A Statistical Analysis of Cricket One Day Internationals

Author: Nikhil Lohia

Team Members:
Aishwarya Pratap Singh, Rahul Aakunuru
and Ayushi Jain

Computer Science and Engineering Program
School Of Computing, Informatics and Decision System Engineering
Arizona State University, Tempe, AZ 85281

April 27, 2017

1 ABSTRACT

Cricket is a game played in over a hundred countries worldwide; One Day International (ODI) is one of the most common forms played. Cricket enthusiasts may be interested in an easy-to-use method to see how their favorite team has been performing throughout the years, while sports analysts may be interested in seeing the trends of different teams' performances throughout the years. In our system, we propose a system of four simple visualizations that come together to tell a story of how the teams have been performing over the period of last 10 years and how their key players have contributed towards these results. The visualizations have been kept fairly simple and easy to understand keeping in mind the nature of audience the system targets.

Keywords— histogram, stacked area chart, cricket

2 INTRODUCTION

Cricket is an extremely popular sport in the world today. The game attracts a huge audience and with it becoming extremely fast paced and increasingly competitive, performance analysis is steadily making its way into the teams dressing rooms. Data visualization and data analysis in cricket has increasingly become a critical part of the game. For instance, players can try to become technically better by analyzing their own performances. Second, teams can make plans for their upcoming matches by analyzing the performance of their competitors. Third, Selectors can analyze team performance to select appropriate teams. Finally, the International Cricket Council can analyze trends to maintain the level of the games taking place.

Each game of cricket can generate a lot of data that can be difficult to comprehend and analyze without some sort of visual representation. Thus, we have created a system of four interactive data visualizations in which users can explore team and player performance and discover underlying trends in the data

3 PROPOSED SOLUTION

One of the main insights of our project is analyzing a players performance over the years. It was my responsibility to visualize this information and I used the concept of stacked area chart. Stacked area graphs are useful for comparing multiple variables changing over an interval and discover trends across a wide range of groups. I am using a spline-based area as it makes the graph look aesthetically pleasing. The years are presented on the X-Axis, whereas the Y-Axis shows the total number of runs scored.
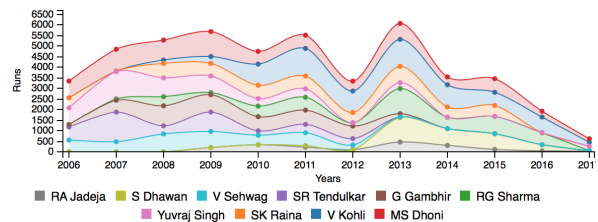

Figure 1: Stacked Area Grapt for India

Stacked area chart is an extremely effective visualization in visualizing the performance of the team based on the sum of the number of runs, the players of the team, have been scoring each year. This also provides an effective means of analyzing the contribution made by a key batsman in the total number of runs scored. This visualization is useful in understanding when the team has peaked and coinciding this with the important tournaments played over the years gives us an understanding of the results of a lot them.

This visualization also helps analyze how the careers of

individual players have evolved over the years. From the debuts and retirements of the players to injuries, that led to a temporary lapse in the performance of a player, can all be visualized using this visualization.

Each color in the graph represents a player and the width defines the number of runs scored by that specific player. The graph follows the Shneiderman's mantra as it gives an overview of the team's performance.

4 RESULTS

The stacked area chart in itself and along with the other visualizations can be used effectively to identify key insights related to the performance of a team and its players.
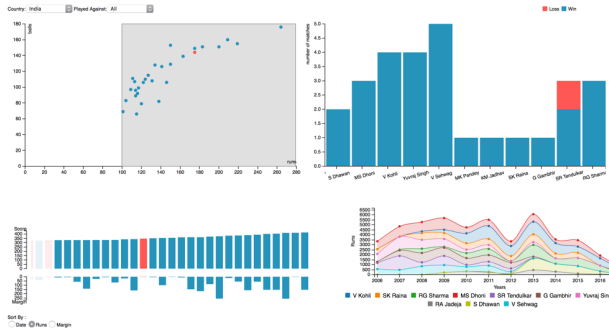

Figure 2: System Overview

4.1 What have been the causes for some of the excep- tionally good and exceptionally bad periods of performance of a team

This case presents itself as a unique study of how the omission of some of the important players of the teams may explain why the team had a slump in their performance at a particular period. For instance, in the year 2016, when three of the most prolific batsmen in the Sri Lankan cricket team, *Jayawardene*, *Sangakara* and *Dilshan*, retired, their performance as team went down to an all time low in the last 10 years.
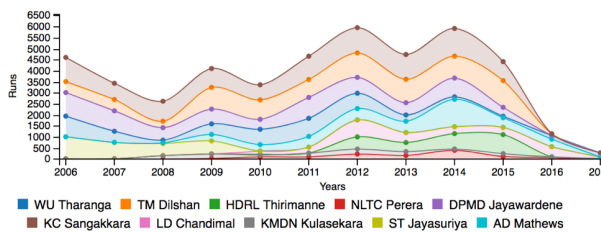

Figure 3: Performance of Sri Lanka

4.2 How were the champions made

This case helps explain why certain teams have performed exceptionally well in some tournaments. India who were the winners of the International Cricket World Cup in the year 2011 and the winners of the ICC Champions trophy in 2013 were at the top of their performance in both those years. They however could not defend their title of the world champions in 2015 due to an indifferent performance by the team. On the other hand, *New Zealand* had an exceptionally good year as a team in 2015 which resulted in them being one of the finalists in the World Cup. This was the first time ever that New Zealand had reached the finals in the tournament.
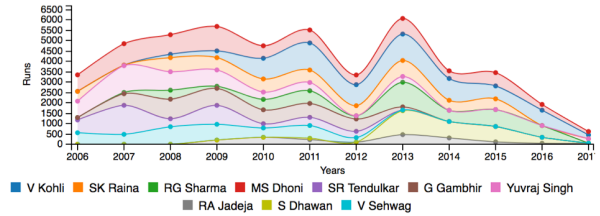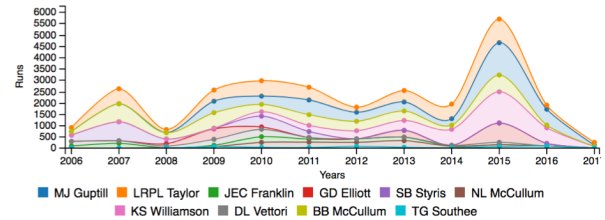

Figure 4: Performance of India

Figure 5: Performance of New Zealand

5 KEY LEARNINGS

A key learning from this project was that the amount of information that can be represented on a unit area of screen is very limited and with increasing it data becomes harder to fetch relevant insights from it. It is really important to express the information in such a way that the used does not get confused and is rather encouraged to explore insights on his own. We tried to explore how the strike rate defines a game outcome.

There are a few more metrics that go on to decide the outcome of a game, but it is interesting to see how with just the batting metric we can explain so many trends and results with respect to the teams. It can be an interesting endeavor to correlate our findings with some of the other metrics in the game.

We have also been able to explore D3, C3 and other re-

lated technologies while developing a fully functional system.

6 CONCLUSION

A good visualization is one which has interactivity and makes the comparison engaging. We present the viewer with data controls and let them uncover things on their own. We have presented a system that analyzes a teams match plan by following Ben Shneiderman's information-seeking mantra[2]. We focused on the fixed parameter of strike rate of a player. However, there still exists a challenge to extract information and gain insights from the bowling statistics of a team. Some teams are known for their exceptional spinners while others are known for

players who can conceive all the wickets in a game. Interactivity among the visualizations is a crucial part of our prototype. However, we may need to search for similar player behavior on more abstract levels. We may want to know how players adapt their behavior when playing with others on their teams; player partnership may be a very interesting area of analysis. This could be another level of exploration.

References

[1] M. Goossens, F. Mittelbach, A. Samarin. *The LATEX Companion*, Addison-Wesley (1994), ISBN 0-201-54199-8.

[2] Shneiderman, Ben. *The eyes have it: A task by data type taxonomy for information visualizations.*, Visual Languages, 1996. Proceedings., IEEE Symposium on. IEEE, 1996.