Data Visualisations in Football

FINAL REPORT

Submitted by

Manan Rajvir (17BCE0112)

George Joseph Chakola (17BCE0118)

Prepared For

DATA VISUALISATION (CSE3020) – PROJECT COMPONENT

Submitted To

Dr. Annapurna Jonnalagadda

(Scope)

School of Computer Science and Engineering

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Table of Contents

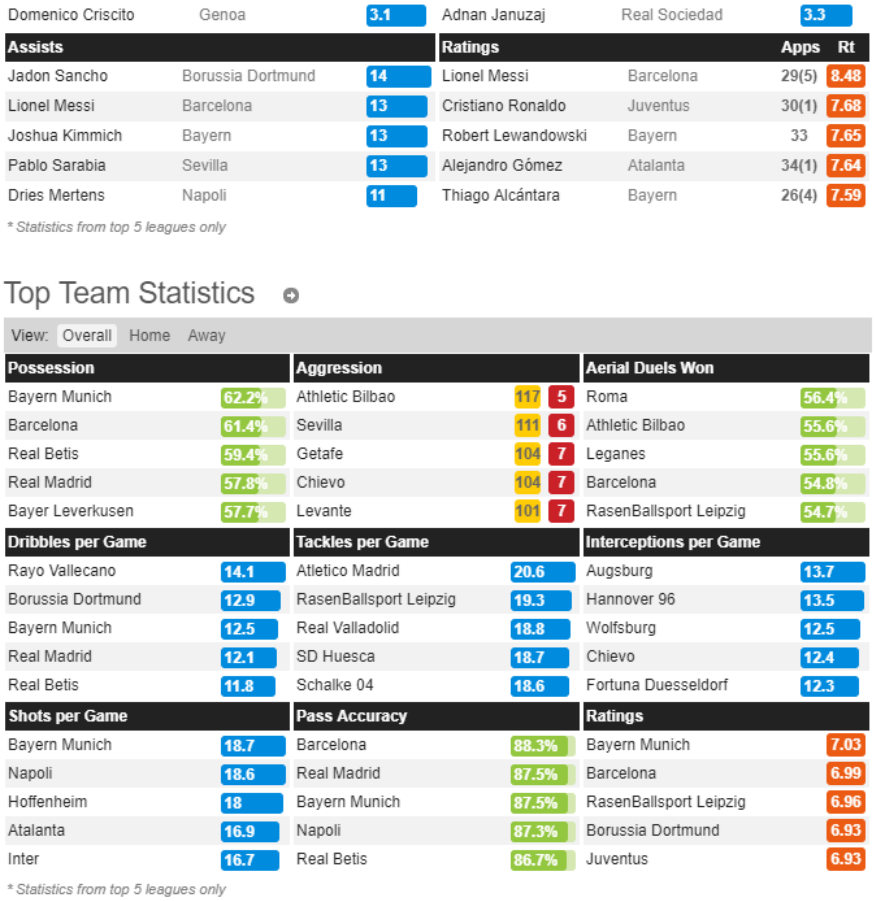
1. Problem Statement

When it comes to sports, there is a lot of data involved in every aspect of the game. For example, in a game like football, the data varies from the stats of the players, the records of the teams, to in game data like possession, passing accuracy, shot accuracy and so on. This data in itself can get overwhelming for a fan/viewer. Therefore our aim is to break down this data and visualize it through various representations that makes it easier for the viewers to interpret and analyse data.

2. Introduction

2.1 Motivation

A normal game of football involves a lot of statistics. This could be the data about the past results of the teams involved, the stats of different players (goals scored, shots, tackles, dribbles, assists, match rating) or live game information like possession, passing percentage, shooting percentage and so on.



Domenico Criscito	Genoa	3.1	Adnan Januzaj	Real Sociedad	3.3

Assists			Ratings		Apps	Rt
Jadon Sancho	Borussia Dortmund	14	Lionel Messi	Barcelona	29(5)	8.48
Lionel Messi	Barcelona	13	Cristiano Ronaldo	Juventus	30(1)	7.68
Joshua Kimmich	Bayern	13	Robert Lewandowski	Bayern	33	7.65
Pablo Sarabia	Sevilla	13	Alejandro Gómez	Atalanta	34(1)	7.64
Dries Mertens	Napoli	11	Thiago Alcántara	Bayern	26(4)	7.59

** Statistics from top 5 leagues only*

Top Team Statistics

View: Overall Home Away

Possession		Aggression			Aerial Duels Won	
Bayern Munich	62.2%	Athletic Bilbao	117	5	Roma	56.4%
Barcelona	61.4%	Sevilla	111	6	Athletic Bilbao	55.6%
Real Betis	59.4%	Getafe	104	7	Leganes	55.6%
Real Madrid	57.8%	Chievo	104	7	Barcelona	54.8%
Bayer Leverkusen	57.7%	Levante	101	7	RasenBallsport Leipzig	54.7%

Dribbles per Game		Tackles per Game		Interceptions per Game	
Rayo Vallecano	14.1	Atletico Madrid	20.6	Augsburg	13.7
Borussia Dortmund	12.9	RasenBallsport Leipzig	19.3	Hannover 96	13.5
Bayern Munich	12.5	Real Valladolid	18.8	Wolfsburg	12.5
Real Madrid	12.1	SD Huesca	18.7	Chievo	12.4
Real Betis	11.8	Schalke 04	18.6	Fortuna Duesseldorf	12.3

Shots per Game		Pass Accuracy		Ratings	
Bayern Munich	18.7	Barcelona	88.3%	Bayern Munich	7.03
Napoli	18.6	Real Madrid	87.5%	Barcelona	6.99
Hoffenheim	18	Bayern Munich	87.5%	RasenBallsport Leipzig	6.96
Atalanta	16.9	Napoli	87.3%	Borussia Dortmund	6.93
Inter	16.7	Real Betis	86.7%	Juventus	6.93

** Statistics from top 5 leagues only*

Figure 1 Football Statistics

However these statistics on in its own is of very little value to the fans, analysts or any other manager studying the data. Furthermore, the volume and variety of data can make it overwhelming for viewers. Therefore we plan to represent this information using different representations and visualizations so that the data can be analyzed more effectively.

2.2 Significance

This significance of this project is thus, that it creates a wide range of visualisations that can be used to analyse various different aspects of the game, rather than focusing on one particular aspect. The project also ensures that the interface and the visualisations in themselves are very simple and anyone can easily make insights from the data shown. This is especially helpful as most of the visualisations take data from datasets that are full of numbers and stats, and will be hard to analyse for the average person. Finally, the project also provides a range of functions that can be help various different people like managers, players, scouts, fans and fantasy league players.

2.3 Scope and Applications

The visualisations created in this project are aimed to help 3 specific groups of people:

A) Football Fans:

The visualisations can be used by football fans who would like to see several different aspects of the game. For example, the fans can view who the most important players are for several different teams. This can be done by using the Viewer Mode which allows the user to view contribution of players in terms of goals and assists for their particular teams. The fans can also analyse the graphs on historical league standings to understand how consistent their favorite team has been over the last decade. Finally, the project also allows the fans to compare their favorite players against each other on the basis of attributes like shooting, pace and defense. Finally, the fans can also make use of the heat map to predict the final score of different matches based on the performances of the teams in their leagues. The visualisations have been kept simple to ensure that the fans can make insights from the complex databases available.

B) Football Managers:

The project offers several functionalities to the managers. First off, the managers can make use of the radar plot while deciding which player to buy. The radar plot can help

the manager decide if a particular player is the kind of player he is looking for and has the necessary attributes. Additionally, the manger can also understand the tactics of the other team and analyse the side on which they are more dominant by analyzing the shot map. The shot map will show whether the opposition team makes most attacks from center or the sides and will help the manager make his defensive tactics accordingly. Finally, the manager can also decide which players from the opposition team to look out for by analyzing the contribution of players in goals and assists.

C) Fantasy Team Players:

The final intended user of this project are the fantasy team players. There is a specially intended viz idiom that helps these player create their teams. The data viz makes use of dynamic tables to help the players select the players based on the criteria they specify and then visualize this team using a formation on the pitch. The fantasy team players can also make use of the heat map to find the most probable score lines and then pick attackers or defenders accordingly.

3. LITERATURE SURVEY

3.1 Related Work

3.1.1 TenniVis: Visualization for Tennis Match Analysis

Most of the research efforts into tennis visualization primarily focus on using ball and player tracking data which can then be used to analyse player strategies and movement. The biggest drawback of this is that analysing and recording such data requires expensive equipment. Therefore, this research focuses on a tennis match visualization system that uses data that can be easily gathered. This includes data such as score, point outcomes, point lengths, service information, and match videos that can be captured ordinary cameras. The current limitation with this set of data is that it is present mainly in statistical form and thus it is hard to make useful insights.

The application, TennisVis provides 2 visualisations - Pie Meter view and Fish Grid view. The first provides a high level overview of the game. It represents data like who won the game, and their dominance. Whereas, the fish grid view provides details at the game level. The game length, critical points and trends are analysed at this level. Additionally, the interactivity of the software allows the user to examine details like percent distribution, point, first serve percentage of point outcomes in forms of bar charts.

The visualization was verified by 2 pilot user studies, and it proved that tennis coaches were quickly able to understand the functioning of the application.

3.1.2 Sports Data Visualization

This research is focused on analysing the growing use of data visualisation in sports. It reflects how for every major sport, analysts extract large amounts of data, which can be used by the media and fans (for entertainment), athletes (to improve individual and team performance), and organizations (to yield a deeper understanding of the sport itself). Besides the normal game itself, data collection has also become an important part of fields like medicine applications and rehab for players. The teams often employ analytical staff that help prepare training plans, predict athlete injury risks, and prescribe personalized recovery strategies.

In recent times, media outlets such as ESPN, FiveThirthEight.com, New York Times have used various visualisations to represent sports data. Infact, the study found out that use of data visualisation in sports has also increased in conferences like IEEE Information Visualization (InfoVis), ACM Conference on Human Factors in Computing Systems (CHI), International Association of Computer Science in Sport (IACSS), and MIT Sloan Sports Analytics.

3.1.3 Visualization of Sports using Motion Trajectories:Providing Insights into Performance, Style, and Strategy

This paper develops a sports visualization system for tennis called LucentVision. LucentVision uses real-time video analysis to obtain motion trajectories of players and the ball, and offers a rich set of visualization options based on this trajectory data. Rather than using technology to resynthesize the sport, this paper focuses on using it for further analysis of players, teams and performances. This application uses 8 cameras placed around the stadium to track the players and the ball which then form the trajectory. Based on this trajectory, several different visualizations can be formed. These include:

A) Coverage Maps - This is a follows the trajectories to analyse the position / play of players: do they play near the baseline or come towards the nets.
B) The dominant hand of the player - backhand and forehand is also analysed by LucentVision.
C) Speed charts and virtual replays of serves can also be traced using the trajectories recorded.

