:5.029
 3rd Qu.:734004                      3rd Qu.: 7.000   3rd Qu.: 7.000   3rd
Qu.:7.000
 Max.   :981024                      Max.   :13.000   Max.   :13.000   Max.
:9.000

  Venue_Name        Toss_Winner_Id   Toss_Decision       IS_Superover
IS_Result
 Length:577         Min.   : 1.000   Length:577         Min.   :0.0000   Min.
:0.0000
 Class :character   1st Qu.: 3.000   Class :character   1st Qu.:0.0000   1st
Qu.:1.0000
 Mode  :character   Median : 5.000   Mode  :character   Median :0.0000   Median
:1.0000
                    Mean   : 5.192                      Mean   :0.0104   Mean
:0.9948
                    3rd Qu.: 7.000                      3rd Qu.:0.0000   3rd
Qu.:1.0000
                    Max.   :13.000                      Max.   :1.0000   Max.
:1.0000

 Is_DuckWorthLewis   Win_Type            Won_By          Match_Winner_Id
 Min.   :0.000     Length:577         Length:577         Min.   : 1.000
 1st Qu.:0.000     Class :character   Class :character   1st Qu.: 3.000
 Median :0.000     Mode  :character   Mode  :character   Median : 5.000
 Mean   :0.026                                           Mean   : 4.991
 3rd Qu.:0.000                                           3rd Qu.: 7.000
 Max.   :1.000                                           Max.   :13.000
                                                         NA's   :3
 Man_Of_The_Match_Id First_Umpire_Id Second_Umpire_Id  City_Name
Host_Country
 Min.   :  1.0       Min.   :470.0   Min.   :471.0    Length:577         Length:577
 1st Qu.: 40.0       1st Qu.:475.0   1st Qu.:488.0    Class :character   Class
:character
 Median :105.5       Median :482.0   Median :493.0    Mode  :character   Mode
:character
 Mean   :139.8       Mean   :484.1   Mean   :495.2
```

```
 3rd Qu.:209.5       3rd Qu.:493.0   3rd Qu.:500.0
 Max.   :460.0       Max.   :511.0   Max.   :521.0
 NA's   :3
```

Code:

```
# finding min and max of first umpire id
min(data$First_Umpire_Id)
max(data$First_Umpire_Id)
```

Output:

```
> min(data$First_Umpire_Id)
[1] 470
> max(data$First_Umpire_Id)
[1] 511
```

Code:

```
# finding mean and median of won by amount
data$Won_By = sapply(data$Won_By, function(x) {
  if (x == "NULL") {
    return(0)
  }
  x
})
data$Won_By = as.numeric(data$Won_By)
mean(data$Won_By)
median(data$Won_By)
```

Output:

```
> mean(data$Won_By)
[1] 17.07972
> median(data$Won_By)
[1] 8
```

Code:

```
# finding quantiles of won by
quantile(data$Won_By)
```

Output:

```
> quantile(data$Won_By)
  0%  25%  50%  75% 100%
   0    6    8   20  144
```

Code:

```
# checking NaN values (if cleanup is required)
sum(apply(data, 2, is.nan))
```

Output:

```
[1] 0
```

Code:

