CricViz: Analysis of Indian Premier League data

Ashwin Sankaralingam, Sreesha Nath, Supriya Naidu

University of Colorado Boulder

ABSTRACT

IPL is a league of cricket teams that play the T20 format of cricket. This league has been around for 10 years and every year and each year there are a total of 70 matches played. We are interested in analyzing this data associated to learn how the teams and the players perform. We visualized the number of games played by each team against other teams, the trend of fours and sixes over the 20 overs, and which batsman and bowler performed the best.

Keywords: Cricket, sports visualization, analysis.

1 INTRODUCTION

Indian Premier League(IPL) is a league of teams owned by celebrities and business tycoons in India. This league convenes once every year to play a season for a span of 3 months. The interesting part about this is that we have cricket players from all over the world shuffled into different league teams. There are 9 teams in the league. They are all paired to play against each other and the best four teams go into the playoffs to win the IPL trophy.

2 RELATED WORK

As the craze for leagues are increasing, it is becoming a bigger appeal, to analyze how the games pan out. This not only helps the fans understand how their favorite players are performing but also helps the teams strategize their games. While researching for this project, we came across several research attempts to understand games like baseball, American football and tennis. In [2] they acquired the data for Red Sox team and performed a ball by ball statistical analysis to see player to player comparison. They sorted their data to find the best player and then focused to analyzing his performance throughout the game. In [1] they did a vantage point analysis based on the position of the players on the field. It was interesting to know the method of preprocessing their data and the architecture they created for data retrieval. In [3] the authors create a full-fledged application to visualize the tennis games. TenniVis allows coaches to understand the best strategies for their players and lay people to see how their hero player performed. [4] discusses the general strategies of sports visualizing and analysis. Even though the paper discusses about processing a lot of data, we were able to apply some of the techniques to our project.

3 DESCRIPTION

We acquired our data for this project from Kaggle. The data is split into 2 comma - separated files, matches and deliveries. The matches file has the details of all the matches played over the 10 seasons. The deliveries file has the ball by ball details for all the overs for all the matches. This gave us more granularity of data and scope for analyzing how the players and teams performed.

3.1 Who played against whom

Our first objective was to see which teams were paired with which other team and how many times. We decided to go with a chord diagram that shows an overview of all the matches played by the teams and on hovering over the arc allocated to the team, the diagram shows the number of matches played by the team in question with the other teams. We chose to use a chord diagram as it a concise overview visualization that can represent multi-levels of data on user actions, like hover, click etc. This can be achieved by encoding data into the length of the arc allocated to the team, the thickness of the chords connecting to the other arcs and different colors add an additional clarity of which arc is connected to which other arc. In our visualization, the length of the arc tells us the total number of matches played by the team in all the seasons, the thickness of the chord outgoing from the arc tells us the proportions of the games played with other teams. Even as we make a choice of using the chord diagram, we understand that using it in a scenario where there are too many teams would lead to clutter.

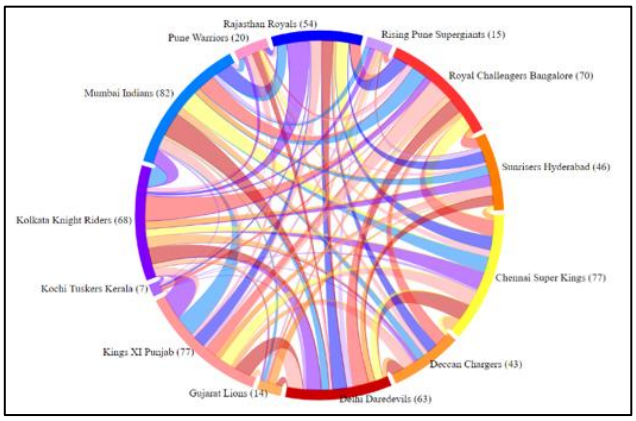

Figure 1: Chord Diagram

3.2 Observing the 4's and 6's

We move on to see the trend of runs scored by the teams in each over. An *over* in cricket comprises of 6 balls. Each point scored is called a *run*. The higher runs that can be scored in a ball are 4 and 6. As ardent cricket lovers, we were interested in seeing when do the players truly take the chance of scoring these runs. We chose to do a bee swarm plot to see how each of the team scores these runs over each over. We used color to categorize the teams and size encoding to show the number of times the teams scored a 4 or a 6 in the specific over.

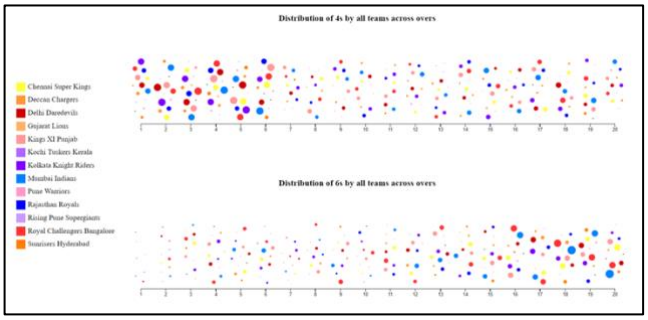

Figure 2: Beeswarm Visualization

This representation is easy to perceive as you can see that the teams score more 4s in the starting overs and more 6s in the last few overs.

One thing to remember here is that there is a high risk of the player getting out when they are trying to hit a 6. Hence the teams attempt to score 6s in the last few overs as they are more inclined to score more runs and less afraid to lose players.

For the visualizations in 3.1 and 3.2 we have used a color scheme that depicts the colors of the jerseys worn by the players of teams in the league.

3.3 Performance of Players

After analyzing the performance of the teams as a whole we proceeded to study the performance of each player. In IPL, the players are auctioned in each season and the owners can bid on the best players if they have information about how the players perform. That serves as the purpose of this study. In cricket, both batsmen and bowlers are important because if the batsmen score a good amount of runs and the bowlers throw poorly placed/paced balls, the efforts of the batsmen would be in vain. Hence, we chose to study the performance of both. We used a bar chart to see the best performance. In cricket, the possible scores that can be scored by a batsman in a ball are 0,1,2,3,4,6. Therefore, we provided the option of clicking on a bar in the bar chart to update the donut chart to see the distribution of the number of different runs scored.

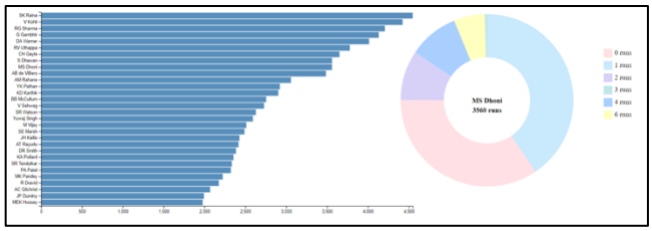

Figure 3: Coordinated bar and donut chart for batsmen

Getting a batsman out is called a *wicket*. There a few different types by which bowlers take a wicket. We show a distribution of the different wickets that a bowler has taken when you click the bar associated with that bowler.

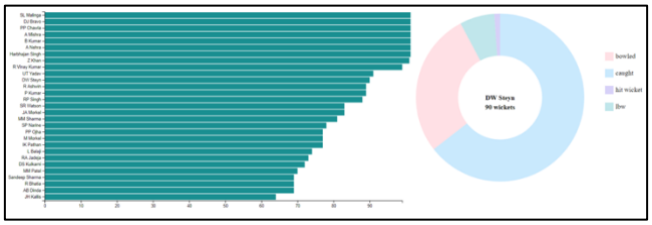

Figure 4: Coordinated bar and donut chart for bowlers

The bar chart is easy to read and its visually easy to pick out the best batsman or bowler. Since the number of runs that can be scored and the types of wickets that can be taken are limited, using a donut chart to see the distribution has proven effective.

4 FUTURE WORK

An interesting way to analyze the given data would be create a virtual pitch and observe the different runs made in each ball by each player in all the teams. It would give more clarity about the consistency of performance of the players.

5 CONCLUSION

Through the course of this project we have seen an aggregation and specialization of the data through the visualizations. We have observed the facets of visual encodings like shape, size, color etc. There is always a downside of using visualizations like chord diagram and donut chart when the categories of data to be represented become too many. But we found them to be truly effective to visualize the questions we wanted answered.

REFERENCES

[1] Toshihiro Tani, Hung-Hsuan Huang, and Kyoji Kawagoe. Sports Play Visualization System for American Football. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2015 Vol I, IMECS* 2015

[2] Cox, Andy & Stasko, John. (2002). SportVis: Discovering Meaning in Sports Statistics Through Information Visualization.

[3] Tom Polk, Jing Yang, Yueqi Hu, and Ye Zhao. TenniVis: Visualization for Tennis Match Analysis. *IEEE Transactions on Visualization and Computer Graphics, Vol. 20, NO. 12, December 2014*

[4] Stein, M.; Janetzko, H.; Seebacher, D.; Jäger, A.; Nagel, M.; Hölsch, J.; Kosub, S.; Schreck, T.; Keim, D.A.; Grossniklaus, M. *How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects*. *Data* 2017, *2*, 2.

[5] J. C. Roberts, C. Headleand and P. D. Ritsos, "Sketching Designs Using the Five Design-Sheet Methodology," in *IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 419-428, Jan. 31 2016.*

[6] Andrea Lau and Andrew Vande Moere Key Centre of Design Computing & Cognition, University of Sydney, Australia. Towards a Model of Information Aesthetics in Information Visualization.

[7] https://www.tableau.com/solutions/workbook/dynamically-track-assets-across-organization

[8] http://buckets.peterbeshai.com/app/#/playerView/201935_2015