Finally, the service landing positions is also analysed by the cameras to show the general trend in a players serves. Therefore the LucentVision system shows how the numerous visualizations made possible which can then be used to analyse player styles, strategies and performance.

3.1.4 SportVis: Discovering Meaning in Sports Statistics Through Information Visualization

SportVis uses visualization to help people discover meaning in the massive amount of statistics generated during sporting events. This research paper focuses on baseball and how different visualization techniques can be used to enhance the analysis. The data for this research has been obtained from Retrosheet and Baseball Databank and has then been used to develop 2 visualizations - Baseline bar display and Player map.

The baseline bar is used to represent attributes related to a particular game. It shows the result of the game, the location of the game and the runs scored by the team. This graph will also highlight the form of the team.

Whereas the player map provides a way to quickly identify ineffective allocations of playing time. The treemap design where each rectangle represents a player, with the area encoding the number of plate appearances. The colour (a green to red spectrum encodes high to low values) of each rectangle encodes a particular rate statistic, which can be selected by the user.

3.1.5 Towards Classifying Visualization in Team Sports

Rather than providing a new visualizing technique, this paper analyses the use visualisations in team sports. The use of visualisation has been divided into 3 groups:

A) Athlete centered: which focuses on supporting the athlete.
B) Spectator centred: which is designed for the audience.
C) Judgement centered: which is geared towards the accurate, fast and effective judgment of team sports by one or more official referees.

The paper also analyses the use of visualisations in clothing, environment, media and wearable technologies. The conceptual model created is designed to allow any team sport visualization to be described. The front and side views of the conceptual model allow certain visualization trends and gaps within team sports to be identified. From the side view, it is found that visualizations which occur close within the vicinity of game-play usually convey information needed for success, such as rules and identity, and hence need to be understood quickly.

3.1.6 SnapShot: Visualization to Propel Ice Hockey Analytics

This paper focuses on developing SnapShot, a system for exploring, discussing, and presenting hypotheses and findings using National Hockey League shot data. The different visualisation techniques are - Shot map, Traditional heat map, Radial heat map.

The shot map highlights shot location, result, and the shooter team distinctions. Other attributes like period, rink can also be represented if needed. The traditional heat map view on the other

hand is designed primarily to highlight patterns of shot location density. Finally, the radial heat map is created to analyse the shot length.

The application was then used by 3 domain experts who claimed how SnapShot can be used to analyse the games better. However, there are several modifications that can be made to this application for future work. This includes:

- Expanding the number of event types that SnapShot supports to include events such as hits, passes, and penalties.
- Adding analytic capacity to the system in order to support visual analytics methods.
- Filtering the results based on whether a shooter's team is winning or losing.

3.1.7 The research develops a visualisation system called SprotsViz.

The visualization system uses both statistical analysis and motion analysis where play classifications are realized using trajectory mining. In this system, the type of attack pattern and the play of each pattern can be understood dynamically and visually. This includes data such as trajectory of players, type of attack and positions of players.

This current limitation of this tool is that it is only applicable to American Football. Even within the same game, the play classification can be improved by making play tendency determination easier.

3.1.8 The harsh rule of the goals: data-driven performance indicators for football teams

This research tries to show how by adopting a data driven approach there is a large potential to boost the understanding of team performance in football. This includes capturing crucial aspects of the passing behaviour of a team from observational data of a football game.

Methodology: First the different events in a football match are noted down. This includes data such as passes, goals, position of ball of field, etc. Based on this data alone, correlation between attributes like average passes made to goals scored, goal attempts is calculated. This data is also plotted onto charts to make the comparison easier. However, the number of passes made isn't important attribute. The distribution of passes will also impact the goals scored / chances created. This graph can be created using the average and variance of passes made by individual players in a team. The distribution of passes over the zone of the pitch is also added to this

visualisation. Finally, a combination of all the visualisations was used to predict the performance of teams in 4 different football leagues.

While the research was able to make a lot of accurate predictions, the limitation of this paper is that it only focuses on the passing. However, the result in a football game is dependent on a lot of factors besides just passing. Furthermore, within the same application, several different factors / attributes can be added. This includes attributes like difficulty of a pass in the performance indicators. Passes, either short or long, have different difficulties depending on additional features, i.e. the zone where they take place and the direction of the pass (forward/backward). Additionally, information about defensive events (tackles, goalkeeping actions, recoveries of ball and so on) have also been ignored in the given study. Defensive actions are indeed crucial in the strategy of a team and they can significantly improve the description of the performance of a team.

3.1.9 Sports Scene Analysis and Visualization from Multiple-View Video

This research, introduces methods for sports scene analysis and visualization from multiple videos captured with multiple cameras.

The first part of this is tracking multiple soccer players. The cameras installed across the ground track the foot position, the projected position and player trajectory. The next stage involves free viewpoint visualisation. In these systems, we employ view interpolation method as a key method for free-viewpoint visualization. For view interpolation of dynamic objects such as players and a ball, all dynamic regions are first extracted by subtracting the background image from the current frame image.

Therefore, the research presented a method for tracking of multiple football players, which is performed by integrating the tracking data from all cameras based on the geometrical relationship between cameras.

3.1.10 Baseball4D: A Tool for Baseball Game Reconstruction & Visualization

Analyses of baseball games has increases exponentially in recent times. A major reason for this is believed to be the success of Oakland Athletics in 2002, who made an improbably playoff run with a team with one of lowest salaries. A major part of this success was the use of statistics to find players who were usually ignored by other teams.

This paper focusses on the developed of a new visualisation tool – Baseball4D. The tool provides functionalities like ball tracking, player tracking, and pitch tracking. In addition, it also shows how these visualizations can be interactively explored and how new metrics for fielding and base running can be derived from this information. The application provides gameplay that combines pitching, hitting, running, and fielding information, putting spatial and temporal data in a unified format. The database used to collect this data is the Major League Baseball Advanced Media (MLBAM) Gameday server.

Besides this, the tool also represents spray charts and heat maps and player positions, scatter plots to analyse catching based on how far the ball was from the fielder and bar charts to represent difficulty of balls handled by infielders.

While the tool in itself provides lot of useful insights, there is still a possibility of further advances. This includes adding more metrics to the system, and also developing ways to depict them in an intuitive way. The tool can also provide complete coverage of a baseball season rather than just individual matches. Handling this volume of data will require new querying methods and changes in interactivity of the visualisations.

3.1.11 Visualization of wearable sensor data during swimming for performance analysis

We realised, that besides the online databases, there is another source of data that can be used to analyse the athletes. This is the data collected by the wearable technologies. This paper therefore focuses on swimming performance monitored by sensors can be used to provide feedback that coaches and athletes can understand. The validity of the process and the data collection was validated by comparing the data of a casual swimmer vs. that of a former Olympian.

Once the data was collected, 5 visualisations were created using Matlab. These were:

- A time series plot
- A time series overlay plot – Consisted of plotting each individual stroke on the same plot which overlays the strokes allowing direct comparison
- Two phase space plots – The phase space portraits are parametric plots with one plot employing the derivative technique and the other plot employing an embedded delay technique
- A ribbon plot – 3Dplot that showed each individual stroke as a ribbon clearly delineating each stroke
- A wavelet scalogram – Applied to an entire lap to produce a 2D scalogram.

The comparison of the 2 datasets showed that all the methods were able to give useful information about the consistency of the stroke cycle and show the difference between the two. This research can further be extended applying the techniques to a larger number of people and then comparing the results for this large dataset.

3.2 Gaps Identified

The research papers analyzed aren't just limited to football. We have analyzed the use of data visualisation in several different sports like Tennis, Basketball, Ice hockey, Baseball and so on. This has been done to understand the usage of data visualisation in this field and help us best determine the data viz idioms that will best suit our project.

However, while going through these research papers, we identified some gaps that we planned on filling. The first major drawback we noticed in these papers was that most of them used expensive equipment like high quality video cameras to record data, especially trajectory motion of the player and the ball. This limits the use of such software and projects to manager and analysts who can afford the equipment. We wanted to make a project that can be used by everyone, by making use of readily available datasets.

The limitations of the results for some of these projects was also in terms of the statistical format of the results. It is difficult to make insights from statistical data and expert analysts might be required to analyse such data.

Three of the research papers mentioned above created apps that helped the user analyse the data. However, the interface and functioning of these apps was difficult to understand and required some prior training. We plan to avoid this problem by using a simple Menu driven

Program that can be used by anyone.

Finally, we also realized that research papers available were restricted to either one aspect of the game (ex: analyse passing) or focused on targeting one kind of user (ex: Managers). However, we wanted to create a project that combines several different aspects of this game – Analysis of teams, Contributions of players, Player stats and attributes, Fantasy XI, Game type analysis and provides it under 1 application. We also wanted to ensure that our project can be used by everyone ranging from football fans and mangers to scouts and fantasy league players. The functionality, interface and the visualizations chosen have thus been selected accordingly to ensure that everyone can understand the data easily.

3.3 Drive to the present work

Therefore the drive for this project was to fill in these gaps present in the current solutions. Both of us are also football fans and thus we would like to create such visualisations that help us analyse games and teams. We also wish to integrate all the visualisations and analysis so that the fans and managers can entire all different aspects of the game, rather than just focusing on one part. Finally, we also wish to ensure that we build an application that can be used by everyone ranging from football managers and scouts to ordinary football fans like us.

4. IMPLEMENTATION

4.1 Framework/Architecture

A) Visualisation 1: Predicting the final score line

This visualisation has been created to view the probabilities of different score lines and creating a heat map to view these probabilities. A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. In this case, the values will be the probability of that particular score line. Heatmaps are also a lot more visual than standard analytics reports, which can make them easier to analyse at a glance. This makes them more accessible, particularly to people who are not accustomed to analysing large amounts of data. This is especially important, given that most people utilising this data will be people who will be more concerned about the relative probability and the highest probable scores, rather than the individual probabilities.

B) Visualisation 2: Radial plot for player comparisons

The players in a team can be compared based on a range of different parameters, ranging from movement, shooting, defense, passing and pace. The visualisation chosen for this is radar graphs. The advantage of using a radar graph is that several different attributes can be compared at the same time. In this case, all the 5 attributes mentioned can be compared for several different players. This is more effective than creating subplots for bar charts/line charts as instead, try to use faceting: display as many chart as the number of individual, it makes easy to compare the shape of each. This also makes a job a manager easier. For example, if a manager is looking for a goal scorer, he simply has to look for players whose graph spikes at shooting and movement. Whereas, if he is looking for a defender, he will look for a player whose graph is spiking at defense.

C) Visualisation 3: Team Analysis

Another functionality provided by this project is that it helps managers and fans analyse an entire team and find out the players who have been the most crucial to the team's goals and general success. For this, a database for all player stats has been used. All the comparisons for goals, assists and appearances of a players within a team have been made plotted using a pie chart. A pie chart can be an effective communication tool for even an uninformed audience, because it represents data visually as a fractional part

of a whole. Additionally, readers or audiences can also see data comparison at a glance, enabling them to make an immediate analysis or to understand information quickly. This type of data visualization chart removes the need for readers to examine or measure underlying numbers themselves, so it's a good way of presenting data that might otherwise appear in a table. This is especially useful for our project, as football fans might not be able to compare the contributions themselves. It also helps us emphasize the importance of a player in a club visually.

D) Visualization 4: Creating Dream 11 and choosing and visualizing formation

This visualization is created so that the manager can decide how is starting 11 should be. He is able to choose his own formation from a list and analyze the game plan. For example, if he wants an attacking team, he can choose 4-2-4 with 4 attackers. Another option available is that he can search for players by giving constraints in overall and position. The other 2 formations that are available are 4-3-3 and 5-3-2, the manager can choose the formation that he prefers and choose players accordingly. This visualisation is also useful for fantasy league players who can easily search for players based on the attributes they want and create their dream team.

E) Visualization 5 : Shot Map

This Visualisation is used for post-match analysis of a match and the manager can find from where the opposition team or their team took the shots from. This will be useful when the manager wants to see whether his players were able to break down the opposition defense and if the players were only depended on long shots. This visualisation can also be used by managers to analyse the opposition team. They can analyse from which side the teams attacks are more dominant and plan tactics accordingly.

This is better than current visualizations because the current ones show where goals were scored from, but not where all shots in the game were shot from.

The visualisation makes use of the map() function. This is the most suitable representation for the given data as it helps the manager understand the team’s play spatially.

F) Visualisation 6: Premier League Analysis:

The final visualisation tool created can be used by fans to view the consistency of their teams over the past few seasons of Premier League. For this a line graph has been selected. The line is the most effective visualisation here as the league rating of several different teams can be added to the same plot. The line use of line graph also enables the user to easily determine the league positions of each team in a particular year. Finally, the line graph highlights the trends and consistency of the teams over a long period of time.

4.2 Algorithm

A) Visualisation 1: Predicting the final score line

For creating this heatmap, the probabilities have been calculated using Poisson's Distribution. Poisson distribution was preferred over a Binomial distribution as Binomial distribution can only have 2 possible outcomes, whereas Poisson can have infinite outcomes. Therefore Binomial would have been useful to predict if a team 'would win or lose' or if a team would 'score a goal or not score a goal'. However, since the number of outcomes (score lines) is infinite, Poisson distribution is preferred. Moreover, Poisson is also preferred when the number of trials get very large and the probability of success is small, as evident in this scenario.

The objective here is to use the data from past matches of the two teams and competition in which there were playing to predict how many goals each team is expected to score. The 3 requirements for Poisson distribution are:

1. The event can be counted in whole number - The number of goals scored can be counted in whole number.
2. The occurrences are independent - The probability of one team scoring a goal is independent of the other team scoring.
3. The average frequency of occurrence (mean) is known - The mean has been calculated in further steps by using historic data of the two teams and the tournament in which they played.

We started by using the data from the league table of Premier League 2018-19 to calculate the average number of goals a team would score and concede in Home and Away games. The data was then used to calculate Average Goals scored =

$\frac{Goals\ Scored}{Matches\ Played}$ and average goals conceded $= \frac{Goals\ Conceded}{Matches\ Played}$.

This data was used to calculate attacking strength and defensive strengths of both teams. This was found because in a football match, the chances of a team scoring a goal depended on both - the attacking strength of the team who scores the goal, and the defensive strength of the team that concedes it. The formulas used were:

Liverpool Attacking Strength (Home)

$$= \frac{\text{Average Home Goals scored (Liverpool)}}{\text{Average Home Goals Scored (in Premier League)}} = \frac{2.89}{1.57} = 1.84$$

Arsenal Defenvisve Strength(Away)

$$= \frac{\text{Average Away Goals conceded (Arsenal)}}{\text{Average Away Goals Conceded (in Premier League)}} = \frac{1.84}{1.57} = 1.17$$

The following formula of Goal Expectancy was used to calculate the mean as it took into consideration 3 criteria – the statistics for a team scoring a goal, the opposing team conceding a goal and the average number of goals scored in the tournament.

$$\boldsymbol{Liverpool\ Goal\ Expectancy\ (\lambda)}$$

$$= Liverpool\ Attacking\ Strengh(Home)\ \times Arsenal\ Defensive\ Strength\ (Away)$$
$$\times \text{Average goals in Premier league (Home)}$$
$$= 1.84\ \times 1.17\ \times 1.57 = 3.38$$

$$\boldsymbol{Arsenal\ Goal\ Expectancy(\ \lambda)}$$

$$= Arsenal\ Attacking\ Strength\ (Away)\ \times Liverpool\ Defensive\ Strength\ (Home)$$
$$\times Average\ Goals\ in\ Premier\ league\ (Away)$$
$$= 1.304\ \times 0.66\ \times 1.252 = 1.078$$

The probability of different score lines was calculated then calculated using the Poisson distribution formula:

$\boldsymbol{p(X = x) = \frac{e^{-\lambda} \times \lambda^x}{x!}}$ **Where λ is the mean for distribution**

For example: Probability of Arsenal winning 2-0 = Probability of Arsenal scoring 2 x Probablity of Liverpool winning 0, wherein the mean (λ) value is the respective goal expectancy of the team as calculate above

$$= p(X = 2) \times p(X = 0)$$

B) Visualisation 2: Radial plot for player comparisons

The database used for project is the ratings and stats used from FIFA 19 (fifa19new.csv). The input for both the player names is taken from the user. The

rad=datarad[c('Name','Finishing','Dribbling','SprintSpeed','ShortPassing','StandingTackle')]

will define the different attributes that we will compare for the players. The rad function is used to add the attributes of the players based on the name entered by the user. Finally, the plot_ly function is used to plot the graph, which of the type Scatterpolar.

C) Visualisation 3: Team Analysis

The team analysis makes use of the Player Stats database for players that play in the Premier League. Each player has been sorted according to the team, and the details include their appearances, goals, assists, yellow cards and so on. The program starts by asking the user for the team they would like to analyse. Based on this, the players, appearances, goals and assists of that particular team are extracted from the database. To create the goals contribution plot first the total goals scored by the team is calculated, then all players who have scored more than 4 goals for the team is extracted separately (others are categorized under the 'others' group). The pie chart is plotted using the plot_ly where the plot type = 'pie', the labels are the player names and the values are goals scored by each player. The individual player names and goals are stored as a non-distributed vector. A similar process is followed to plot the assists and appearances of players for each team.

D) Visualization 4: Creating Dream 11 and choosing and visualizing formation

To visualize the formations the following steps were done:

i. First, a list of formation will be shown to the user. The user can enter what

formation they want

ii. Then a football field will be visualized using the annotate_pitch() function of the ggsoccer package.

iii. Along with the annotate_pitch the ggplot() function is used and the matrix of co-ordinates of the points of player position is kept as a parameter.

iv. The geom_point() function is used to print the points and give attributes for the point

To make the dream 11

There are 2 methods available: Manual entry and Advanced search

i. Manual entry is executed by looping 11 times and the user can add players to their team by manually entering their name. When the name is entered it will search the dataset and bind the details to the team table.

ii. In Advanced entry, the manager can set minimum overall and also search by position. After the search from dataset is done it will display a list of players satisfying the conditions, this is done by creating an empty data frame and binding the details of the valid players row by row. From this list the manager can select the player he wants by entering his name. After choosing the 11 players, mean analysis is done

Average overall = sum of overall of 11 players /11

E) Visualization 5: Shot Map:

To visualize the shots taken the following steps were carried out:

i. First, a list of matches will be shown to the user. The user can enter what formation they want

ii. Then a football field will be visualized using the annotate_pitch() function of the ggsoccer package.

iii. Along with the annotate_pitch the ggplot() function is used and the matrix of co-ordinates of the points of shot position is kept as a parameter.

iv. The geom_point() function is used to print the points and give attributes for the

point

v. The pitch is flipped and cut into half so as to get a better experience for the user

F) Premier League Analysis:

The final visualisation tool created can be used by fans to view the consistency of their teams over the past few seasons of Premier League. For this the league positions were first extracted from the league past Premier league table. The data was then plotted as line graph wherein each teach was assigned a different color to easily differentiate the positions of different teams.

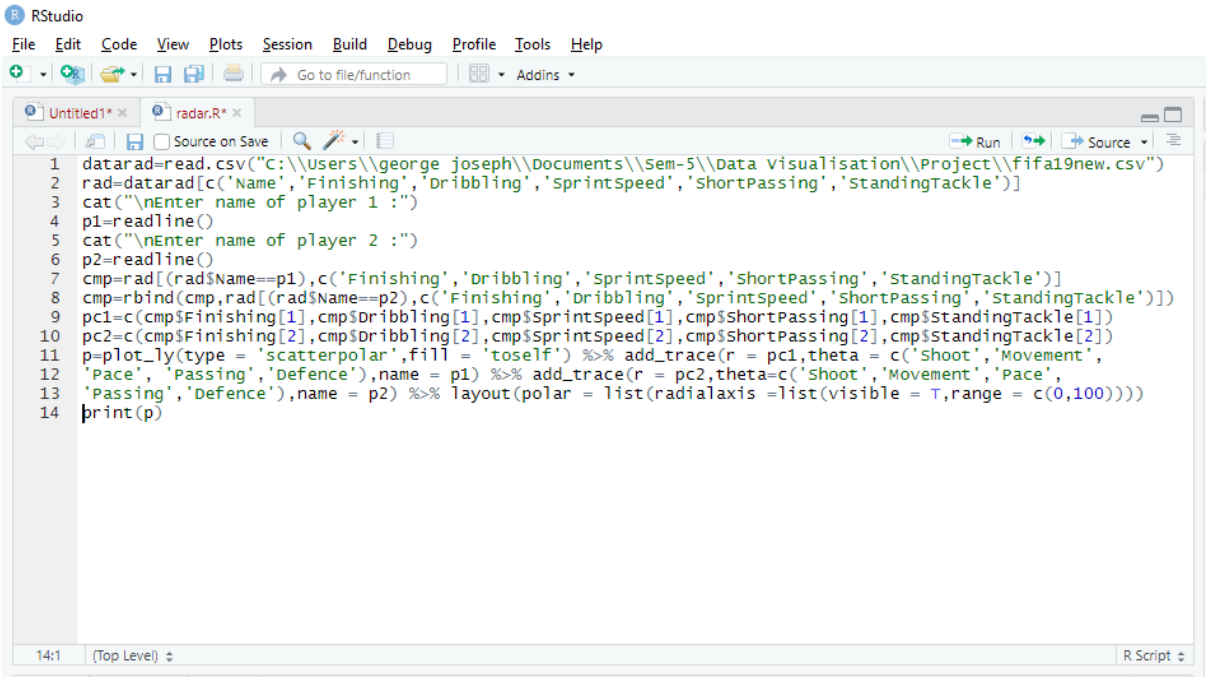
4.3 Complexity Analysis

4.4 Program

A) Visualisation 1: Predicting the final score line

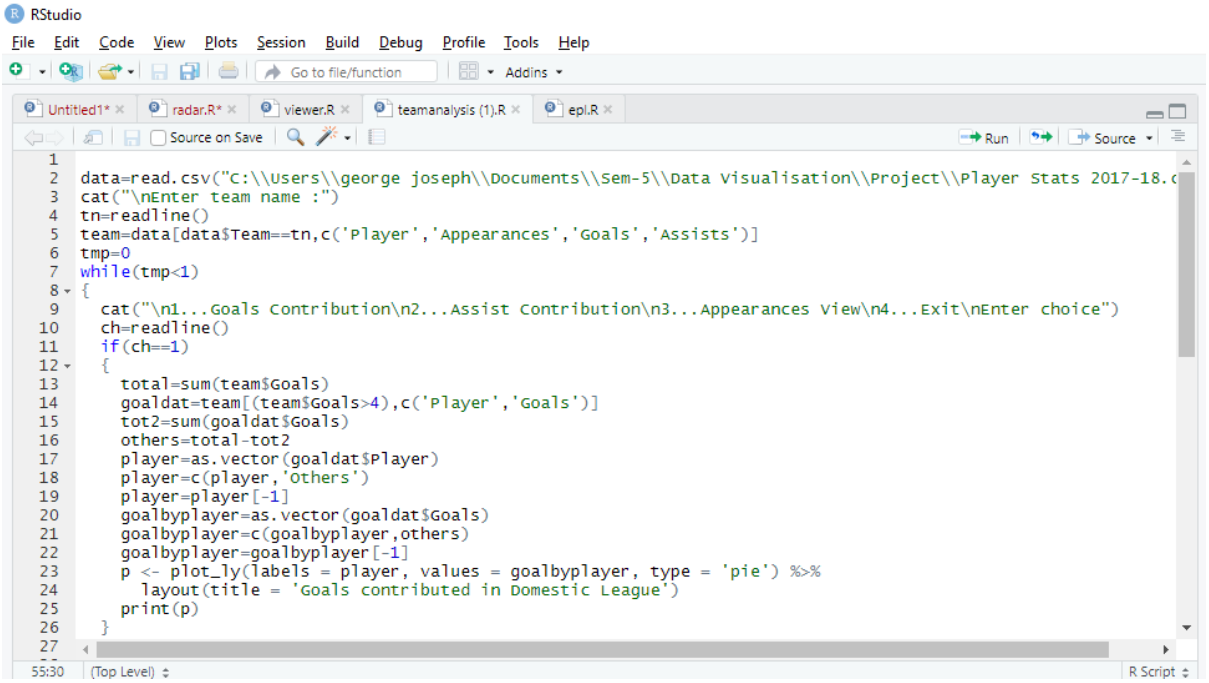
```
Python 3.7 (32-bit)
Python 3.7.3 (v3.7.3:ef4ec6ed12, Mar 25 2019, 21:26:53) [MSC v.1916 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> import seaborn as sb
>>> import matplotlib.pyplot as plt
>>> df = pd.read_csv("E:\\VIT\\Year 3\\Data Visualisation\\Project\\heatmap.csv")
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'pd' is not defined
>>> import pandas as pd
>>> df = pd.read_csv("E:\\VIT\\Year 3\\Data Visualisation\\Project\\heatmap.csv")
>>> print(df.head(10))
   0.01156  0.012478  0.006732  0.002414  0.000646  0.000136
0  0.03910    0.0422  0.022770  0.008165  0.002185  0.000460
1  0.06596    0.0712  0.038412  0.013774  0.003686  0.000776
2  0.07446    0.0804  0.043362  0.015549  0.004161  0.000876
3  0.06290    0.0679  0.036630  0.013135  0.003515  0.000740
4  0.04250    0.0459  0.024750  0.008875  0.002375  0.000500
5  0.02414    0.0261  0.014058  0.005041  0.001349  0.000284
>>> heat_map = sb.heatmap(df)
>>> plt.show()
>>> heat_map=sb.heatmap(data,xticklabels=False,yticklabels=False)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'data' is not defined
>>> heat_map=sb.heatmap(df,xticklabels=False,yticklabels=False)
>>> plt.xlabel("Goals Scored By Arsenal")
Text(0.5, 47.24444444444444, 'Goals Scored By Arsenal')
>>> plt.ylabel("Goals Scored by Liverpool")
Text(74.44444444444444, 0.5, 'Goals Scored by Liverpool')
>>> plt.show()
```

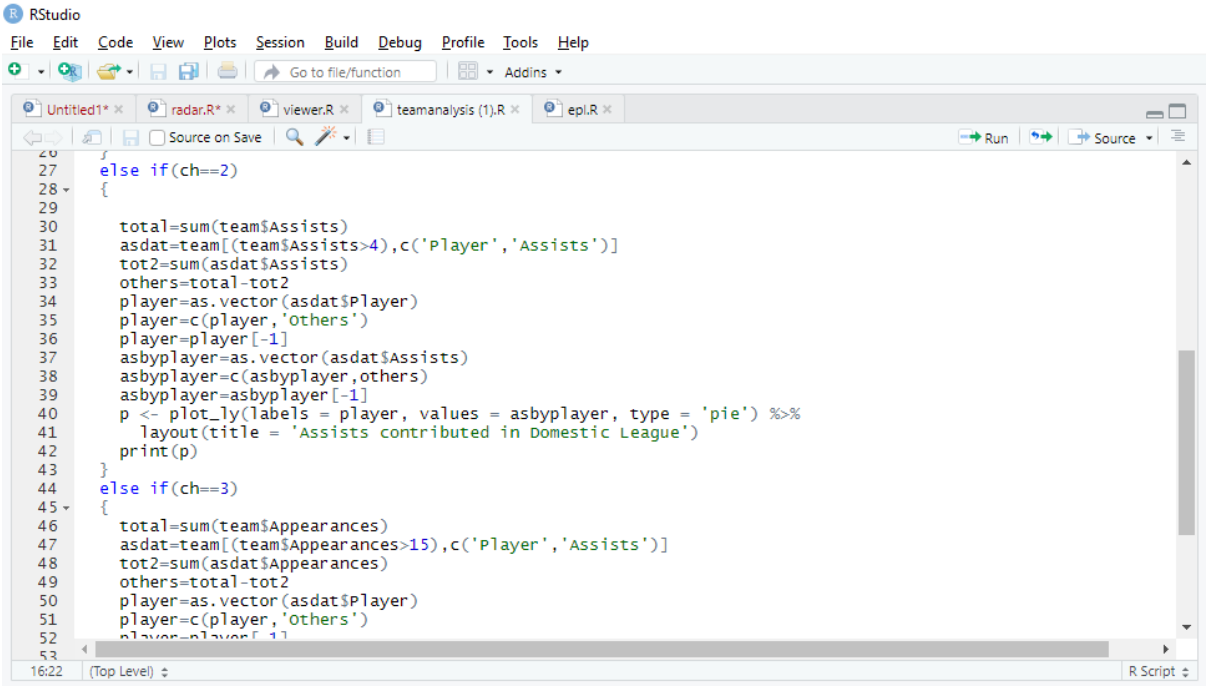
B) Visualisation 2: Radial plot for player comparisons

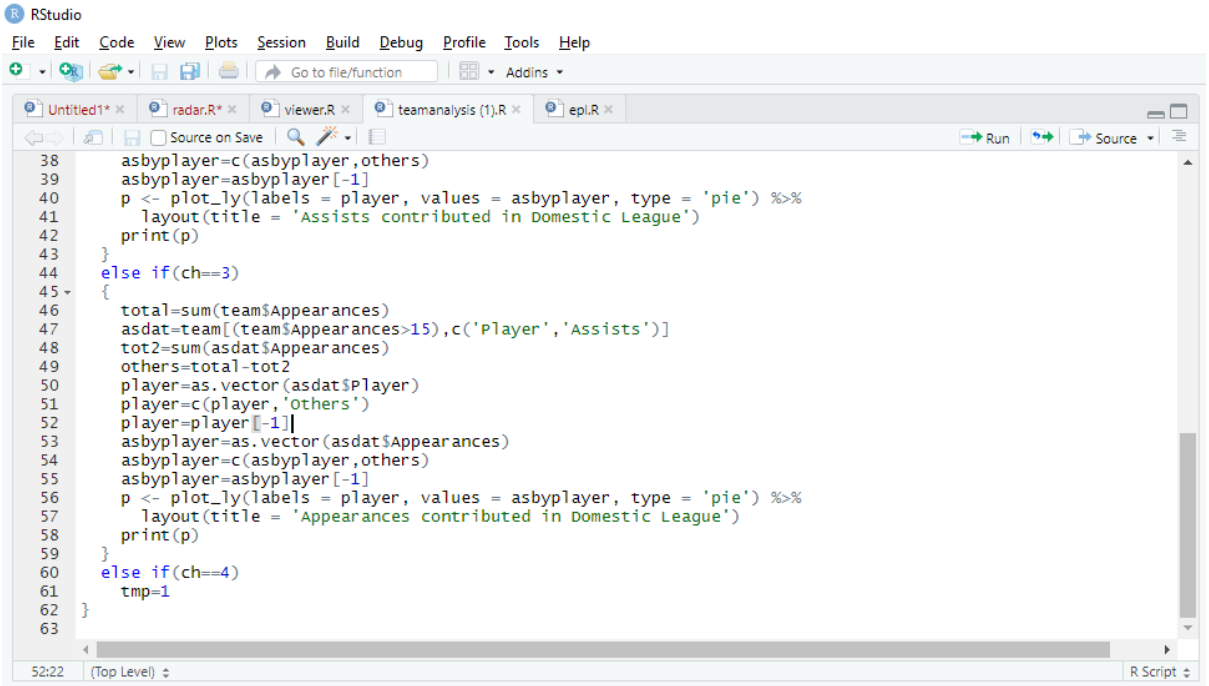


The dataset used for this program is from the Player ratings of these players. This dataset had 88 different columns for the attributes and 1697 rows for different players. Out of this, the 5 most attributes were selected.

C) Visualisation 3: Team Analysis

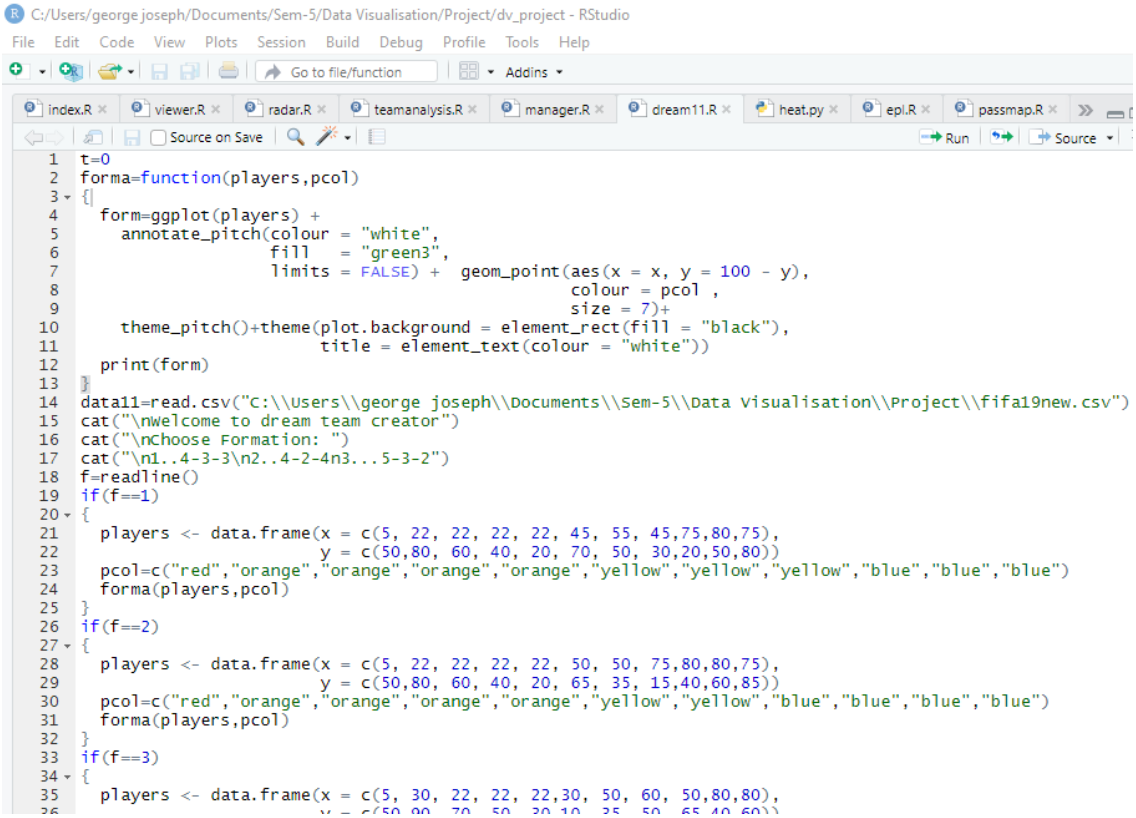




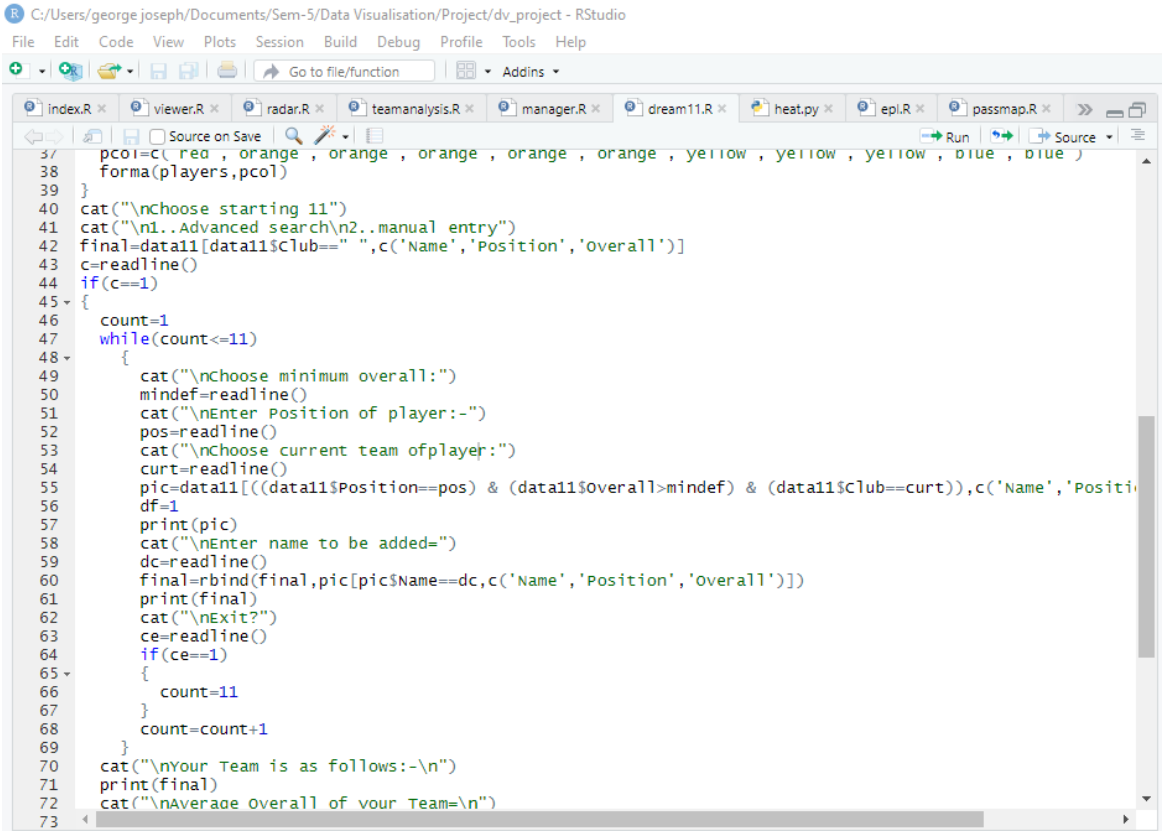


The dataset used for this program was from Player Stats of the Premier League table. This table consisted of 57 different columns which represent the attributes and data for 4187 players across the 20 teams.

D) Visualization 4: Creating dream 11 for managers and choosing and visualizing formation

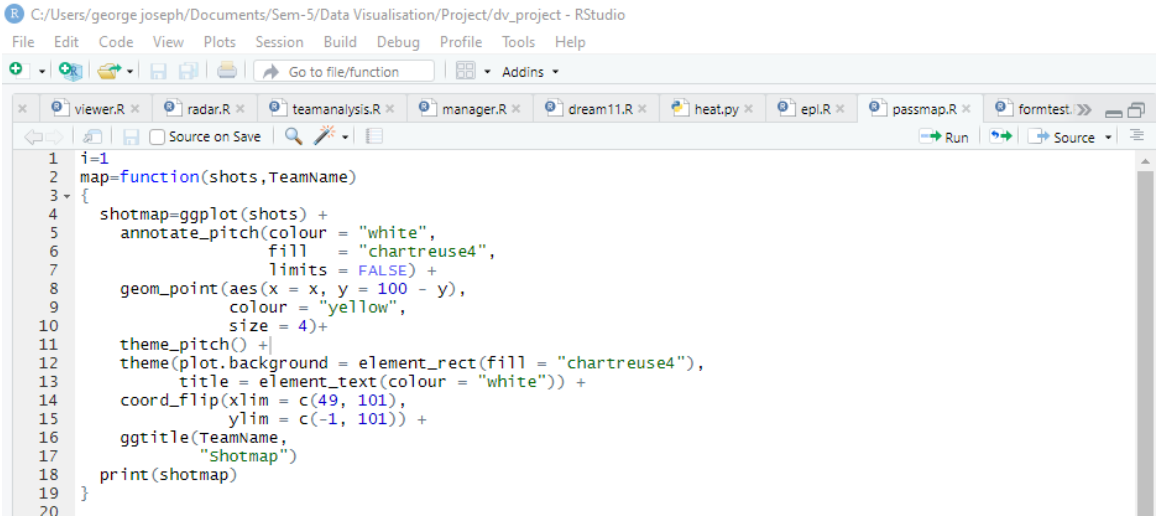


```
t=0
forma=function(players,pcol)
{
  form=ggplot(players) +
    annotate_pitch(colour = "white",
                   fill   = "green3",
                   limits = FALSE) +  geom_point(aes(x = x, y = 100 - y),
                                                 colour = pcol ,
                                                 size = 7)+
    theme_pitch()+theme(plot.background = element_rect(fill = "black"),
                        title = element_text(colour = "white"))
  print(form)
}
data11=read.csv("C:\\Users\\george joseph\\Documents\\Sem-5\\Data Visualisation\\Project\\fifa19new.csv")
cat("\nWelcome to dream team creator")
cat("\nChoose Formation: ")
cat("\n1..4-3-3\n2..4-2-4n3...5-3-2")
f=readline()
if(f==1)
{
  players <- data.frame(x = c(5, 22, 22, 22, 22, 45, 55, 45,75,80,75),
                        y = c(50,80, 60, 40, 20, 70, 50, 30,20,50,80))
  pcol=c("red","orange","orange","orange","orange","yellow","yellow","yellow","blue","blue","blue")
  forma(players,pcol)
}
if(f==2)
{
  players <- data.frame(x = c(5, 22, 22, 22, 22, 50, 50, 75,80,80,75),
                        y = c(50,80, 60, 40, 20, 65, 35, 15,40,60,85))
  pcol=c("red","orange","orange","orange","orange","yellow","yellow","blue","blue","blue","blue")
  forma(players,pcol)
}
if(f==3)
{
  players <- data.frame(x = c(5, 30, 22, 22, 22,30, 50, 60, 50,80,80),
```

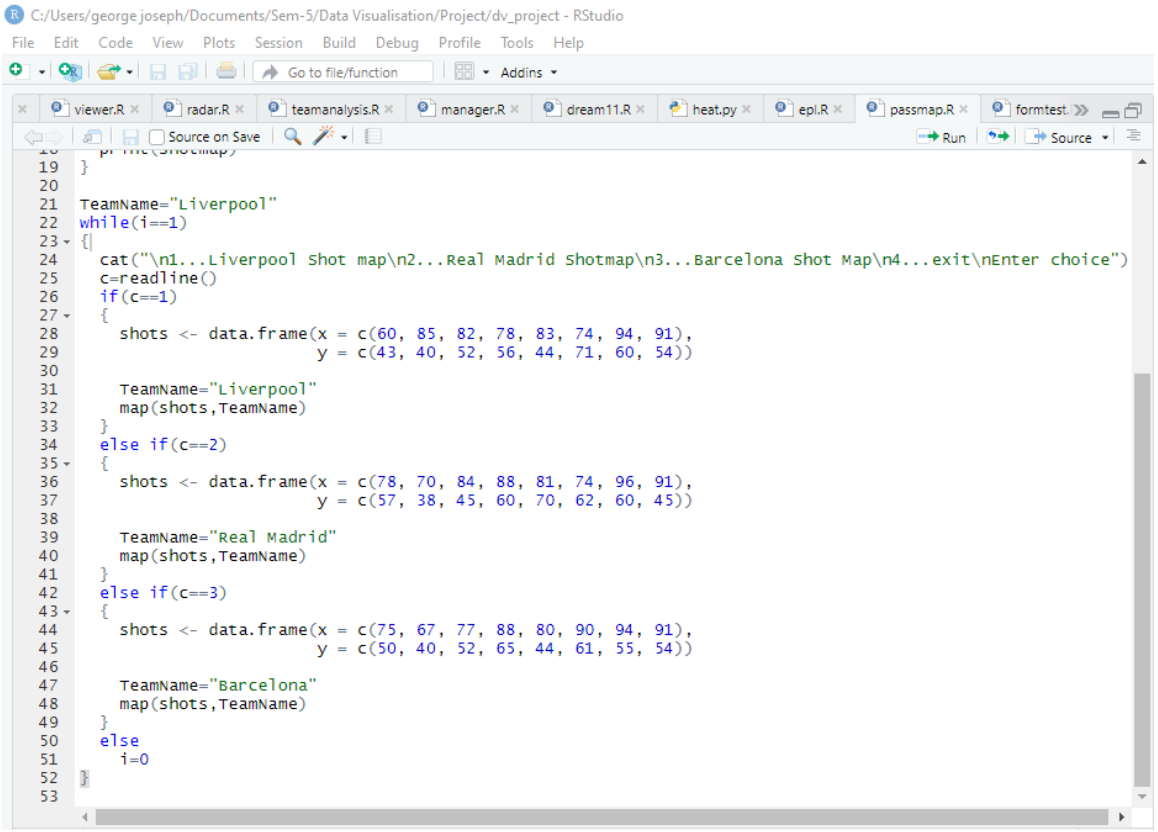


```
  pcol=c("red","orange","orange","orange","orange","yellow","yellow","yellow","blue","blue")
  forma(players,pcol)
}
cat("\nChoose starting 11")
cat("\n1..Advanced search\n2..manual entry")
final=data11[data11$Club==" ",c('Name','Position','Overall')]
c=readline()
if(c==1)
{
  count=1
  while(count<=11)
    {
      cat("\nChoose minimum overall:")
      mindef=readline()
      cat("\nEnter Position of player:-")
      pos=readline()
      cat("\nChoose current team ofplayer:")
      curt=readline()
      pic=data11[((data11$Position==pos) & (data11$Overall>mindef) & (data11$Club==curt)),c('Name','Positi
      df=1
      print(pic)
      cat("\nEnter name to be added=")
      dc=readline()
      final=rbind(final,pic[pic$Name==dc,c('Name','Position','Overall')])
      print(final)
      cat("\nExit?")
      ce=readline()
      if(ce==1)
      {
        count=11
      }
      count=count+1
    }
  cat("\nYour Team is as follows:-\n")
  print(final)
  cat("\nAverage Overall of your Team=\n")
```

E) Visualization 5: Shot Map




```r
i=1
map=function(shots,TeamName)
{
  shotmap=ggplot(shots) +
    annotate_pitch(colour = "white",
                   fill   = "chartreuse4",
                   limits = FALSE) +
    geom_point(aes(x = x, y = 100 - y),
               colour = "yellow",
               size = 4)+
    theme_pitch() +
    theme(plot.background = element_rect(fill = "chartreuse4"),
          title = element_text(colour = "white")) +
    coord_flip(xlim = c(49, 101),
               ylim = c(-1, 101)) +
    ggtitle(TeamName,
            "Shotmap")
  print(shotmap)
}
```




```r
}

TeamName="Liverpool"
while(i==1)
{
  cat("\n1...Liverpool Shot map\n2...Real Madrid Shotmap\n3...Barcelona Shot Map\n4...exit\nEnter choice")
  c=readline()
  if(c==1)
  {
    shots <- data.frame(x = c(60, 85, 82, 78, 83, 74, 94, 91),
                        y = c(43, 40, 52, 56, 44, 71, 60, 54))

    TeamName="Liverpool"
    map(shots,TeamName)
  }
  else if(c==2)
  {
    shots <- data.frame(x = c(78, 70, 84, 88, 81, 74, 96, 91),
                        y = c(57, 38, 45, 60, 70, 62, 60, 45))

    TeamName="Real Madrid"
    map(shots,TeamName)
  }
  else if(c==3)
  {
    shots <- data.frame(x = c(75, 67, 77, 88, 80, 90, 94, 91),
                        y = c(50, 40, 52, 65, 44, 61, 55, 54))

    TeamName="Barcelona"
    map(shots,TeamName)
  }
  else
    i=0
}
```

F) Premier League Analysis

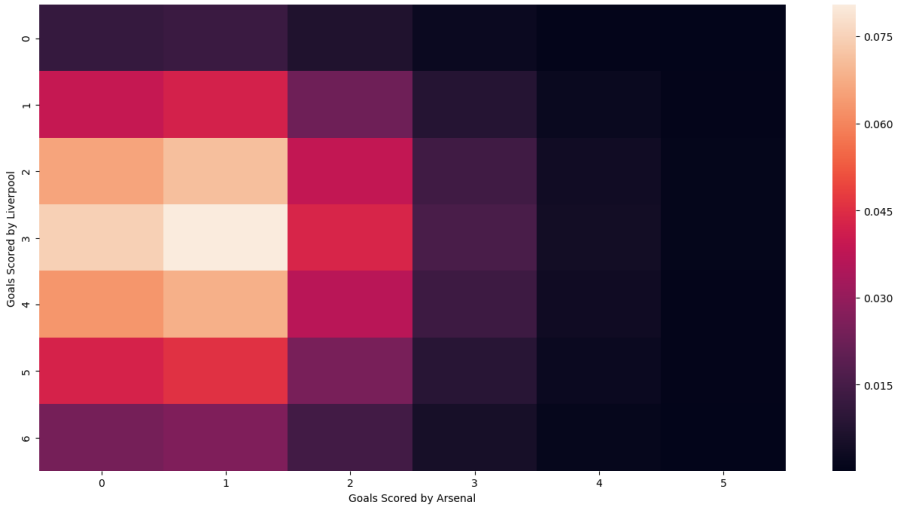
RGui (64-bit) - [C:\Users\samsung1\Downloads\epl.R - R Editor]

File Edit Packages Windows Help

```
leicester = c(1,12,9,9,4)
arsenal = c(2,5,6,5,3)
tottenham = c(3,2,3,4,9)
manchester_city = c(4,3,1,1,2)
manchester_united = c(5,6,2,6,12)
southampton=c(6,8,17,16,17)
westham=c(7,11,13,10,8)
liverpool=c(8,4,4,2,1)
chelsea=c(9,1,5,3,5)
season = c(2015,2016,2017,2018,2019)
plot(season,leicester,col="blue",ylim=rev(c(1,20)),type="l",ylab="League Position",xlab="Seaon")
lines(season,arsenal,col="red",ylim=rev(c(1,20)),type="l")
lines(season,tottenham,col="purple",ylim=rev(c(1,20)),type="l")
lines(season,manchester_city,col="brown",ylim=rev(c(1,20)),type="l")
lines(season,manchester_united,col="orange",ylim=rev(c(1,20)),type="l")
lines(season,southampton,col="black",ylim=rev(c(1,20)),type="l")
lines(season,westham,col="green",ylim=rev(c(1,20)),type="l")
lines(season,liverpool,col="gray",ylim=rev(c(1,20)),type="l")
lines(season,chelsea,col="cyan",ylim=rev(c(1,20)),type="l")
teams=c("Leicester","Arsenal","Tottenham","Manchester City","Manchester United","Southampton","Westham","Liverpool","Chelsea")
legend("bottomleft",teams,cex=0.8,fill=c("blue","red","purple","brown","orange","black","green","gray","cyan"))
```

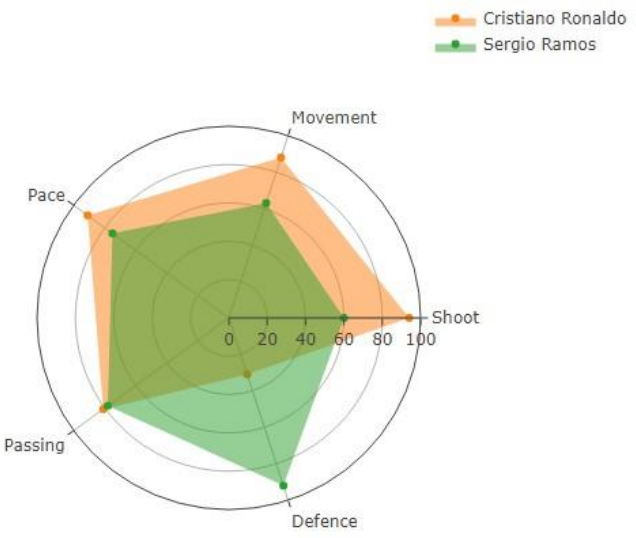
5. RESULT ANALYSIS

A) Visualisation 1: Predicting the final score line



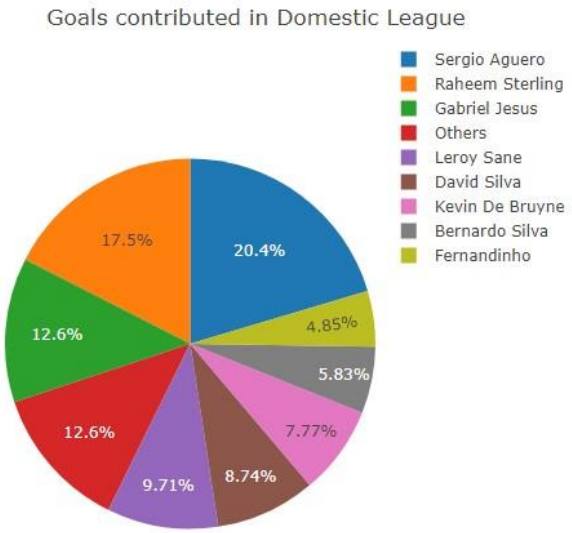
The heatmap above uses different hues and gradients to show the possibilities of different score lines. For example, the brightest box in the center shows the highest probable score line: Liverpool (3) – Arsenal (1). The darker the box, the lower the possibility of that score line. Therefore, analysis like these can be used by fans who would like to predict the outcome of the game. This can also be used by fantasy league players, for example for the following game, the players will avoid having Arsenal defenders in their team as they might concede 3 goals and will instead put Liverpool attackers.

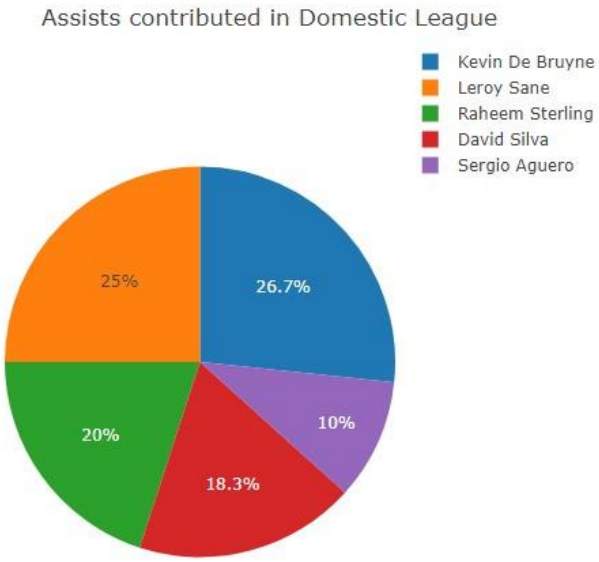
B) Visualisation 2: Radial plot for player comparisons



The radar plot above shows the attributes of 2 players: Cristiano Ronaldo and Sergio Ramos. The 2 sets of data are differentiated by using different colors. The use of radar plot helps us add more such players on the same graph and compare them easily. This visualization is most suitable for managers and scouts while looking for new players. For example, if a manager is looking for a player that is fast and has good shooting, they will prefer Ronaldo, whereas a manager looking for someone who is good at defense will pick Ramos.

C) Visualisation 3: Team Analysis





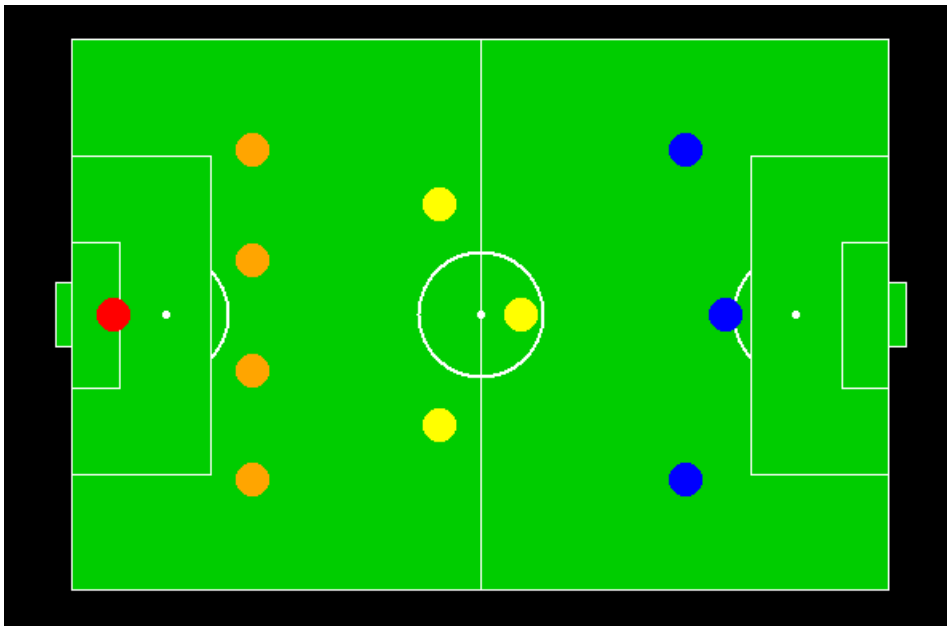
The first pie chart shows the Goal Contribution and the second pie chart shows the Assists Contribution for Manchester City in the domestic league. The use of pie chart quickly helps

the user realize that Sergio Aguero, Raheem Sterling and Gabriel Jesus contribute for more than half of the clubs goals. Managers can also use this data, to ensure that they ask their defenders to be vary of these 3 players. The 2nd chart shows that Leroy Sane and Kevin De Bruyne are the most important players for the club in terms of goals created. The pie chart makes it really easy to make these comparisons and analyze who are the most important players for the club. For example, if the same data was provided in terms of table:

Player	**Goals**
Bernardo Silva	6
Danilo	3
David Silva	9
Fernandinho	5
Gabriel Jesus	13
Ilkay Gundogan	4
Kevin De Bruyne	8
Leroy Sane	10
Nicolas Otamendi	4
Raheem Sterling	18
Sergio Aguero	21

This table shows that Sterling and Aguero have scored a lot of goals, but a person cannot visualize that Aguero + Jesus + Sterling have scored more than half the clubs goals at first glance. This analysis is made much easier with the help of the pie chart.

D) Visualisation 4: Creating dream 11 for managers and choosing and visualizing formation



The map shown above will help the manager analyse how his team will play spatially.

```
Welcome to dream team creator
Choose Formation:
1..4-3-3
2..4-2-4n3...5-3-2
1

Choose starting 11
1..Advanced search
2..manual entry
1

Choose minimum overall:
80

Enter Position of player:-
ST


Choose current team ofplayer:
Real Madrid
             Name Position Overall
31           Isco       ST      88
37         G. Bale      ST      88
80  Marco Asensio       ST      85
106    K. Benzema       ST      85
173 Lucas Vázquez       ST      83

Enter name to be added=
G. Bale
      Name Position Overall
37 G. Bale       ST      88
```

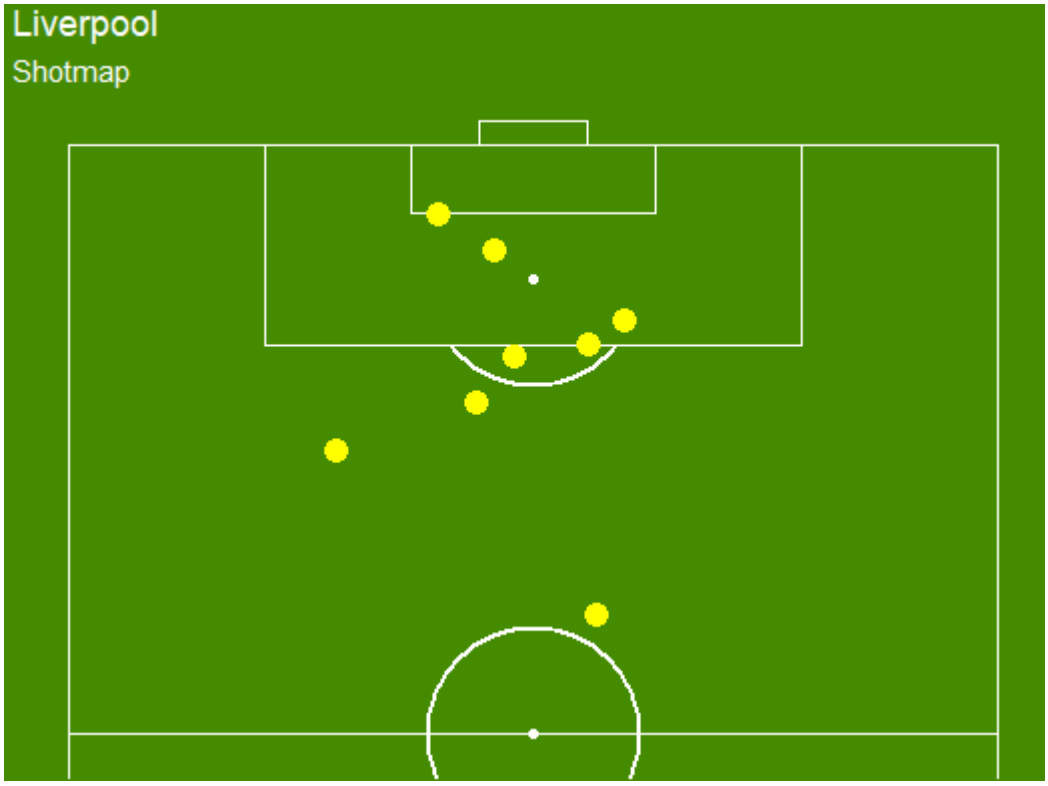
After choosing the 11 players the team is as follows:

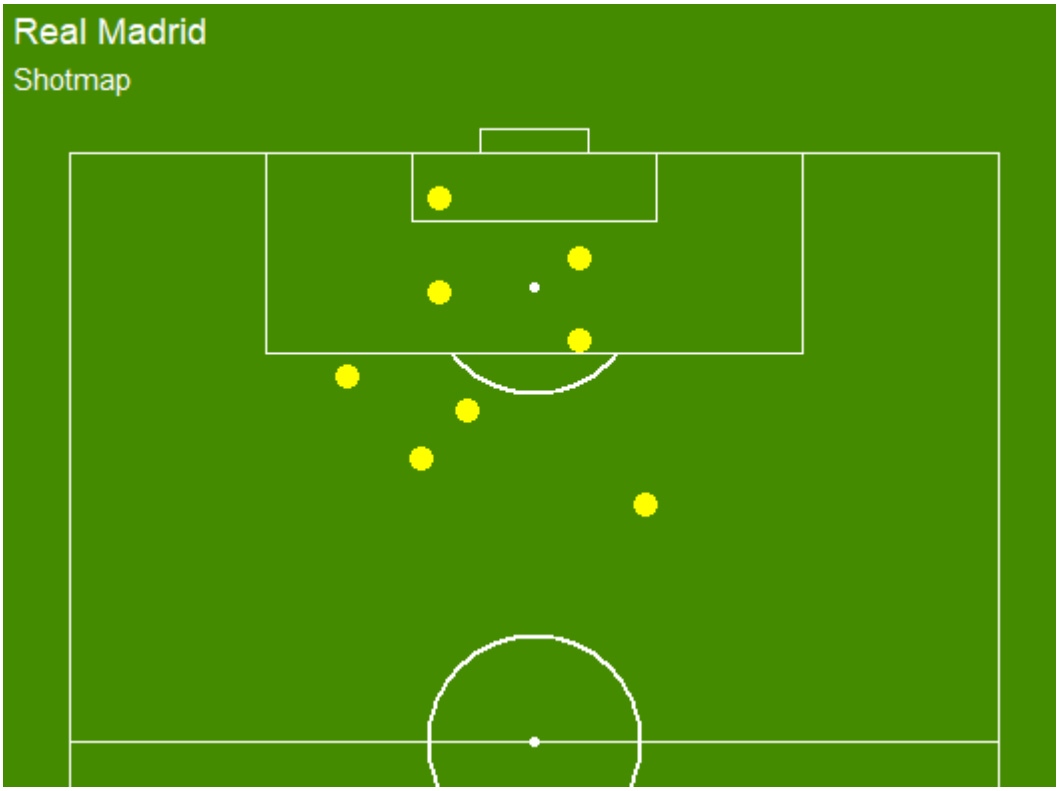
Name	Position	Overall
G. Bale	ST	88
E. Hazard	ST	91
Cristiano Ronaldo	ST	94
T. Kroos	CM	90
P. Pogba	CM	87
J. Rodriguez	CAM	88
D. Godin	CB	90
Carvajal	RB	84
Jordi Alba	LB	87
J. Oblak	GK	90

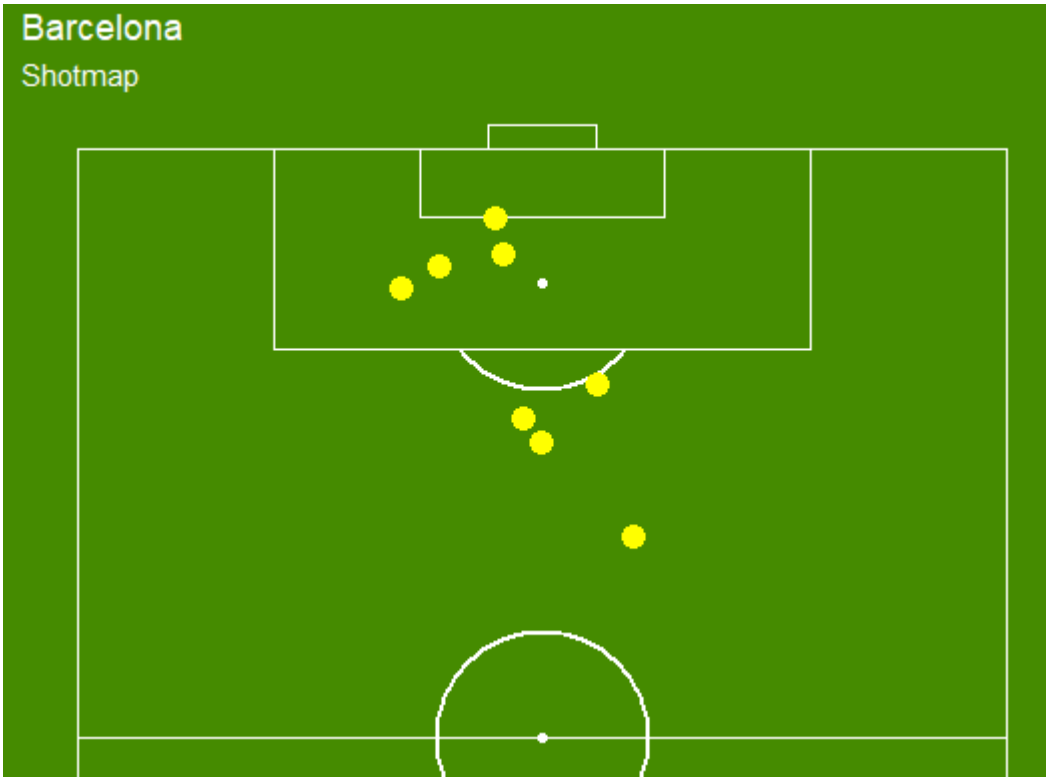
This visualisation helps the fantasy league players as they can search players easily based on the overall rating of players and their teams. This is especially helpful for inexperienced users who won't be able to search through the entire dataset of players available.

E) Visualization 5: Shot Map

```
1...Liverpool Shot map
2...Real Madrid Shotmap
3...Barcelona Shot Map
4...exit
Enter choice
1
```



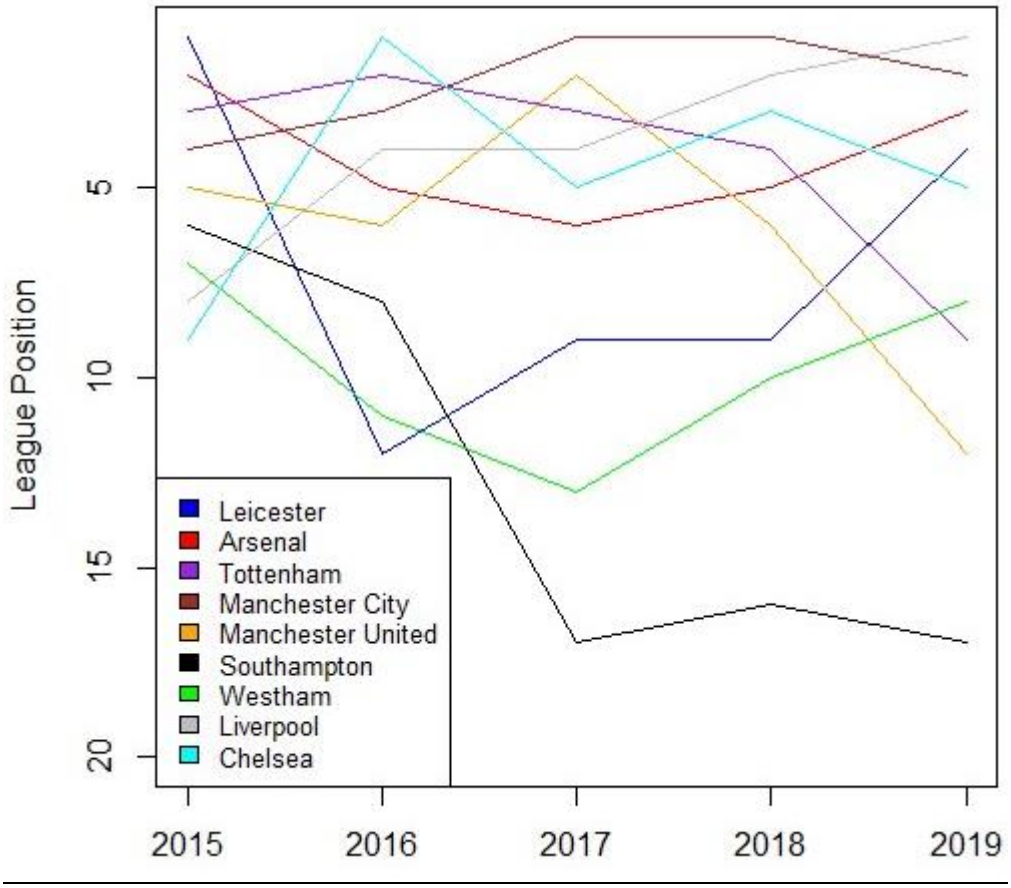




The shot map is a very effective visualisation tool for the managers to understand the tactics of their team and the opposing team. For example, several insights can be made by comparing the shot maps of Real Madrid and Barcelona. For example, Real Madrid take most of their long shots from the left side of field while Barcelona take their long

shots from right side of their pitch. Additionally, we can see that Barcelona have a tendency close from the left hand side of the field. A comparison between Liverpool and Barcelona can also be made as Liverpool take most of their shots from inside the box, whereas Barcelona take a lot of their shots from outside the box.

F) Premier League Analysis:



The line chart helps the user make useful insights easily regarding the consistency of teams. For example, Chelsea won the league in 2016, but after this year their performance dipped. Whereas the performances of Southampton have been constantly falling. The Manchester City fans on the other hand can see that their team has been very consistent over the past few years. The line graph helps highlight all these trends effectively.

6. FUTURE WORK

The project can be further improved to add more functionalities. For example, the project uses only historically recorded data. This can be extended by adding more dynamic data recording live during games or adding video data (trajectory and player movements of different players). Another example of this could be getting data from wearable technologies used by the players. The prediction model can also be improved by making it dynamic, wherein the probability of the game and the final score changes after every goals scored. The project can also analyse further aspects of the game, like set pieces like free kicks and penalties. Finally, the project can also be extended to help other people like physicians who will be interested in the health data of the football players or goalkeepers who want to understand the probabilities of a person taking a penalty to his right. As seen in the research papers, dynamic and interactive visualisations can also be made for football events like the FIFA World Cup.

7. REFERENCES

1. Polk, Tom et al. "Tennivis: Visualization For Tennis Match Analysis". IEEE Transactions On Visualization And Computer Graphics, vol 20, no. 12, 2014, pp. 2339-2348. Institute Of Electrical And Electronics Engineers (IEEE), doi:10.1109/tvcg.2014.2346445. Accessed 9 Sept 2019.

2. Basole, Rahul C., and Dietmar Saupe. "Sports Data Visualization [Guest Editors' Introduction]". *IEEE Computer Graphics And Applications*, vol 36, no. 5, 2016, pp. 24-26. *Institute Of Electrical And Electronics Engineers (IEEE)*, doi:10.1109/mcg.2016.85. Accessed 9 Sept 2019.

3. Pingali, G. et al. "Visualization Of Sports Using Motion Trajectories: Providing Insights Into Performance, Style, And Strategy". Proceedings Visualization, 2001. VIS '01., 2001. IEEE, doi:10.1109/visual.2001.964496. Accessed 9 Sept 2019.

4. Cox, Andy, and John Stasko. "Sportvis: Discovering Meaning In Sports Statistics Through Information Visualization". College Of Computing / GVU Center Georgia Institute Of Technology, Accessed 9 Sept 2019.

5. Mitchell Page, Andrew Vande Moere. "Towards Classifying Visualization In Team Sports". International Conference On Computer Graphics, Imaging And Visualisation (CGIV'06). IEEE, doi:10.1109/cgiv.2006.85. Accessed 9 Sept 2019.

6. Pileggi, Hannah et al. "Snapshot: Visualization To Propel Ice Hockey Analytics". IEEE Transactions On Visualization And Computer Graphics, vol 18, no. 12, 2012, pp. 2819-2828. Institute Of Electrical And Electronics Engineers (IEEE), doi:10.1109/tvcg.2012.263. Accessed 9 Sept 2019.

7. Tani, Toshihiro et al. "Sports Play Visualization System For American Football". International Multiconference Of Engineers And Computer Scientists, vol 1, 2015, Accessed 9 Sept 2019.

8. Cintia, Paolo et al. "The Harsh Rule Of The Goals: Data-Driven Performance Indicators For Football Teams". 2015 IEEE International Conference On Data Science And Advanced Analytics (DSAA), 2015. IEEE, doi:10.1109/dsaa.2015.7344823. Accessed 9 Sept 2019.

9. Saito, H. et al. "Sports Scene Analysis And Visualization From Multiple-View Video". 2004 IEEE International Conference On Multimedia And Expo (ICME) (IEEE Cat. No.04TH8763). IEEE, doi:10.1109/icme.2004.1394492. Accessed 9 Sept 2019.

10. Dietrich, Carlos et al. "Baseball4d: A Tool For Baseball Game Reconstruction &Amp;Amp; Visualization". *2014 IEEE Conference On Visual Analytics Science And Technology (VAST)*, 2014. *IEEE*, doi:10.1109/vast.2014.7042478. Accessed 10 Sept 2019.

11. Rowlands, David D. et al. "Visualization Of Wearable Sensor Data During Swimming For Performance Analysis". *Sports Technology*, vol 6, no. 3, 2013, pp. 130-136. *Informa UK Limited*, doi:10.1080/19346182.2013.867965. Accessed 10 Sept 2019.

12. Moon, Bo, and Richard Brath. "Bloomberg Sports Visualization For Pitch Analysis". Accessed 10 Sept 2019.