```
# check different host countries
levels(factor(data$Host_Country))
```

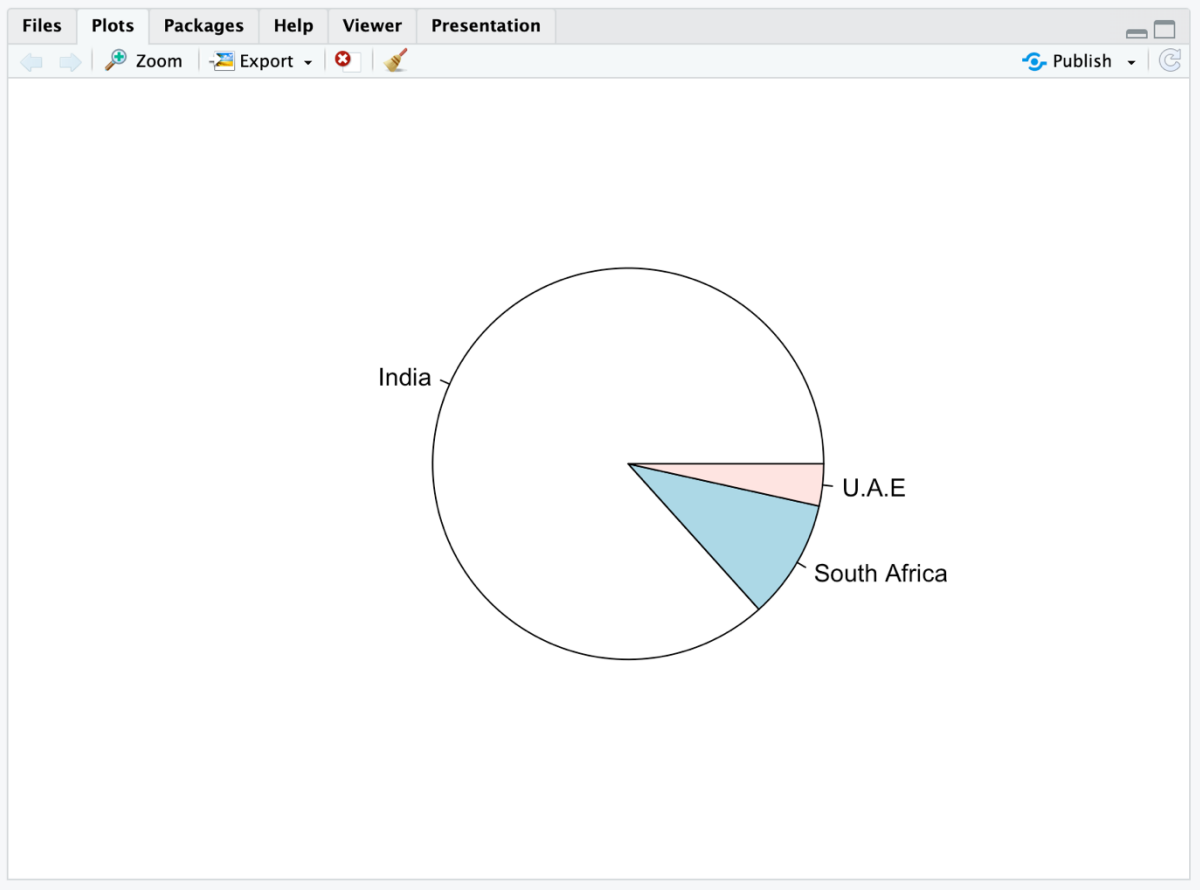
Output:

```
[1] "India"        "South Africa" "U.A.E"
```

Code:

```
# plotting by host country
country_counts = data %>%
  group_by(Host_Country) %>%
  summarise(count = length(Host_Country))
pie(country_counts$count, labels=country_counts$Host_Country)
```

Output:



Code:

```
# check different win conditions
levels(factor(data$Win_Type))
```

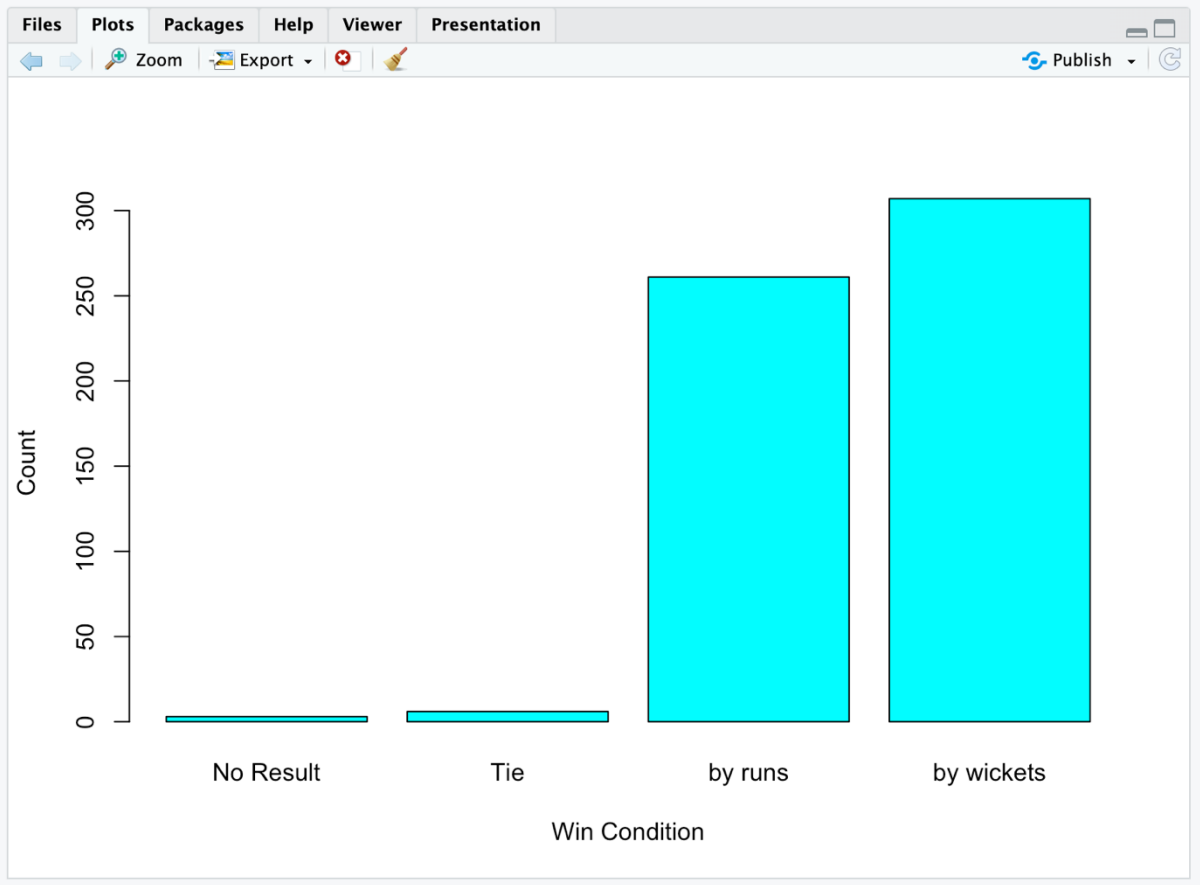
Output:

```
[1] "by runs"    "by wickets" "No Result"  "Tie"
```

Code:

```
# plotting by win condition
win_cond_count = data %>%
  group_by(Win_Type) %>%
  summarise(count = length(Win_Type))
barplot(win_cond_count$count, xlab="Win Condition", ylab="Count",
names.arg=win_cond_count$Win_Type, col="cyan")
```

Output:



Code:

```
# check different city names
levels(factor(data$City_Name))
```

Output:

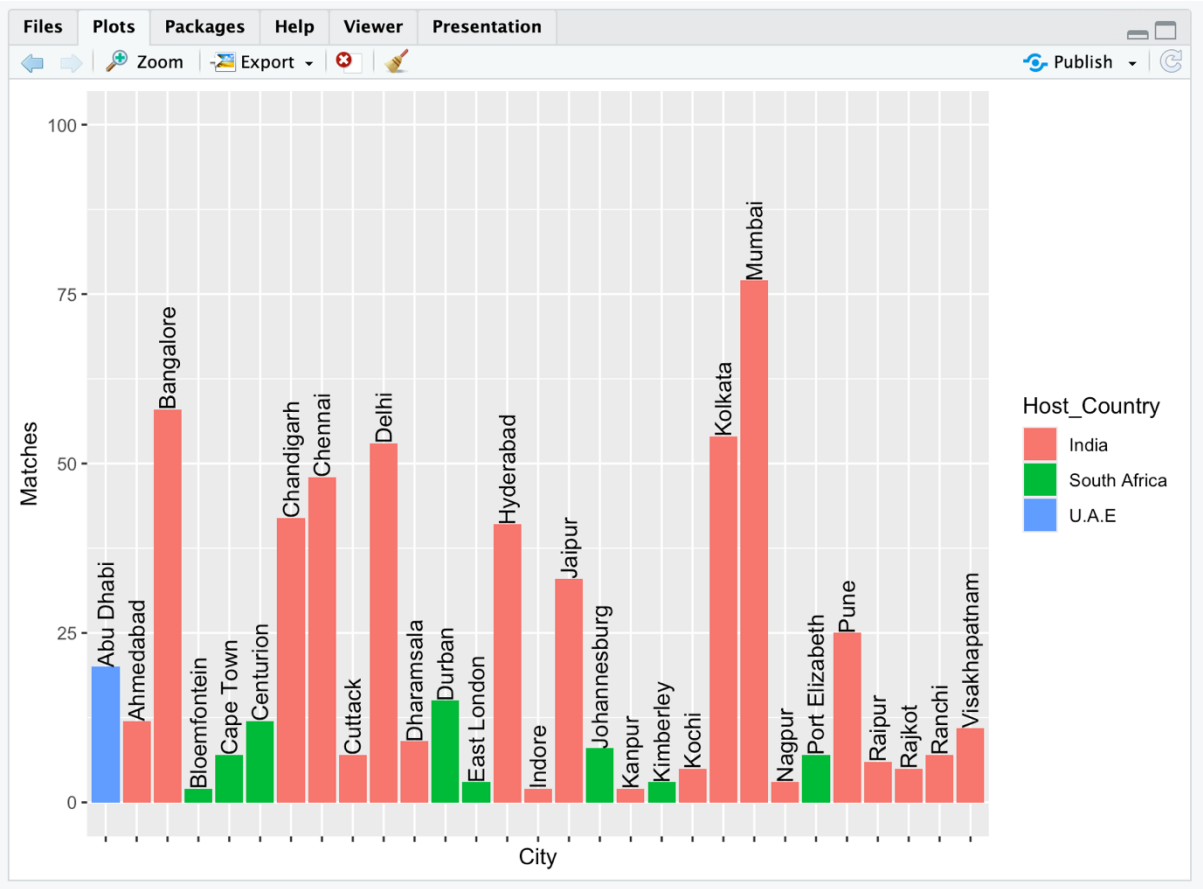
```
[1] "Abu Dhabi"      "Ahmedabad"      "Bangalore"      "Bloemfontein"   "Cape Town"
 [6] "Centurion"      "Chandigarh"     "Chennai"        "Cuttack"        "Delhi"
[11] "Dharamsala"     "Durban"         "East London"    "Hyderabad"      "Indore"
[16] "Jaipur"         "Johannesburg"   "Kanpur"         "Kimberley"      "Kochi"
[21] "Kolkata"        "Mumbai"         "Nagpur"         "Port Elizabeth" "Pune"
[26] "Raipur"         "Rajkot"         "Ranchi"         "Visakhapatnam"
```

Code:

```
# plotting by city name, color by host country
city = data %>%
  group_by(City_Name, Host_Country) %>%
```

```
  summarise(count = length(City_Name))
ggplot(city, aes(x=City_Name, y=count, fill=Host_Country)) +
  geom_bar(stat="identity") +
  geom_text(aes(label=City_Name), vjust=0.5, angle=90, hjust=0) +
scale_x_discrete(labels=NULL) +
  ylim(0, 100) +
  labs(x="City", y="Matches")
```

Output:



Code:

```
# casting match date column to date type
match_dates = data.frame(date=as.Date(data$Match_Date, format="%d-%b-%y"))
head(match_dates, 2)
```

Output:

```
        date
1 2008-04-18
2 2008-04-19
```

Code:

```
# finding the matches played per month
match_dates = match_dates %>%
  mutate(month=month(date)) %>%
  group_by(month) %>%
  summarise(count=length(month))
```

```
head(match_dates)
```

Output:

```
# A tibble: 4 × 2
  month count
  <dbl> <int>
1     3    29
2     4   261
3     5   285
4     6     2
```

Code:

```
# plotting matches played by month
ggplot(match_dates, aes(x=month, y=count, fill=month)) +
  geom_bar(stat="identity") +
  geom_text(aes(label=month.name[month]), vjust=-0.5) +
  labs(x="Month", y="Matches Played")
```

Output: