Project Report on **T20 WORLDCUP 2016**

MTH208A Data Science Lab - I

20/11/22

Submitted By:

Abhraraj Haldar - 221255

Pravin Raman Bharti - 221375

Rajdeep Adhya - 221385

Vanshika Gupta - 211146

Under The Guidance Of:

Dr. Dootika Vats Department of Mathematics

DEPARTMENT OF MATHEMATICS AND STATISTICS

INDIAN INSTITUTE OF TECHNOLOGY

KANPUR

DATA

The data is based on the 2016 ICC World Twenty20, the sixth edition of the ICC World Twenty20, the world championship of Twenty20 International cricket. It was held in India from 8 March to 3 April 2016 and was the first edition to be hosted by India.

The dataset contains 33 rows and 13 columns which means we have observations on a total of 33 matches played between various teams.

Various variables in our dataset are: Team1, Team2, Date, Stadium, Toss_winner, Toss_decision, Team1_runs, Team1_wickets, Team2_runs, Match_winner and POM.

The table below is showing the first 6 rows and 6 columns of our dataset.

Table 1: T20 World Cup 2016

X1.35	Team1	Team2	Date	stadium	Toss_Winner
1	Zimbabwe	Hong Kong	March 08, 2016	Vidarbha Cricket Association Stadium, Jamtha, Nagpur	Hong Kong
2	Afghanistan	Scotland	March 08, 2016	Vidarbha Cricket Association Stadium, Jamtha, Nagpur	Afghanistan
3	Bangladesh	Netherlands	March 09, 2016	Himachal Pradesh Cricket Association Stadium, Dharamsala	Netherlands
4	Ireland	Oman	March 09, 2016	Himachal Pradesh Cricket Association Stadium, Dharamsala	Ireland
5	Zimbabwe	Scotland	March 10, 2016	Vidarbha Cricket Association Stadium, Jamtha, Nagpur	Zimbabwe
6	Hong Kong	Afghanistan	March 10, 2016	Vidarbha Cricket Association Stadium, Jamtha, Nagpur	Hong Kong

We also have other Lists of data frames based on the scoreboard of four semi-finalist teams for which we have given names such as: eng_Mat, Ind_Mat, nz_Mat and Wi_Mat having lists of 24, 20, 20, and 24 data frames respectively.

	Players	Runs
1	Rohit Sharma	5
3	Shikhar Dhawan	1
5	Virat Kohli	23
7	Suresh Raina	1
9	Yuvraj Singh	4
11	MS Dhoni (c)†	30
13	Hardik Pandya	1
15	Ravindra Jadeja	0
17	Ravichandran Ashwin	10
19	Ashish Nehra	0
21	Jasprit Bumrah	0
22	Extras	4

	players	wickets	ECON	0's	4's	6's
1	Ravichandran Ashwin	1	8.00	11	3	2
3	Ashish Nehra	1	6.66	10	2	1
5	Jasprit Bumrah	1	3.75	13	1	0
7	Suresh Raina	1	4.00	10	0	0
9	Ravindra Jadeja	1	6.50	11	2	0
11	Hardik Pandya	0	10.00	1	1	0

OBTAINING THE DATA

We have obtained our data by web scrapping of website named www.espncricinfo.com. Firstly, we got data on 35 matches out of that we have deleted data on two matches which got canceled. So, we worked on data on 33 matches. Then we made data frame as described in Data part of this report.

IDENTIFY ANY BIASES IN THE DATA

As the data is based on actual scores and realistic conditions of matches played by various teams so there may not have any biases.

Still here are some biases which we can think of:

1. We assumed that all matches were played in identical conditions but there can be some biases in terms of weather in a particular stadium.
2. We assumed that all the players were in identical conditions when we observed their performances but there may be some biases as players who were injured doesn't give their best.

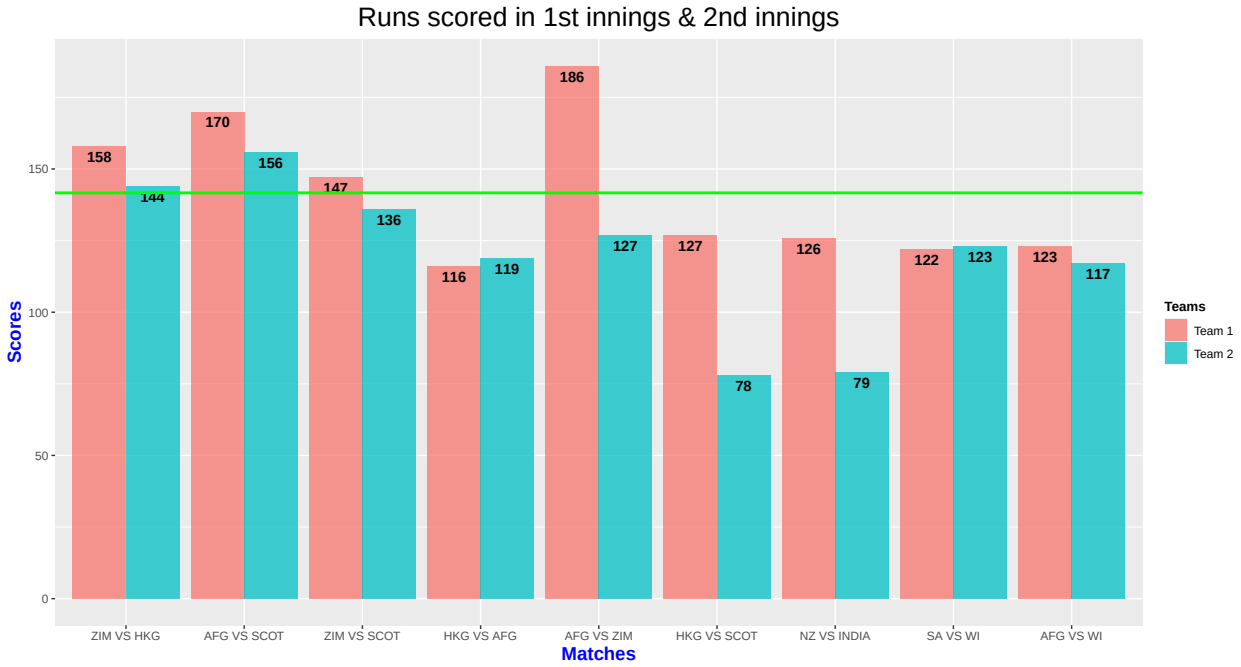
INTERESTING QUESTIONS TO ASK FROM THE DATA

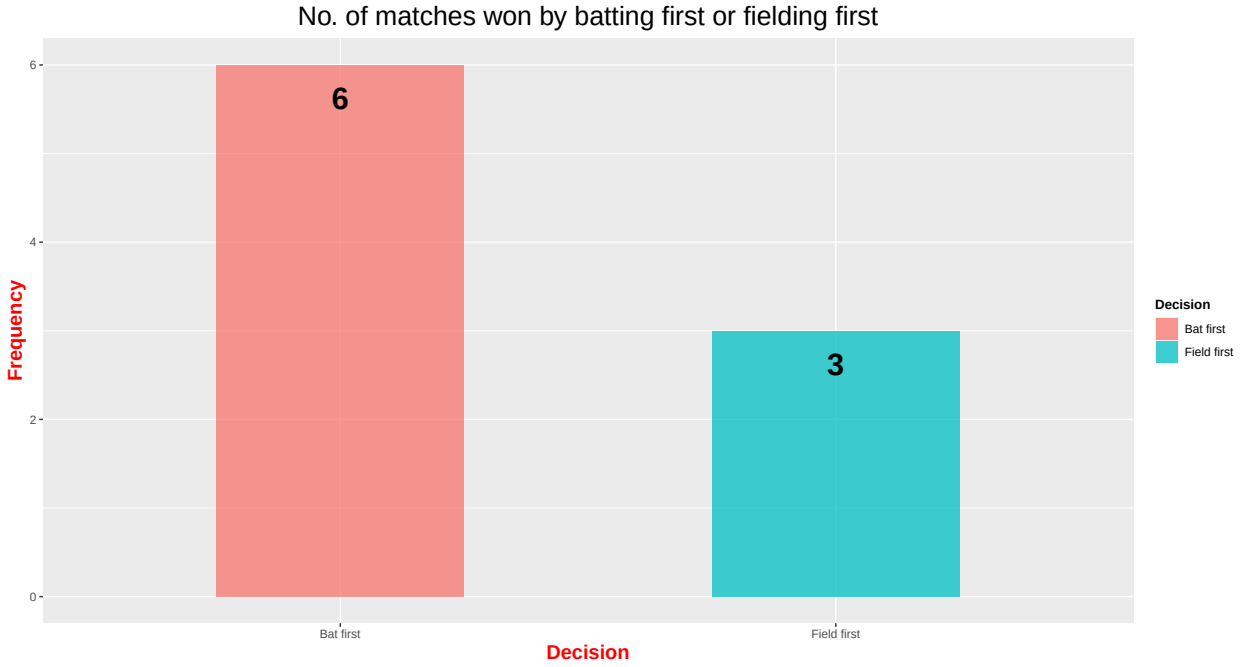
1. Did a particular stadium could play an important role in winning the match through a toss win?

Sometimes a particular stadium or more specifically its pitch plays an important role in winning a cricket match. The bounce and pace of the pitch continuously changes as the match progresses.

For example, in case of Vidarbha Cricket Association, the first innings score in most of the cases is greater than the 2nd innings score which signifies that chasing is relatively difficult in this pitch. A possible reason of which may be the pitch becomes slower and its turn increases after the 1st innings and the batsman faces difficulty in chasing even small a total. The plot below showing us the above fact to us.

Vidarbha Cricket Association Stadium, Jamtha, Nagpur

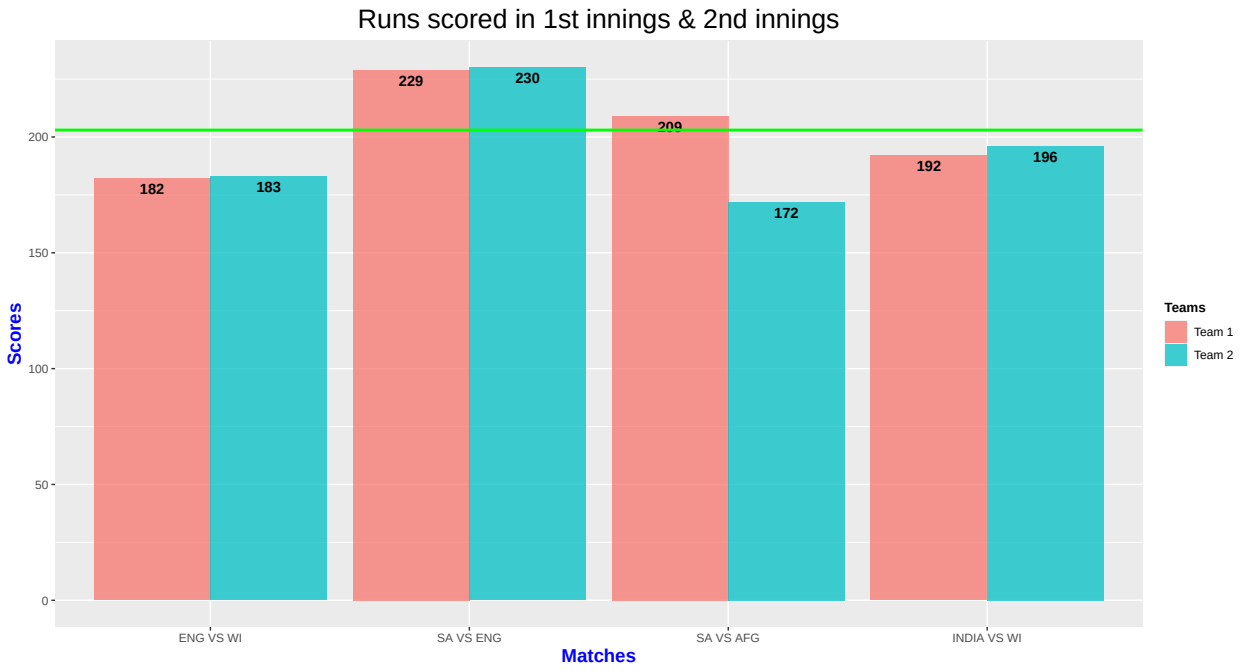




1) As the pitch becomes slower and turn increases possibly it helps the spinners to be dominant in the second innings. As a proof we can see in the Match 13 between India and New Zealand in the 2nd innings the key spinners of kiwis i.e. Nathan McCullum, Mitchell Santner, Ish Sodhi have taken a total of 9 wickets with an Economy less than 5.5 become a key cause of the destruction of Indian Batting line up. Also, in the match between Afghanistan and Zimbabwe famous leg spinner of Afghanistan Rashid Khan performed brilliantly, taken 3 wickets with an economy of 2.75 and destructed Zimbabwe with in 127 runs which proves the dominance of spinners in the 2nd innings.

2) Also, there are some cases where the pitch is a batting pitch and a very high total is also chased in the ground which proves that there are also some cases where the batsmen get help in the pitch like Wankhede Stadium where the total of 229 runs in the 1st innings by SA is chased by England in the match 18.

Wankhede Stadium, Mumbai



All the instances prove the role of the pitch of a particular stadium in a match.

2. May the toss be an important factor in winning a match throughout the tournament?

Here we consider a ratio.

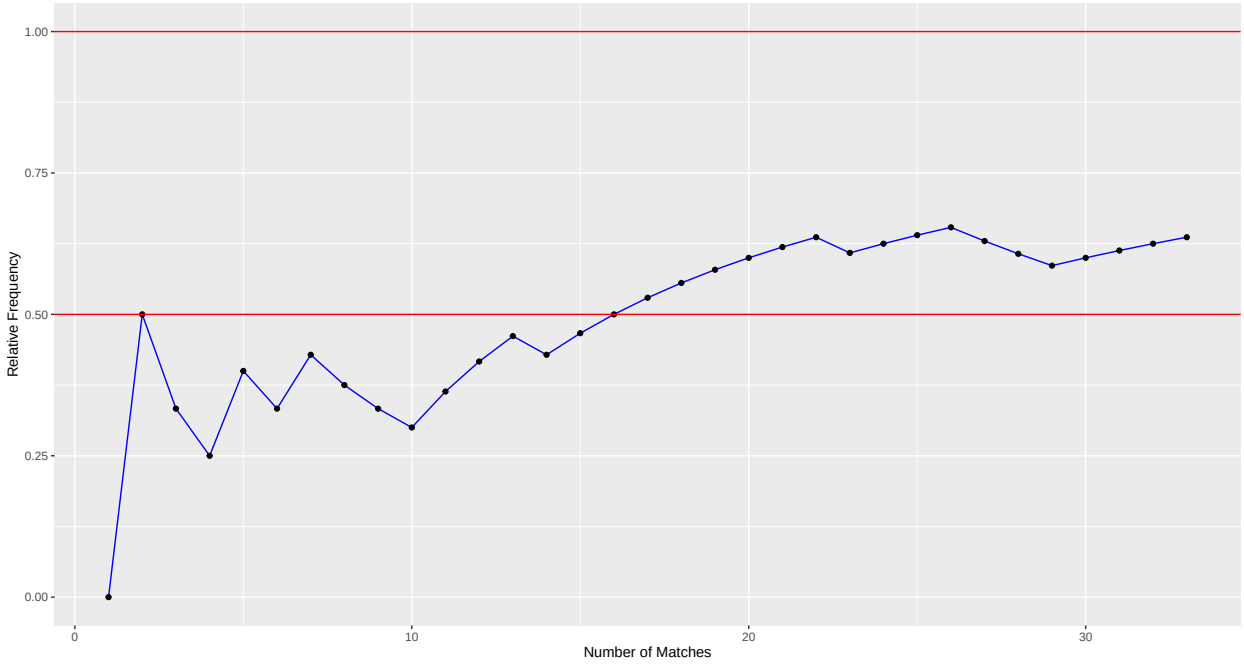
In the denominator, it is the total number of matches and in the numerator, it is among those total number of matches, the number of matches where the team who wins the toss also wins the match.

Here in the first match, the team who wins the toss can not win the match. So, the ratio is 0. In the 2nd match, the team which wins the toss, also wins the match. So now the ratio becomes 1/2. And if we consider 4 matches, we can see that among those 4 matches, in 2 matches, the team which wins the toss, wins the match. So, the value becomes $2/4 = 0.5$.

If we gradually increase the number of matches, the fluctuations decrease. So, a kind of regularity is happening. We have only 33 matches. So, if we can take into account a greater number of matches, we may see proper regularity.

So, from this plot, we can tell in a crude sense that if a team wins the toss, there is a 60 to 70 percent chance to the team of winning the match.

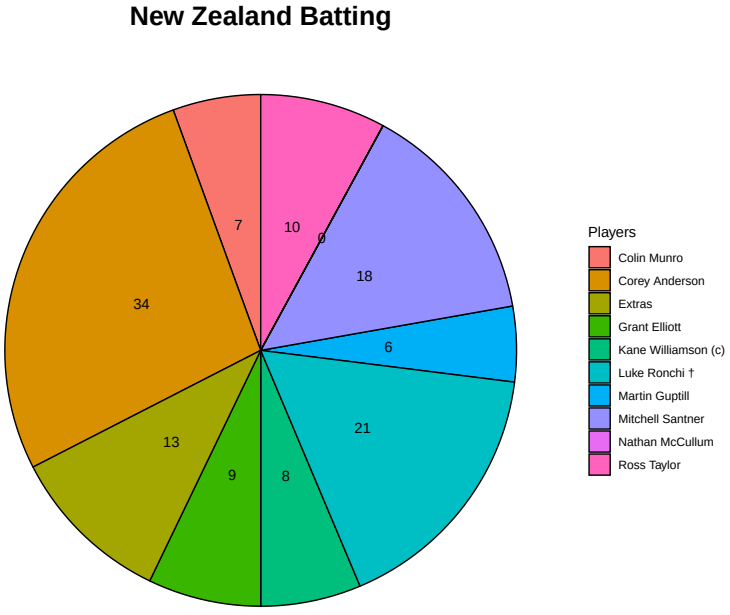
The line diagram below is showing us that crude analysis.



IMPORTANT VISUALIZATIONS

We make pie charts of runs scored by the individual players of 4 Semi-Finalist teams in all the matches.

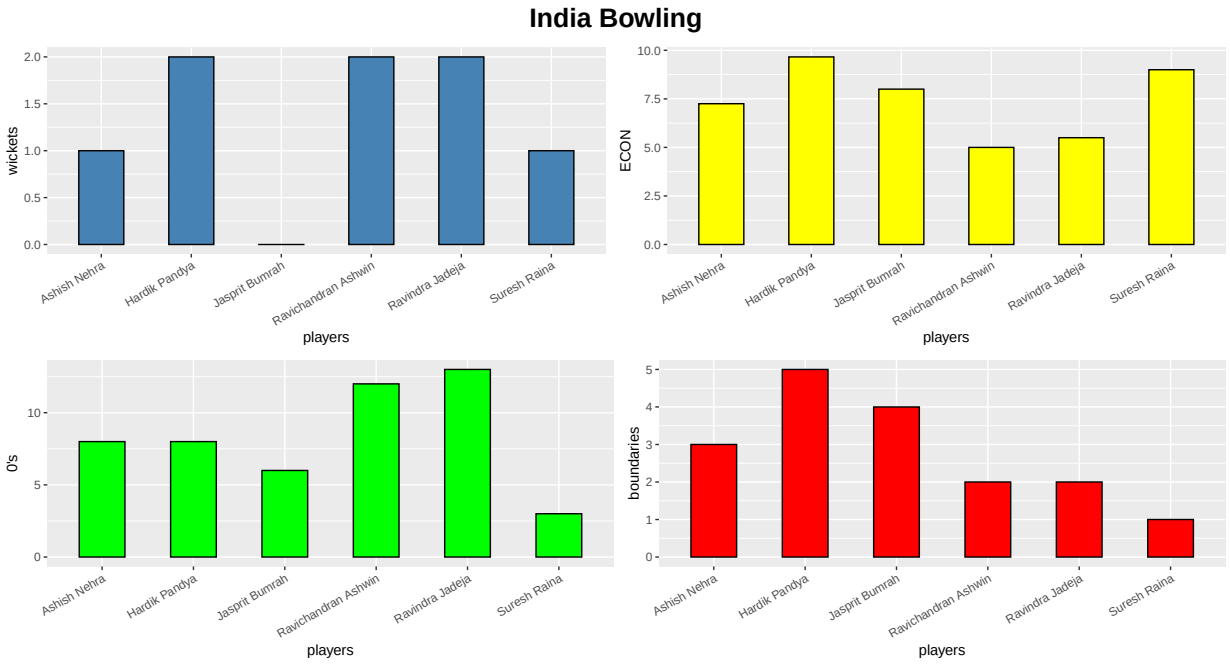
The reason for making pie charts is so that, we can see the contribution of individual players in achieving the total score.



In the above New Zealand VS India match, we can see that, all the New Zealand players had some contribution in scoring a good total.

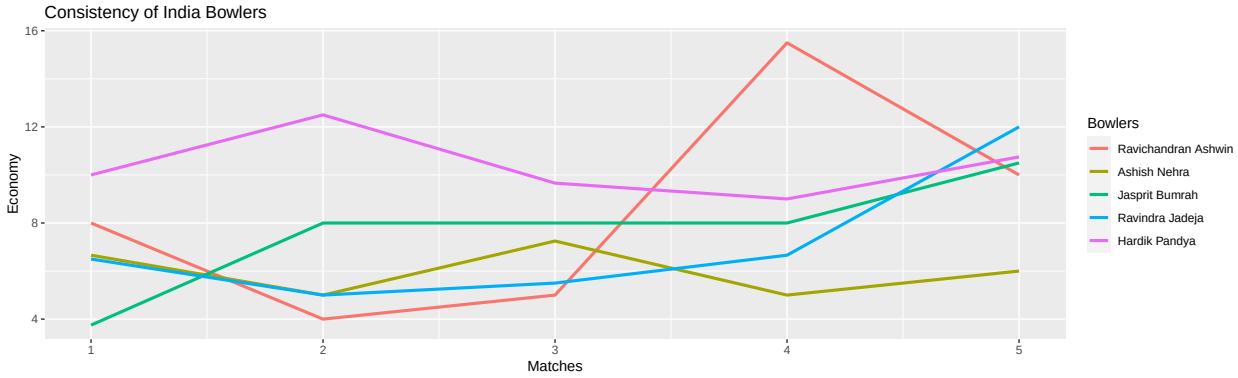
Similarly for the 4 semi-finalist teams, we have also shown the number of wickets taken by the bowlers, their economy, number of dot balls, number of boundaries they got in all the matches, through bar diagrams.

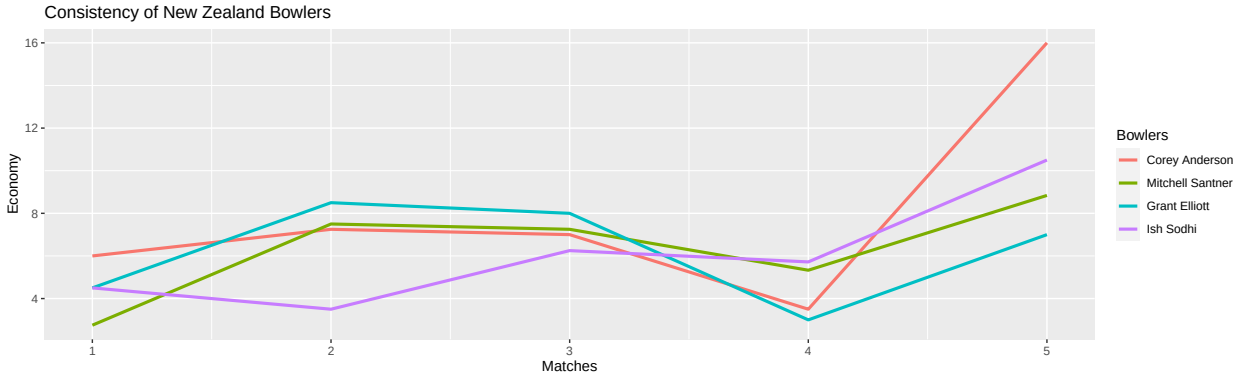
Following is the bowling status of the Indian bowlers in India VS Bangladesh match.

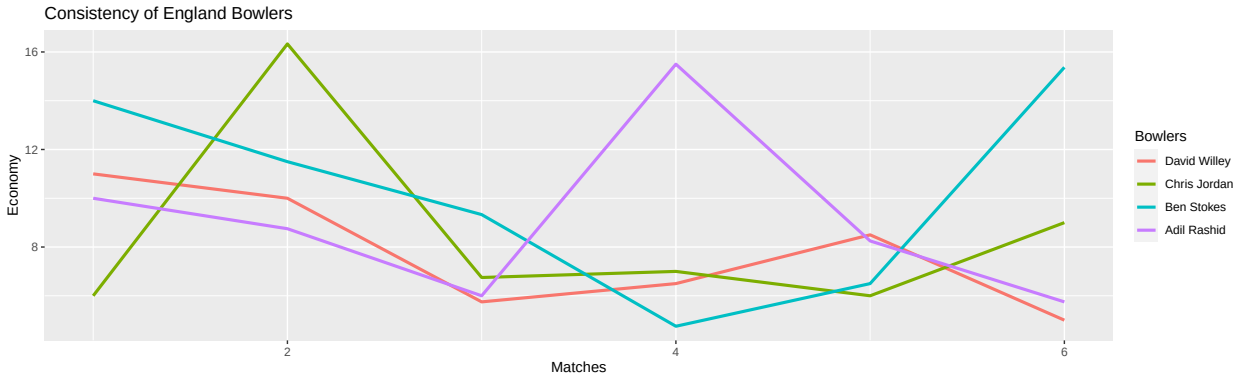


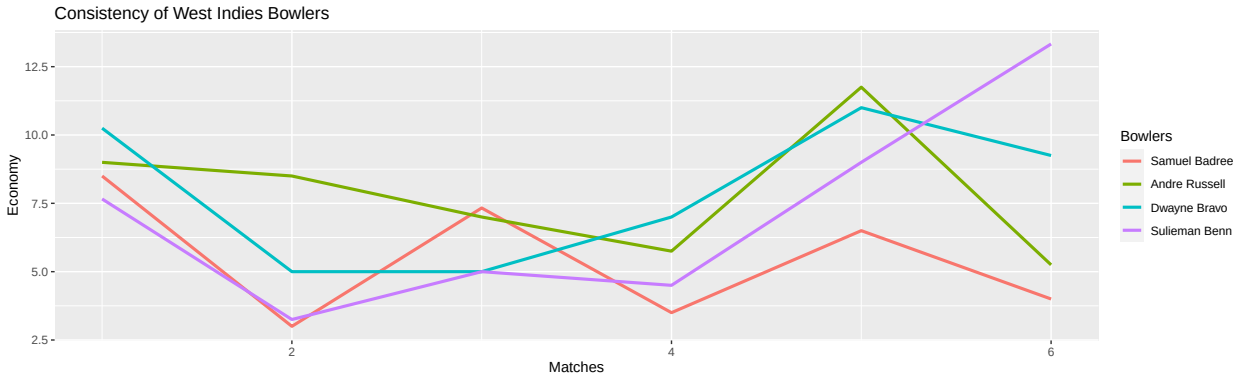
From the above visualization, we can see why Ravichandran Ashwin became the Player Of The Match in the India VS Bangladesh match.

We have shown the consistency of the bowlers of the 4 semi-finalist teams by their economy.









Frome these line graphs, we can clearly visualize the consistency of the New Zealand bowlers, except the last match. Also, we can see the brilliant consistency of the Indian bowler, Ashish Nehra through out the matches.

FINAL CONCLUSIONS

Analysis related to stadiums and players is very crude as we have collected data on very few matches, so we might not make any fair conclusions on the aspects for which we have fewer data points.

1. In the toss factor of our shiny app we have plotted the number of matches vs relative frequency where the relative frequency is the proportion of matches where the team which wins the toss also wins the match. We have observed that if a team wins the toss there is a 60 to 70% chance of winning the match.
2. We have also observed the consistency of bowlers in four semi-finalist teams. The observation might have some biases as the bowlers may not have given their best in each match.
3. Along with this we can also comment on runs that are expected by any team in a particular stadium by looking at its previous average of runs made.

REFERENCES

https://www.espncricinfo.com/series/world-t20-2015-16-901359