

DECISION MAKING VIA DATA ANALYSIS

FINAL PROJECT DOCUMENTATION

Project Members:

1. Abhishek Ghorpade
2. Prudhvi Paruchuri

Dataset: We chose the ESPN Cricket Players Data set which is taken from the [ESPN](#) website. The data includes statistics of the player from the following formats of the game:

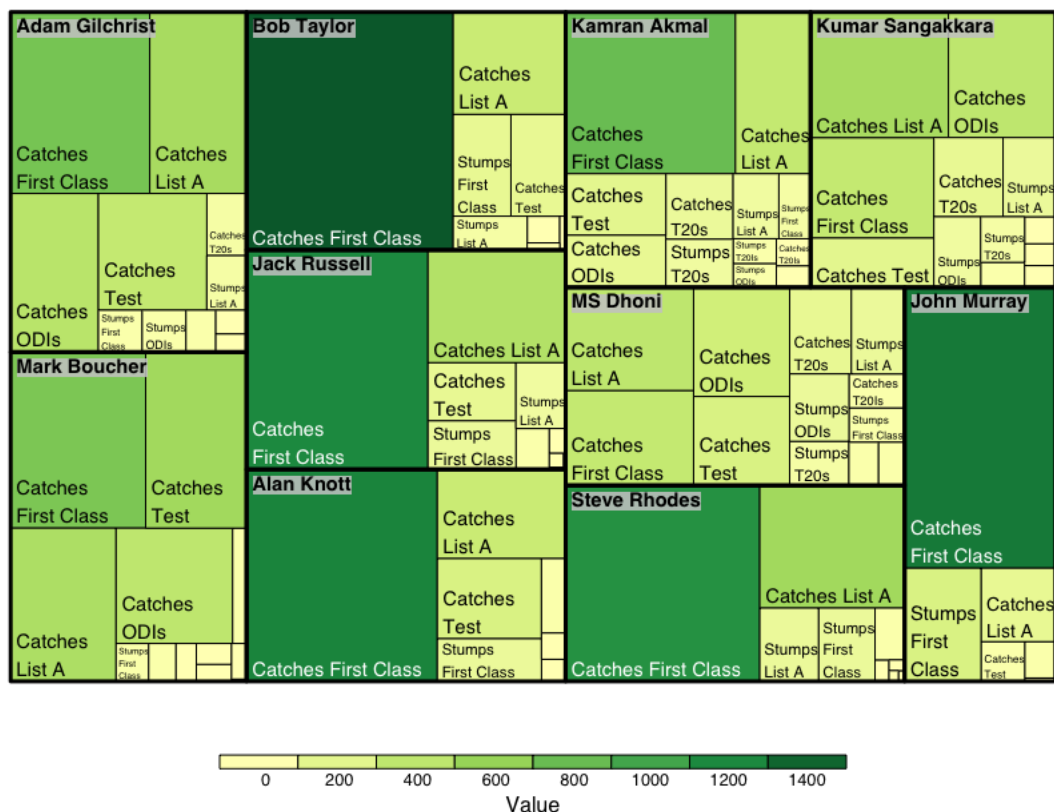
1. One Day International (ODI) Format (Played for 50 overs each; in one day)
 2. Twenty- Twenty (T20) Format both International and Domestic Data (Played for 20 overs each: in 3 hours)
 3. Test (International) Format (Longest format in the game; played for 5 days)
 4. First Class Format (Domestic 3-4 Day Match)
 5. List-A Format (Domestic Limited Overs Match)
- Batters have runs, catches, stumps, highest score, strike rate, etc.
 - Bowlers have wickets, economy (average), runs given, etc.

Data Cleaning: As per the data cleaning part, all the data was either in integer format or character. So, first we filled all the null characters and '-' with NA value. Then we converted all integer format columns to numeric format. Apart from this, we did clean while answering the questions proposed.

Link to dataset: [Kaggle](#)

Proposed Questions:

1. Who are the top ten wicketkeepers in the game and what are the statistics (catches, stumpings) they have made?
2. How do these bowlers ("Harbhajan Singh", "Muttiah Muralitharan", "Shaun Pollock", "Dale Steyn") vary in their innings played and wicket taken in international format?
3. Who is the best allrounder from the top five best allrounders in the game and how do their stats vary? Given they are allrounders show how many innings they played, wickets, runs, catches, stumping they have made.
4. Out of these batters ("Virat Kohli", "Sachin Tendulkar", "Sir Alastair Cook", "Ricky Pointing", "Jacques Kallis") how do their runs made, and innings played vary in all international formats of the game?
5. Which country has the players with the most boundaries (to see the strongest bunch out of all)?



Insights:

- As for the scoring we gave, the top player was Adam Gilchrist, who had great scores in all formats of the game. Also, the tree map shows the same analysis as he has a fair number of stumping and catches in all formats of the game.
- Apart from Adam, Kumar Sangakkara, Mark Boucher, MS Dhoni also has a good number of catches and stumping and runs close in the game to be the best wicketkeeper.
- From the tree map, we can see that Bob Taylor has the highest number of First-Class Catches followed by Alan Knott, Jack Russell, Steven Rhodes, and John Murray. When it comes down to First Class catches, we need them behind the stumps.

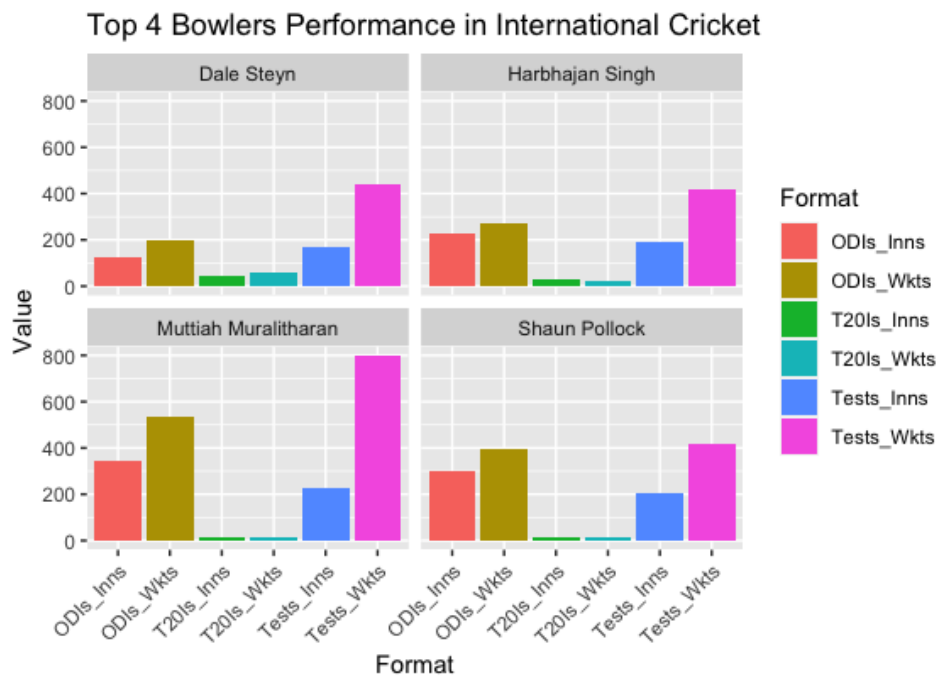
Question 2:

Type of Plot: Bar Chart

Bar chart is selected to show players innings vs wickets for bowler across different formats of the game. This will help us in comparing different players bowling performance in different formats along with their personal performance.

Representation:

Top bowlers in different formats of the game, they have been selected based on their overall high numbers of wickets and a greater number of games played.



Insights:

- First let us talk about Dale Steyn, he has taken maximum wickets in test matches. He has player almost similar number of ODI and test games and has best performance in test matches. As per overall statistics he is one of the best fast paced bowlers the game. He also has maximum T20 wickets among the four players in comparison to other 3
- Next up, Harbajan Singh has good ODI wickets and maximum wickets in test cricket. His bar graph shows most played matches in ODI and better performance in test cricket
- Finally, comparing Saun Pollock with Muthiah Muralidharan, Muralidharan has the most wickets in test cricket and highest wickets in ODI. They have played only a few T20 matches.

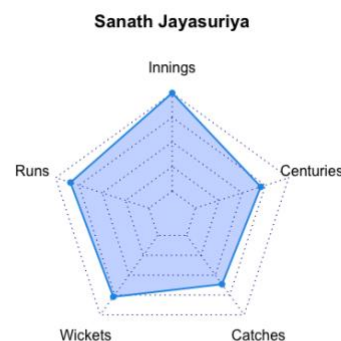
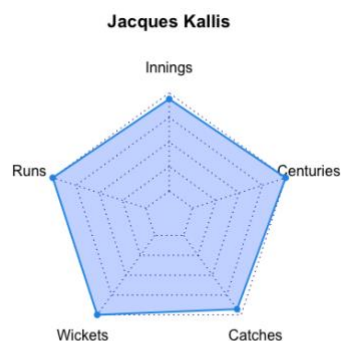
Question 3:

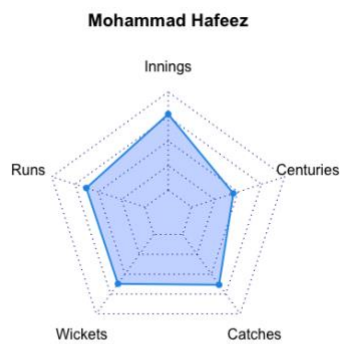
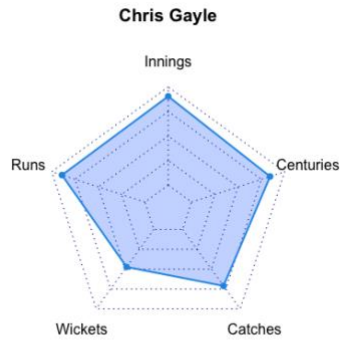
Type of Plot: Radar Chart

We chose a to use Radar Chart because we wanted to see difference among five variables in top five players selected by score. Radar Chart will help use see where each player is excelling as compared to the other and which one has the highest variable score of all. This will let us compare where each player is lacking and where each player is outperforming.

Representation: Top 5 Allrounders in the World

These five players have been selected on a basis of score. The scoring technique was simple, for runs and wickets they made in their career players get twice the amount of runs and wickets as their score and catches, innings played, and centuries have no bonuses. Now based on these scores, we picked the top 5 players and compared all these characteristics.





Insights:

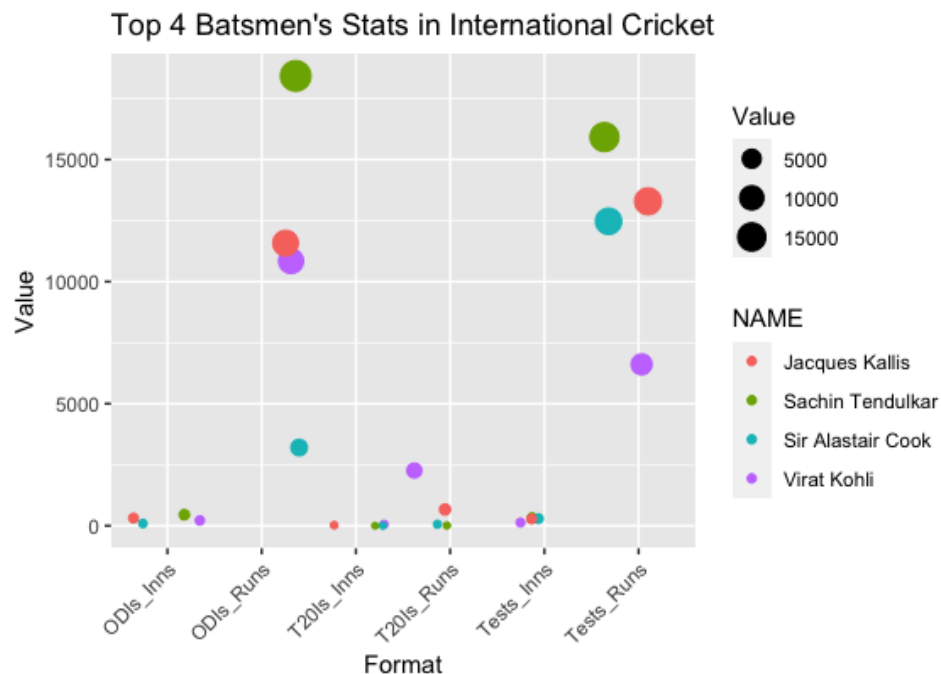
- First, let us talk about Jacques Kallis, hailed as the best all-rounder in the game. As per the statistics and people, he is the best all-rounder in the world. He has played more matches, scored runs, took wickets down and made lots of tons. The Radar Chart is filled with his talent.
- Next up, Chris Gayle and Sanath Jayasuriya, both have played a lot of games, scored the same number of runs and centuries, the one place where they lack is at wickets and catches. Sanath Jayasuriya has more wickets than Chris Gayle and vice-versa.
- Dilshan, on the other hand, is like Jayasuriya, his teammate, but has fewer overall wickets, he was good in fielding and batting.
- Mohammad Hafeez has an overall balanced radar chart after Jacques Kallis, he has a fair number of runs, wickets, centuries, catches and, innings played.

Question 4:

Type of Plot: Scatter plot with players comparison of total runs in different format of game

For this type of question which is based on statistics we have chosen scatter plot with facet of players which gives us understanding of each player performance in each type of game. We also wanted to compare the players' individual performances and their performance compared to other top players

Representation: We have shown numbers in y axis and different format and runs are shown in x axis. We have done facet on players to compare them with each other. Color coding is added to make visualization cleaner.



Insights:

- First we compare Jacques Kallis's individual performance. He is an all-round player and has most runs in test format. He has played fewer T20 games but still managed to score highest runs.
- Next up is Sachin Tendulkar, he is the player to score maximum runs in both ODI and test cricket with most matches played as well. The graph clearly shows his top statistics
- Finally, we compare Sir Alister Cook with Virat Kohli where Cook has the most runs in test matches, however Kohli has the most runs in ODI and T20. Both are star batters with maximum runs out there.

Question 5:

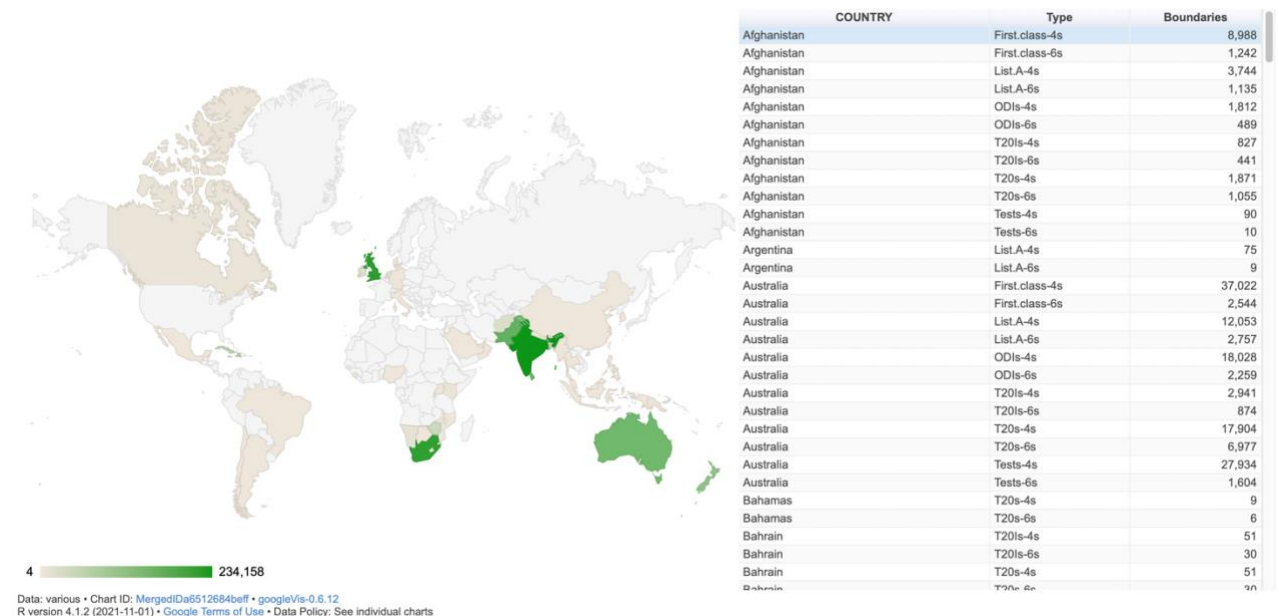
Type of Plot: Geographic Heat Map with Country and Format based 4's or 6's Information

For this type of question, which is spread across the globe, we must take a geographic map, we chose to take the Geographic heat map from google Vis's library as we also wanted to make it interactive to also show each countries cumulative boundaries along with all the boundaries values for each format of the game.

Representation: Cumulative Boundaries Contribution

The first thing is to select data for 4's and 6's which are boundaries. Then fill NA values with 0 and gather with a Format and Boundaries key value pair for all the formats 4's and 6's. Now we can group by country summaries boundary values to cumulate and get a count of boundaries for each country per format.

There are a few countries like West Indies which are represented as three countries (Cuba, Jamaica, Haiti) and China (China + Hong Kong).



Insights:

- The map overall, shows how widespread the game is around the world, there are players from most parts of the world.
- We can see all the countries which were once British (now United Kingdom) ruled have more overall boundaries. The top countries like India, Pakistan, United Kingdom,

Australia, South Africa, Bangladesh, New Zealand, Jamaica (West Indies) were all once British colonies.

- Only, countries like United Kingdom, West Indies, South Africa and, Zimbabwe on the western side have more boundaries, while most of the power hitters are from the eastern side of the map from countries like India, Bangladesh, Australia, Pakistan, and New Zealand.
- The table on the right shows us more in-depth details about each country. For instance, Afghanistan has a greater number of First Class 4's than ODI 4's.