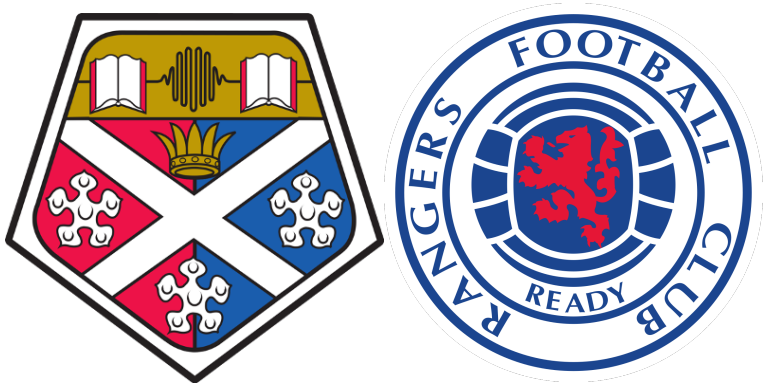
DEPARTMENT OF MANAGEMENT SCIENCE

University of Strathclyde Business School



Providing an Insight into the Usage of Data Visualisation Techniques for Tactical Analyses in Football

Ross McRae

201979395

MSc Data Analytics

Academic Supervisor: Matthew Hutcheson

University of Strathclyde

01/11/2020

1 Declaration

Except where explicitly stated, all the work in this dissertation - including any appendices - is my own and was carried out by me during my MSc course. It has not been submitted for assessment in any other context.

Signed: 

Print Name: Ross Andrew McRae

Date: 01/11/2020

2 Summary

Data analysis is an extremely important factor in gaining advantages over opposition teams in the footballing world. Football clubs use the area of data analysis in various different area - for example in physiological aspects and tactical analyses (Hill-Haas et al. 2011). In order to provide a comprehensive insight into such analyses, an open-sourced data-set was utilised. The aim was to provide a client with an insight into the use of data analysis in the footballing world, in order to give them a starting point to build upon.

Research was conducted in order to acumen information that would be best suited towards the clients needs. Initially, these needs were assumed to be an issue of comprehension as opposed to complexity of the subject area, so simpler areas of analyses were chosen to provide reasonable foundations for the clients knowledge that are able to be built upon.

Within the data testing section of this report, passmaps and shotmaps were the main focus alongside some initial exploratory analysis to provide information on the most common statistics that coaches will use to aid in their team picking and such like. These styles of statistics were initially explored to provide an insight into the best forms of passmaps and shotmaps to create to aid in the coaches decision making aspects. This gathered knowledge was then brought together in the form of an exemplar case study. This done to equip the client with an extensive and understandable beginning into the usage of data analyses for a specific match-day.

This report aimed to create a comprehensive and easily understood set of data visualisations and varying form of data analyses that could be utilised in both a short-term - match-day - and long-term - in terms of full season analyses for a multitude of factors. The client consisted of a coach at Rangers F.C, who had some resources for the purpose of data analyses but did not have the know-how as to how to best utilise this data for their own usage.

3 Acknowledgements

I would like to pass on my earnest gratitude to my academic supervisor - Matthew Hutcheson - who has provided much appreciated aid and direction, which has allowed me to complete this project to the best of my ability.

A special thank you to my Mum and Gran for all the support, advice and encouragement you have gave me over the entirety of my academic career as it is wholeheartedly appreciated - as have all the cups of tea and coffee, which have gone down an absolute treat.

Finally, thank you to my friends - Erin and David specifically - who alongside my family have had to deal with the absolute ball of stress that I became throughout my entire academic career. A special mention also to the boys - Connor, Will, Graeme and Cameron - thank you for all those gaming nights which allowed me to de-stress and relax as I don't think I could've completed this project without them.

4 Ethics

This dissertation initially had involved the use of sensitive data provided by Rangers Football Club - due to unforseen circumstances, the project had to be altered due to the inability to access the data promised. As part of the project - both initially and currently - an interview was carried out with Mr. Michael McKenna, who provided valuable insight into the world of footballing analytics and his permission was granted to use such information as part of this report.

The storage of this data or information, is the most likely to incur an ethical issue. An ethics form and participant guidance sheet was provided to the participant, in order to detail the required information about the storage of data and so on. An example of this participant guidance sheet is detailed on the next page.

The dataset used for analysis within this project is a public dataset provided by the authors of "A Public Data Set of Spatio-Temporal Match Events in Soccer Competitions" (Pappalardo, Cintia, Rossi, Massucco, Ferragina, Pedreschi & Giannotti 2019). This meant that no personal information would be found within the dataset and therefore complete anonymity would be achieved.

DEPARTMENT OF MANAGEMENT SCIENCE

University of Strathclyde Business School

Participant Information Sheet

MSc Project of Ross McRae

You are being invited to attend an interview conducted by Ross McRae, on behalf of the current MSc Project ran by the Department of Management Science at the University of Strathclyde. This information sheet is provided to you in order to provide insight as to what you will be doing as part of this interview.

The aim of this interview is to provide expert judgement to be utilised in the design of a generalised template for use of data analysis and visualisation within sports analysis – particularly in the fields of grassroots and women's football.

The interview should last no more than 1 hour. The interview will be conducted through the use of Zoom, where the session will be recorded – subject to confirmation from participant. This recording will be kept until the conclusion of the project, after such time it will be deleted.

This interview is voluntary and as such you may decide to withdraw at any time without providing any explanatory reason. In the case that this is exercised, all data or information which has been provided will be withdrawn and destroyed.

The data and/or information will be utilised in order to best determine a suitable design for use in data visualisation and analysis of sports datasets. The data will be kept confidential and will be stored on the University OneDrive and will be deleted once the project has been completed. The information provided will be used alongside your name and position in order to best classify the expert judgement – this may be removed depending on confirmation of participant.

This study will be conducted by Ross McRae, MSc Data Analytics student at the University of Strathclyde, as part of his final year project. The project is being supervised by Mr Matthew Hutcheson – a teaching associate also at the University of Strathclyde.

If you have any questions about this study, do not hesitate to contact either one of us:

Ross McRae – Email: ross.mcrae.2019@uni.strath.ac.uk

Mr Matthew Hutcheson – Email: matthew.hutcheson@strath.ac.uk

Contents

List of Figures

Project Context

1 Client Background

This study aims to delve into the use of data analytics in football, specifically for Rangers W.F.C - or Rangers Women's Football Club.

Rangers Football Club is a professional football club established in Govan in March 1872 - which makes them the joint fourth oldest Scottish football club alongside Dumbarton FC and sitting behind Stranraer, Kilmarnock and Queen's Park. Rangers FC are native to Glasgow and play their football in Ibrox Stadium which is so named from the district of residence situated in Govan (BBC 2012).

Rangers are best known for being the second most successful club in terms of decoration - this includes the likes of league titles and domestic cups, for example - behind on Al Ahly, who play their football in the Egyptian Premier League. They sport a wealthy trophy cabinet consisting of 54 League Titles, 33 Scottish Cups and 27 Scottish League Cups. They also boast the significant feat of achieving 7 domestic trebles - with a domestic treble being winning the League Title and both respective domestic cups all within the same season (UEFA.com n.d.).

In 2008, Rangers FC developed their respective female team, so named Rangers Women's Football Club - originally named "Ranger's Ladies." Rangers WFC currently play at the highest level of women's football in Scotland (*Rangers Womans Team – Rangers Football Club – Info and club news* 2008).

2 Project Background

The aim of this collective group effort was to display an open source of data, which was readily available for any sue within data analysis. Open-source data is not often freely accessed, due to the ownership of afore mentioned datasets being privatised - whether this be freelance companies or the individual football clubs.

Data analytics is a hastily growing area within the footballing world, and has became more significant in recent years. For years, the reliance of winning football games was put on the amount of effort and passion that every player gave in the respective games. Although passion and effort still play a hugely important part in what determines the outcome of a football match, data analytics in varying forms have became much more prominent in the recent era (Memmert 2019).

There are a multitude of areas where data analysis is used - ranging from tactical analysis to physical analysis for example. Tactical analysis has became a highly important asset in the recent era of football. Tactical analysis is used to perform analyses of formations, set-pieces and passing networks to but name a few (De Silva et al. 2018).

The client is very much aware of the power and knowledge to continue to progress in the modern era of football, that comes from the use of data analytics but seems to be fairly unsure as to how to harness it. The client had numerous varying ideas as to how they would like to approach this area and what they would like to gather and obtain as a result of the utilisation of data analytics.

Due to the limitation around the volume of data that the client had, a more novel approach was decided to be the best route to take this project for the client. After several conversations with the client, it was decided that the best use of the project would be to show the capability of data analysis, with the

knowledge of the type of data they could gather and the volume of such data. This then lead to the identification of a number of studies and open source datasets which can be used, and whose methods can be adapted and made accessible to the client.

3 Project Motivation

The main intention of this project is to provide a significant insight into the utilisation of data analytics to allow the client to keep progressing further forward into the modern era of football. This is turn, with the knowledge of the type of data and volume of data, has narrowed the project down to looking in depth into the usage of player passing networks for use in tactical analysis.

The aim is to make player passing networks more readily accessible to the client. This, so that, the client is able to utilise and harness their data to the extent that is available to them.

The secondary intention is to provide a translation from academic literature to the client. This is the aim, as the academic literature is of significant value to the analysts and coaching staff who are currently part of the infrastructure and included at Rangers. This is to allow the knowledge and information to be passed to those member of staff, who otherwise, would find this academic literature inaccessible.

4 Project Plan

Task name	Stard date	End date	Progres s	WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 5	WEEK 6	WEEK 7	WEEK 8	WEEK 9	WEEK 10	WEEK 11
PROJECT	**17/08/2020**	**01/11/2020**	**100%**											
Update Literature Review	17/08/2020	24/08/2020	**100%**											
Exploratory Analysis	24/08/2020	07/09/2020	**100%**											
Supervisor Meeting	28/08/2020	28/08/2020	**100%**											
Passmap Creation	07/09/2020	21/09/2020	**100%**											
Shotmap Creation	21/09/2020	05/10/2020	**100%**											
Supervisor Meeting	02/10/2020	02/10/2020	**100%**											
Exemplar Match Report	05/10/2020	19/10/2020	**100%**											
Write Dissertation	19/10/2020	01/11/2020	**100%**											

Figure 1: Updated Project Plan

The initial plan of the project, was to utilise a data-set provided by the client, in order to derive correlations between varying factors to be considered within a match and the physiological output of the player. Unfortunately, issues arose which meant that this could no longer be the case, and as of mid-August, the project description had to be changed and thus, a new plan was invoked. The following plan describes an updated timeline for the revised project, beginning on the amended start date.

Client Report

1 Executive Summary

This report contains detailed exploration of multiple types of analyses which can be used within the world of tactical analyses. Initially, average amounts of clearances, free kicks, fouls, throw-ins and touches were explored to provide a sufficiently comprehensive starting point for the client to gain knowledge into the typical statistics that are important in tactical analyses.

Passmaps were then created in order to determine the best way to analyse passes made by a specific team. Multiple types of passmaps were created and analysed alongside utilising the expert judgement obtained from the interview earlier in the project. The passmaps were adjudged to analyse individual players or a team in 10 minute incremental phases to derive the best forms of information from them.

Shotmaps were the next to be constructed, with again, the expert judgement being used to determine the best possible method to generate these shotmaps. It was decided to forge these shotmaps in 10 minute increments for the entire squad, in order to analyse positions of the pitch to exploit.

An exemplar match-day analyse was then conceived in order to provide ample insight for the client in terms of the typical report that would aim to be created after a particular match. Analyses of both passmaps and shotmaps were created for a singular Arsenal game. The passmaps covered the entire Arsenal playing squad in 10 minute incremental phases and the shotmaps were analysed for only the forward players, with all scoring at least one goal - in order to calculate goals to shots ratio.

2 Acknowledgements

I would like to pass on my earnest gratitude to my academic supervisor - Matthew Hutcheson - who has provided much appreciated aid and direction, which has allowed me to complete this project to the best of my ability.

A special thank you to my Mum and Gran for all the support, advice and encouragement you have gave me over the entirety of my academic career as it is wholeheartedly appreciated - as have all the cups of tea and coffee, which have gone down an absolute treat.

Finally, thank you to my friends - Erin and David specifically - who alongside my family have had to deal with the absolute ball of stress that I became throughout my entire academic career. A special mention also to the boys - Connor, Will, Graeme and Cameron - thank you for all those gaming nights which allowed me to de-stress and relax as I don't think I could've completed this project without them.

3 Introduction

Data Analysis in the footballing world is a hastily growing asset in the current climate (Curry 2020). Vast amounts of money are being poured into the industry in order to give football clubs an advantage in every game that they play - whether that be against a high or low ranked team (Müller et al. 2017). It's use is becoming more and more prominent in the current footballing world - with teams ranging from European Champions levels all the way down to teams in the fourth tier of the English Football League system.

Within the footballing world of course, data analysis is not always the deciding factor in winning games - sometimes pure passion and hunger is needed to win games. Data analysis in football is a contributing factor, however, as it allows the coach to make tactical and physiological changes depending on the needs of the team and the ability of the opposition. Data analysis in football has the aim to increase the winning probability of the team utilising the analyses - not to be the deciding factor of games (Tenga & Larsen 2003).

4 Literature Review

Data analysis became a severely useful and coveted aspect within the sporting world. The use of data analytics ranges from simple ideology to more complicated forms of analysis. The simple analysis consisting of areas such as: calculating expected goals of a specific team or match, to producing a points per game evaluation in order to decide the outcomes of domestic leagues; the latter was most recently executed to decide the outcome of the Scottish Professional Football League due to the implications of the Covid-19 pandemic. The more complicated forms of analyses consisting of topics such as analysing entire domestic leagues in order to identify potential player transfers that can aid in improving a certain team - whether that be a player to go straight into the first team or a player for the future, or even such areas as injury analysis and prediction (Moura et al. 2012).

Considering in today's era, that there is so much access to readily available data, there have been many various topics researched in order to further understand and improve the area of data analysis in the sporting world (Ortega et al. 2016). However, due to football, particularly, being a seriously competitive sport, a lot of sensitive day that individual clubs will collect in order to assist in decision making and such like, this data is not readily accessible. This has seen a rise in somewhat freelance companies, rising up and collecting their own data in similar ways to what football clubs will do. Companies such as SciSports and Opta have seen a significant rise to fame in recent times, with the two companies being used by both football clubs and football players alike. This section will aim to delve into the use of data analysis within the footballing world.

4.1 Data Analysis in Football

There are numerous routes and directions that can be followed when looking into the use of data analytics in football. One such direction is the area described within "Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science" Rein & Memmert (2016). This papers showcases the use of data analytics in the context of setting up the formation and tactical choices of a specific football team. It investigates the use of analyses of team tactics requiring significant loads of data, as opposed to formalising a team setup through only visual experimentation and observations. It is presented that data is becoming more readily available through technological enhancements but "cleansing" the data becomes an issue at the status of gathering such large data. The main aim of this paper was to showcase the use of such techniques as machine learning can be utilised in order to present the relevant data in such a way to ensure correct production for tactical analysis.

Another use of data analytics in football, is utilising a system which allows for the analysis of movement and events which happen within the game. The system created within the paper (Sacha et al. 2014) offers the ability to analyse these features within the game for a multitude of different options - ranging from the analyses of a single player to a more event structure analyses. For example, the system shows the capability to analyse the formations of a 4 defender system through the utilisation of the defensive effectiveness of the players used in said back 4.

Data analysis in football has also evolved into allowing the use of data to perform analyses on the physical aspects of the game, both for male and female versions of the game. The study used for the sixth FIFA Women's World Cup showcases exactly that (de Football Association et al. 2011). This paper investigates the physical performances displayed within the tournament and aims to display the distances that have been covered by individual players - both organised

into outcomes of games and in order of position. This aim was kept in mind to provide valuable insight as to how to improve or assist within the training regime of the players.

4.2 Spatio-Temporal Analysis in Football

A more specific example of data analytics within the footballing world, is the analyses of spatio-temporal events in football games - this meaning events that belong to both an area of space and time. One such example of this is the study which was carried out to be a descriptor of spatio-temporal data through the utilisation of a publicly accessed set of data (Pappalardo, Cintia, Rossi, Massucco, Ferragina, Pedreschi & Giannotti 2019). This study showcases a multitude of spatio-temporal analyses - including the likes of passes, fouls and interceptions to but name a few - and details around how this data can appropriate this style of analayses in terms of improvement of both the general game of football and the individual teams.

One major issue in the field of spatio-temporal analysis, is the ability to mine or cleanse the large load of data that has been collected to quantify the area of interest. One such study aimed to experiment with large scale data and perform multiple analyses - including both player and team - in order to create a "role-based" system of a progressively changing match, as they describe (Bialkowski et al. 2014). For example, they used their system to detect the formation of a team to then produce a visualisation of said formation, to further emphasize the benefit of this system in the case of large scale match analysis.

In the footballing world, a significant aspect of data analysis is the identification of events in match-time. As a football match at the absolute minimum consists of 90 minutes of gameplay, tracking and gathering the data involves a huge amount of trawling. This study involves focusing on visual analysis of football matches to observe and comment on a concrete approach to aid in this

identification (Khan et al. 2018). This system is then able to classify events and categorise them into spatio-temporal areas.

4.3 Player Passing Network Analysis

Player passing networks are a huge asset in terms of tactical analyses for football. They provide a spatio-temporal reference to how each player is performing and contributing to their assigned role, position and formation. They are also used to exhibit the players contribution to the team in terms of passing, and who, perhaps, is under-performing in terms of the match (Buldú et al. 2018). One study was able to establish the use of network theory in the area of coordinating a player passing network (Gonçalves et al. 2017). This provided a visual display of the individual teams formation and so called "network" that the team followed in terms of playing.

One significant aspect of player passing networks, is the comprehension of how teams play their football - which is especially important to the opposition coach, as this allows them to develop their game plan to attempt to win the match. A study was carried out in order to delve in the understanding of that exact asset (López Peña & Touchette 2012). The study aimed to determine the correlation between the network that a team used for passing during a match, and the result of said match - specifically in youth football. It was found that a strongly connected system, alongside varying other factors produced a higher probabilistic chance of a win.

Player passing networks can be determined to be an extremely significant asset in terms of data analysis in football. All areas can benefit from the use of this, as it provides a significant advantage in terms of tactical analysis both for a specific team and opposition team. These sorts of analyses give the coach in charge vital information that can aid in the decision making process - whether that be picking the team for a match, or determining the best position that a

player can be played to utilise all of their strengths (Cintia et al. 2015).

5 Research Design

From the literature, it can be observed that one area in particular that is a huge asset when performing tactical analyses in football is player passing networks. Working alongside Rangers FC, it was found that the client generally can collect and gather a significant amount of data but seems to struggle in terms of performing such analyses or mining the data to utilise the data in varying ways. It was established that Rangers FC tend to only utilise data in terms of physical analyses - specifically heart rate data, distance covered and sprint speeds - the reason is apparent as to allow for in depth knowledge for such areas as injury prevention.

The forms of analyses which have been identified previously can be seen to significantly improve the forms of tactical and physical analyses that are being performed in the current climate at Rangers FC. These forms of analyses which have been seen to perform well in the literature, could be feasibly integrated into the data analytics infrastructure involved at Rangers FC, without the need to exploit any further resources or financing to perform such analyses.

This research will aim to specifically delve deeper into the areas of player passing networks, to provide a route into further use of data in terms of tactical analyses that Rangers FC could follow and implement. The study will establish as to whether these forms of analyses could be feasible or perhaps, if more work or resources are needed, it will look to recommend this to the club, in order to keep the club moving in the correct direction in improving their data analytics framework.

6 Methodology

Within this section, the methodological choices of this project will be discussed. The aim is to provide a clear and insightful picture into the choices that have been made and an explanation as to why such choices have been made.

The methodology of this project included such things researching into the data-set and recreation of player passing networks, as well as including an interview with an Exercise Physiologist from SportScotland.

6.1 Platform

For this project, RStudio was chosen to carry out the coding aspects of both the passmaps and shotmaps. The visualisation aspect of these analyses would be carried out throughout the use of many packages, with the most important being both "ggplot" and "ggsoccer", which allowed for the creation of pitches and visual shots and passes to be shown on said pitches (Wickham & Wickham 2007).

6.2 The Data-set

The data-set that will be used as part of this project, was provided as part of a large public data-set article, revolving around soccer match events (Pappalardo, Cintia, Rossi, Massucco, Ferragina, Pedreschi & Giannotti 2019). The data-set aimed to be the largest available data-set consisting of so called "Spatio-Temporal" events.

The data-set was gathered through the use of an event-tagging software. This means that to gather this type of data, a full match data-set will only be available after, at the very least, 90 minutes - not including for stoppage time or any additional time in varying competitions (Pappalardo, Cintia, Ferragina,

Massucco, Pedreschi & Giannotti 2019).

The article utilised numerous data-sets in order to perform the analyses described within the paper. This project utilises three of the provided data-sets in order to create passmap and shotmap visualisations. This section aims to delve into the spatio-temporal events data-sets that would be utilised in this project.

6.2.1 Matches

The first stage of this project revolved around filtering the "Matches" data-set to only show matches of interest. The decision that was made at the start of the updated project, was to work around creating analyses for Arsenal Football Club to keep consistency and allow for comprehension throughout.

This data-set consists of numerous columns which contain the following information: competition name, date, duration, gameweek, team labels, match-day number, season identification, status of match, and the venue where the match has been played. To begin the visualisations, this data-set was refined in order to showcase only the matches which resulted in an Arsenal victory - as this would allow for analyses for the coach or manager could utilise to determine what tactical decisions are working and resulting in a victory. The key observation from refinement of this data-set was the match identification, which could be utilised in the other data-sets in hand.

6.2.2 Players

The next stage of the project was centred around making use of the "Players" data-set to refine the data-set to only showcase players that belong to Arsenal Football Club. This would allow for a more understandable approach in the production of the passmap and shotmap visualisations.

The data-set again consists of several pieces of information, which are as follows: birth area, birth date, national team identification, domestic team identification, first name, middle name, last name, preferred foot, height, passport area, position, commentary name, weight and player identification - which has been assigned via Wyscout. The refinement of this data-set revolved around filtering the data-set to display the players that currently play for Arsenal. This allowed for the viewing of the current Arsenal Squad, as of the 2017/18 EPL season. The key point that could be observed from this was their player identification, which again could be utilised for refinement in other data-sets.

6.2.3 Events

The final stage of the project, refers to extracting the spatio-temporal events of intertest - in this case passes and shots. This would be used in turn with the previously filtered data-sets to only show passes and shots made by the Arsenal squad only in matches that resulted in a victory for Arsenal Football Club.

This data-set encompassed the following information: event identification, event name, sub-event identification, sub-event name, tags, event time in seconds, identification, match identification, match period, player identification, team identification, and positions of the event which has taken place. This final part, allowed for further refinement to extract the exact information needed to create the passmap and shotmap visualisations.

6.3 Expert Judgement

Within this part of the project, an interview was carried out in order to determine the various types of information that would be helpful and that typical analysis of data would look to find. The interview would be carried out over Zoom due to COVID prohibiting the ability to have a face-to-face interview.

The interviewee in question was Mr. Michael McKenna, who is an Exercise Physiologist employed by SportScotland. Mr. McKenna has both a BSc and a MSc in Physiology with the masters specialising in Exercise Physiology.

Discussions took place within this interview, revolving particularly around the initial plans of the project, but with information that could be construed in order to correlate with the updated project plan. These discussions included of mainly scripted questions to be answered but allowed for an open discussion to take place if the need arose.

One of the main key points of discussion that could be altered to correlate with the updated project, was to analyse smaller increments of the match in hand, most notably into phases of 10 minutes. This is what is most commonly used in terms of monitoring fitness levels in a game, so this knowledge has been utilised in terms of creating both the passmaps and shotmaps. In terms of the physiological analyses of a match, splitting the game up into increments of 10 minutes allows for the determination of phases of play which will impact the player's fitness level the most and provide reasoning as to why - usually this is down to tactical aspects of the game, e.g. if they are running more at a specific period. In turn, this can be manipulated into terms of passmaps and shotmaps, in order to analyse the phases of games where teams are more likely to score, or where players are more likely to be involved in the tactical build-up.

There was also discussion as to whether positions or injury record play a part in how the analyses should be carried out. Again, this was carried out in the context of physiology but can again be construed for use in correlation to the updated project. By this, some forms of analyses will be carried out regarding positions of the players to be analysed.

At the beginning of the project, when the initial plan was to centre the project around football physiology, discussions took place with the client in order to discuss how best to set out the project. As time proceeded and the client was

unable to provide the promised data - due to unforeseen circumstances that were the fault of neither parties involved.

At the time there were no specific routes to follow advised by the client, with freedom of reign passed to myself as to how to best pursue the project. There were slight suggestions to perhaps look into the correlations between match fitness and heart rates, further going onto injury prediction if the data promised had have sufficed.

When the circumstances arose that meant that the project had to be altered, discussions took place between myself and the client, and also myself and Matt, in order to best determine the correct route forward.

6.4 How Are We Using The Data-set

Within the project, there are some key points that we are looking to obtain. The main points which are to be prioritised, are to create visualisations of both passes made and shots made by Arsenal Football Club.

The data-set to be utilised within this project, has been published online as one of the largest public data-set of spatio-temporal events.

7 Data Testing

Within this section, the dataset provided would be analysed and utilised in order to correlate the correct data needed to create both passmaps and shotmaps. Further down the line, the aim will be to create and exemplar match-day analyses, in order for the client to gain some comprehensive knowledge into the interpretation,

Unfortunately, the volume and specificity of the data needed to ensure an accu-

rate passing network could be made was not made part of the dataset. Instead this section will aim to create passmaps and shotmaps that will provide the most useful forms of information, and will suggest as to how this can then be further analysed to create such networks.

7.1 Exploratory Analysis

Within this section, the aim is to provide an introductory look at some typical properties that would be of use to the client in terms of match-day analyses. These analyses will provide insight into the all games played by Arsenal, in order to examine the averages of certain feats - these being: clearances made, free kicks, fouls made, throw-ins and touches. By extracting this information and calculating the average amount of each of these, will allow for the determination of whether Arsenal are under or over-achieving in these aspects, and if this has any correlation to the outcome of the match - for example, we would expect a team that has made more touches in the final third, to perhaps have scored more goals as the majority of their possession has been around the opponent's goal area.

The English Premier League season consists of 38 games, with each team playing each other twice. This will be the value used to perform the analyses for the future explanatory sections.

7.1.1 Clearances

Clearances are, of course, an extremely significant aspect of the defensive aspect of a football game. Without a doubt, the best teams tend to play their football out from the back, not relying on clearing the ball up the field. This facet tends to be more important with weaker teams, perhaps those battling relegation. These teams will be the ones who tend clear the ball up the pitch, in order to

regain their formation and tactical instructions, before the next attack from the opposition.

In terms of clearances, (n.d.*a*) Arsenal F.C. average around 13.13 clearances per game, in the English Premier League. Arsenal can be considered to be one of the so-called "Top 6" - this means that they are considered one of the most likely teams to consistently finish in the top 6 places of the English Premier League. From this, it could be assumed that in a typical game, Arsenal should be the driving force in the game, controlling the majority of possession and overall being comfortable on the ball. They should not allow many shots - with the other members of the Top 6 perhaps exempt from this inference. As opposed to those teams which are on the lower end of the spectrum, which will, presumptuously, give way to numerous more shots and may not feasibly be as comfortable on the ball, resorting to clearing the ball to allow themselves time to regain composure.

7.1.2 Free Kicks

Free kicks in the footballing world are an immensely useful both in terms of analyses and in various aspects of the real-time match. Within the match, the importance of free kicks cannot be underestimated - as is often the case. Free kicks provide copious amounts of opportunities to control the play, to create a goal-bound chance and so on. Indubitably, these opportunities will range depending on the positions of the free kick, with most goal-bound chances created from positions within the final third.

Through analyses, Arsenal F.C. were found to average 9.89 free kicks per game, with this consisting of only free kicks given in the opposition half. From this, it was established that Arsenal F.C. had only scored twice from a free kick throughout the entire season. Inflating the average value to cover an entire English Premier League season, Arsenal F.C. would obtain 376 free kicks per

season. Using the knowledge of only 2 goals having been scored from free kicks within the 2016/17 EPL season, this would emit to a 0.5% chance of scoring from a free kick. This would suggest to the coaches in charge - indeed, not without further analyses to identify the trend over multiple seasons of the clinical nature of the free kicks - that improvement would be needed on free kicks. This improvement could come via altering the free kick taker, the routine of the free kick, or even the positional discipline of the players taking part in said free kick. Another important detail needed to ensure the validity of this statistic is the nature of the free kick - whether it be indirect or direct - however, the percentage calculated could still be utilised in analyses for free kick improvement.

7.1.3 Fouls

Fouls are the second most common type of game interruptions in a typical football match, with an average of 33 fouls per game. This statistic, of course, refers to total fouls in the game, both for and against a specific team and for both teams, in total - so it would be wise to assume that on average there would be 16.5 fouls per team. Although fouls are severely frowned upon in the footballing world as they can cause injury to the receiving player - they are not without their advantages. This form of foul is commonly known as a "professional foul." A professional foul defines any foul which is used to bring an advantage to the committer of the foul's team. The most common type of example of a professional foul is when a player is brought down to halt a counter attack. This example often goes hand-in-hand with what is known as a "tactical yellow card" and provides abundant opportunity for the team which has committed the foul to recoup and recover their formation and retreat back into position to cover their men or zones as organised within the coaches game-plan.

Through exploratory analyses of the data-set, it was established that Arsenal averaged only 9.66 fouls per game over the course of the 38 games in the English

Premier League season. This is significantly under the average amount of fouls seen previously to be 16.5. This difference could be explained through a variety of speculations, but the exact reason could be interpreted from video analyses of the exact match. One such explanation as to why Arsenal perhaps under-perform in regards to committing fouls, is the most typical game-plan is to control possession and pressure the opponents in the opposition half, which would not rely on committing as many fouls as perhaps the opposition would in terms of both normal fouls and professional fouls alike.

7.1.4 Throw-Ins

Over the standard EPL season, through analyses, Arsenal were found to average around 19.66 throw-ins per game, which would be rounded down to 19 for proper comparisons. Through an independent study mentioned previously, it was found that there were on average 40 interruptions per game corresponding to throw-ins, with this assumed to be split evenly between both teams - meaning 20 throw-ins for each respective team per game. From this, it can be established that Arsenal tend to under-perform in terms of the amount of throw-ins per match that they obtain, albeit only by 1 throw-in.

Throw-ins are often overlooked in football, with most people assuming that they are only a break in play, and provide no tactical advantage. This, however, is quite simply not the case. Throw-ins are useful both for attacking and defending phases of the game. In terms of defending phases, throw-ins allow for the defending team to achieve a bit of "breathing space" where they are able to take a few seconds to re-align and get back into formation fairly quickly. On the other hand, in the attacking phases of the game, they can provide ample opportunity to gain further yards, and perhaps even result in an attempt on goal - this is seen from numerous teams now utilising the tactic of long throw-ins, mainly into the 18 yard box, to attempt a header or some such, as an attempt on goal.

7.1.5 Touches

Analyses of touches can be extremely important both for the analyses of home and opposition teams. Analyses of touches and touch maps can produce insight into many different aspects of the game - such things as: whether the home or opposition uses zonal or man marking, whether a certain player is advancing too far or not far enough up the pitch towards the forward thirds and so on.

On average, (n.d.*b*) the Top 6 clubs, end the season on approximately 30,000 touches over the course of 38 games - which allows for the knowledge, that Top 6 teams average around 790 touches per game. From the analyses produced for the touches made by Arsenal, it was found that they averaged around 761 touches per game. Now, from this, it could be said that perhaps the likelihood is that Arsenal would not be able to finish within the Top 6 in the 2016/17 season, but the reality is that they finished 5th place - 5 points above 6th place Manchester United. From what has been discussed, this perhaps does not seem very realistic in terms of touches, however, for further analyses, the touches made by the other 5 teams which finished in the Top 6 that season would have to be analysed, as this could have possibly been a year in which all teams under-performed in terms of touches made.

There are, of course, limitations in terms of analysing this statistic. One such being, that it does not take into account the style of play that a team favours. For example, Leicester City won the league in the 2015/16 season, and were 5000 to 1 outsiders, but their form of counter-attacking football proved to work against some of the biggest teams in the country, allowing them to claim the title. This style of play was adopted by other teams in the next season, as teams knew how it worked, and could attempt to replicate what Leicester City had done the season before. Counter-attacking football relies on fast-paced transitions, preying on mistakes and high-defensive lines, and does not necessarily require a significant amount of touches to be made.

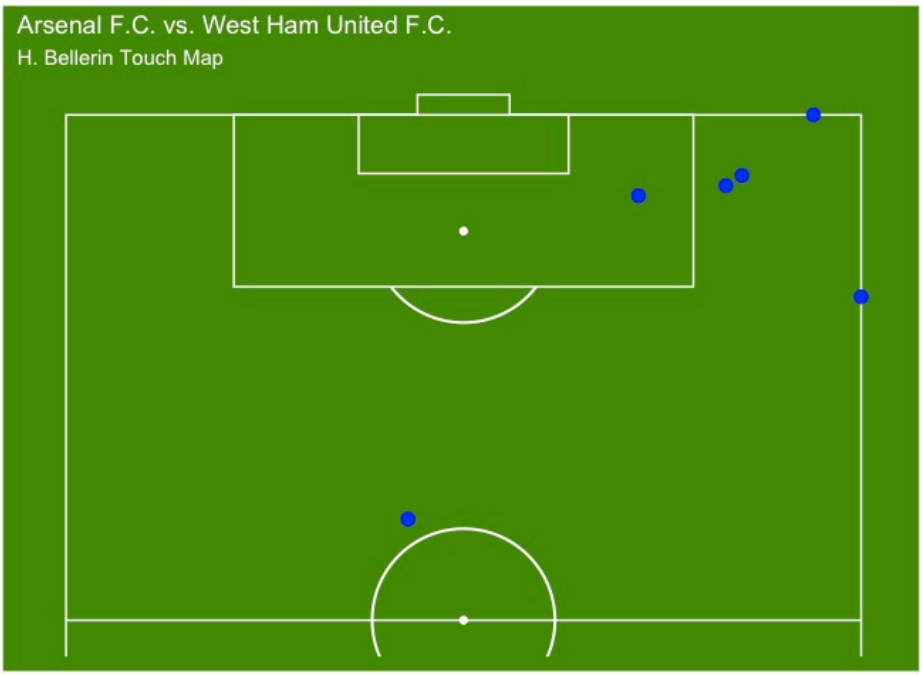


Figure 2: H. Bellerin (RB) Touchmap Versus West Ham United

Producing touch maps can be a powerful method of analyses, but require to be refined to fit the special needs of the client. Previously, it was seen that Top 6 teams average around 790 passes per game. This, for obvious reasons, would not be very helpful to plot on to a pitch-map as it would include the touches of all players in all locations they have made touches. Primarily, the client would make notice of what would be of interest to them, whether that be touches made in the box, touches made to showcase zonal or man marking and so on.

Within 2, the example has been to identify touches made by Hector Bellerin in the game between Arsenal F.C. and West Ham United F.C., which shows the touches made in the opposition half. From 2, it can be seen that the majority of Bellerin's touches appear to be on the forward right-wing, with a touch also appearing on the right-hand side of the 18 yard box. This follows along

with the most common Arsenal F.C. game-plan, which pushes the full-backs further up the wing to provide support to the wingers from supporting crosses or passing the ball back to the edge of the box to force gaps in the opposition defence. This touch map, observed in 2, can be used in terms of client analyses in order to identify lack of positional or tactical discipline, or to examine any implementations of update play-styles during a given phase of the game.

7.2 Passmaps

This section aims to create accurate passmaps that can be utilised to provide information to the reader or viewer. Passmaps are created in order to best highlight the influence a player has on the game in hand.

Passmaps can be used to show the connection between 2 players through passes if the average position is known. They can also be used to denote the average formation of the team being analysed (Shaw & Glickman 2019).

Overall, in the hastily growing sector of football data analysis, passmaps are a significantly insightful and useful visualisation for many such reasons. This, being one of the main reasons that passmaps will be created in this section.

7.2.1 Entire First XI Passmaps

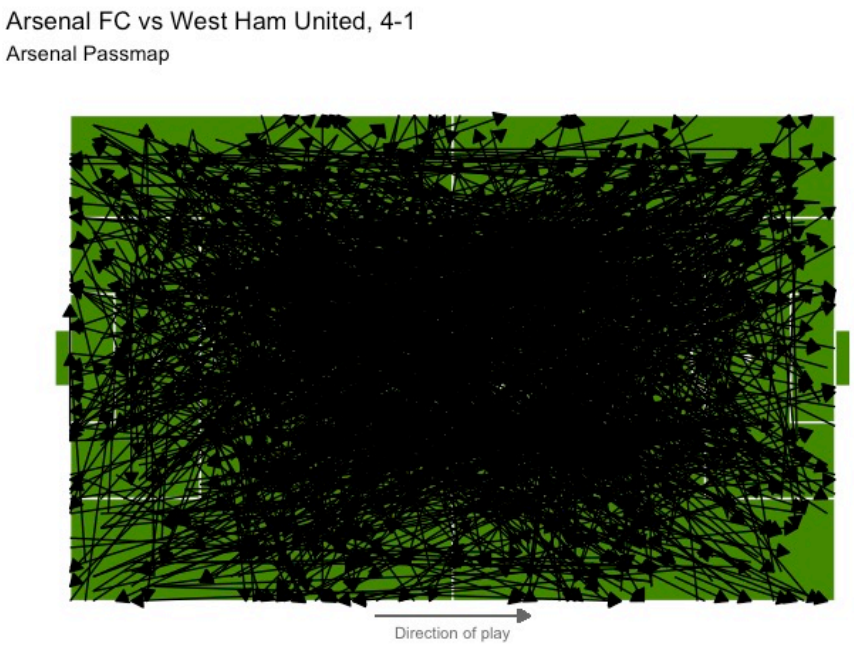


Figure 3: Arsenal Entire Playing Squad Passmap Versus West Ham United

in 3, all passes made by the Arsenal playing squad, including substitutes can be observed. This form of passmap is not as useful as it could be as the passes are all clustered together and we are unable to see the individual passes from their starting position to the ending position.

7.2.2 Individual Player Passmaps

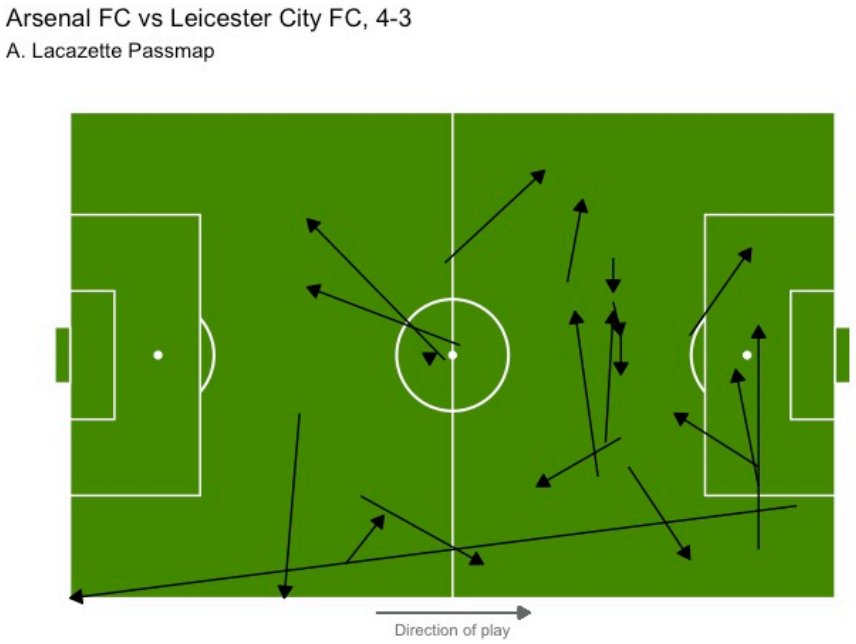


Figure 4: Alexandre Lacazette (ST) Passmap for Both Halves Versus Leicester City F.C.

Within 4 above, the passes made by only Lacazette can be seen. This map shows all the passes made within both halves of Arsenal's 4-3 victory over Leicester City. This visual is much more understandable as opposed to 3, previously.

Initial observations would suggest that Lacazette, perhaps, was not involved in as much of the play in the match as you would expect from a player at such a high level. This is largely due to Lacazette being a traditional "number 9" who relies more on other players in the team passing the ball to him, as opposed to Lacazette passing the ball to other players.

The main issue with this passmap is the fact that the direction of play can be

quite confusing. Although the visual has been labelled to denote the direction of play, the direction of play will switch at half-time. Without knowing which of these passes were made in both the first and second halves, it could be suggested that perhaps Lacazette made a pass backwards, whereas he could have infact played it forward. This leads to a significant amount of confusion in tactical analysis, this will be taken into account in the next section.

7.2.3 Individual Player Passmaps in Individual Halves

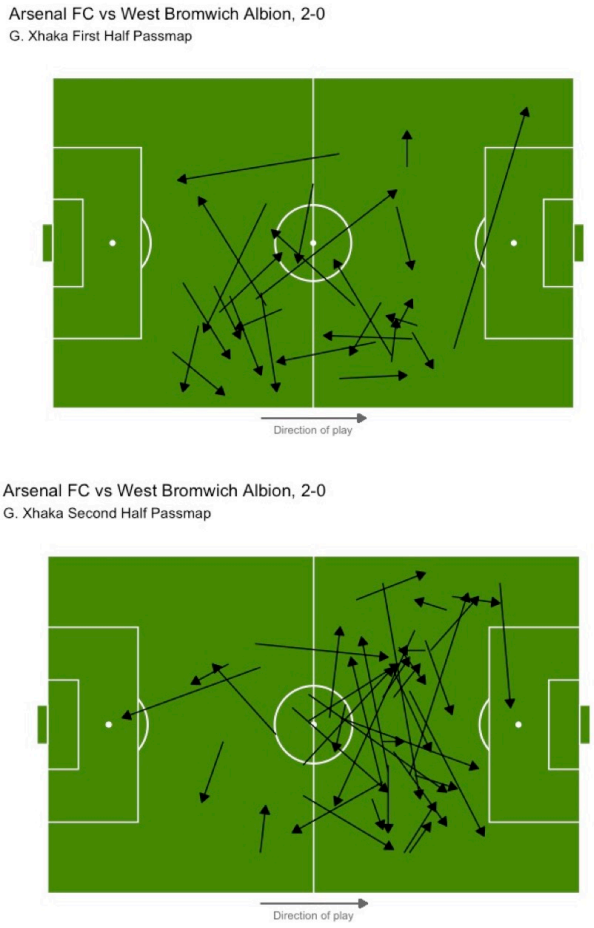


Figure 5: Granit Xhaka (CM) Passmaps for Both Individual Halves Versus West Bromwich Albion

As mentioned in the previous section, the direction of play in both individual halves must be taken into account. In 5, the direction of play is denoted by the label and follows from left to right, and this is so for both halves - with the data reflecting this directional situation also.

These passmaps showcase passes made by Xhaka in both individual halves. Now, this form of visualisation allows for valid and significant insight into the tactical analysis and involvement of said player. They provide a comprehensive volume of information that can be utilised in varying forms, more likely for the tactical analysis, and allow the club manager to best determine what style of player he is, what he brings to the team and how best to utilise him in the manager's chosen tactical system or formation.

Again, however, these passmaps are very useful but can always be improved. The next section will aim to further enhance these visualisations in order to make them into their most comprehensible form, ultimately for the club manager or first-team coach to be able to make tactical decisions.

7.2.4 Individual Player Passmaps in 10 Minute Increments

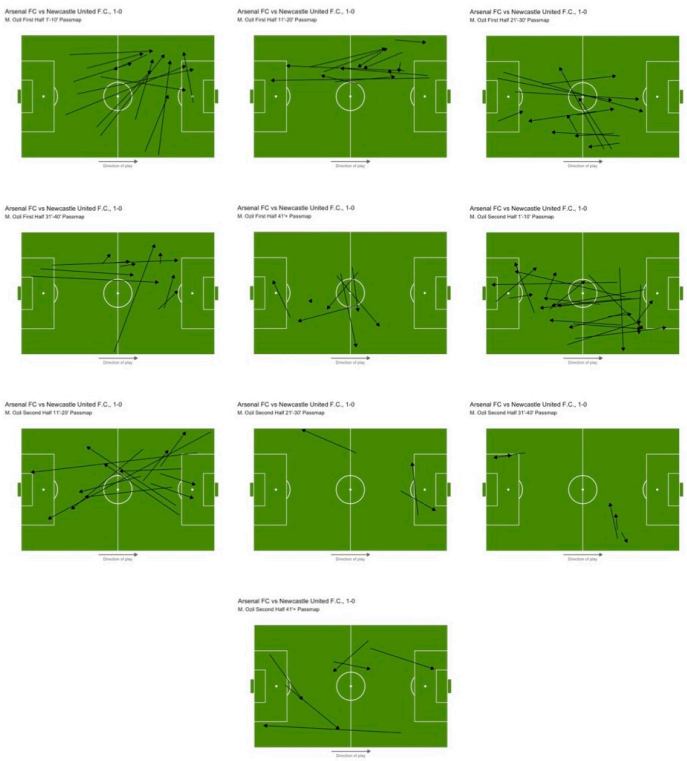


Figure 6: Mesut Özil (CAM) Passmaps for Both Individual Halves in 10 Minute Increments Versus Newcastle United F.C.

Within the Expert Judgement section, it was learned that the most useful forms of analysis are split into 10 minute increments - this allows for better viewing and more in-depth analysis of the match in hand. The aim of this section was to create pass maps of an individual player in 10 minute increments to best fit

this judgement.

In 6, passes made by Mesut Özil versus Newcastle United can be observed. Özil tends to fit the mould of a typical "Number 10" who's style of play revolves around play-making and assisting the striker or "Number 9." By splitting the match into 10 minute increments we can see how the different phases of play are carried out.

For example, in the visual for the first 10 minutes in the first half, we can see that the majority of Özil's passes were directed towards the left wing - notably where Aubameyang plays, who is notoriously one of Arsenal's top scorers. Seeing this phase of play, we can determine that the Newcastle Right-Back - i.e. the player who would be defending against Aubameyang - perhaps, would've been identified as a player who would struggle defending against Aubameyang, and hence the play would've been dictated down the left wing, in order to take advantage of this.

7.3 Shotmaps

Within this section, shots made by Arsenal in varying situations will be created and analysed. Shotmaps in football can be utilised for several reasons, for such things as average positional data maps, and more recently expected goals systems.

This section aims to create and analyse such shotmaps. All shotmaps will be based on the match of Arsenal F.C. versus Everton F.C. which resulted in a 5-1 victory to "The Gunners" or Arsenal. These shotmaps can be utilised in order to provide a visualisation of the average positions of shots made, and even further can produce a map of average key areas where shots are more likely to result in goals.

7.3.1 Entire First XI Shotmaps

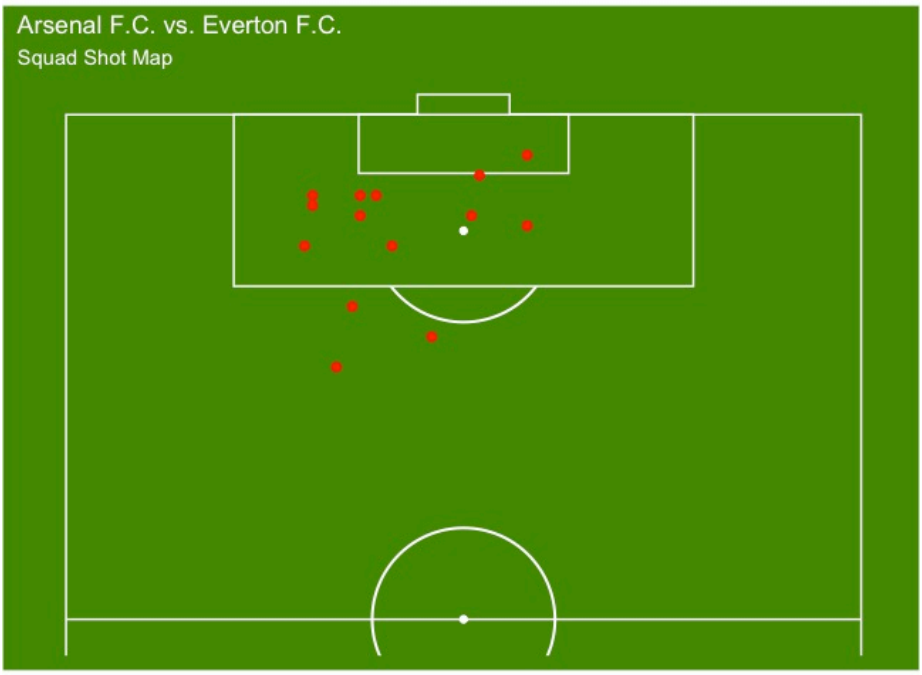


Figure 7: Entire Playing Squad Shot Map Versus Everton F.C.

In 7, the total shots made by the Arsenal playing squad - including substitutes - can be observed. This visual, albeit helpful, in providing an insight as to the local positions of the shots which have been made as well as providing tactical indications for the coach or manager.

The tactical analyses of this allows for the coach or manager, to gain insight as to where the shots take place and where the areas of weakness are that can be improved upon. Within 7, it can be seen that the majority of the shots have been taken from the inside left flank of the pitch, with all but three of them coming from within the 18-yard box. This allows for the coach or manager to become aware of a key question that they must ask themselves: is the right-sided

player struggling to impact the game or is the right-sided defender struggling to deal with our left-sided player? By asking themselves this, it allows for decisions to be made upon substitutions or tactical changes of play.

7.3.2 Individual Player Shotmaps

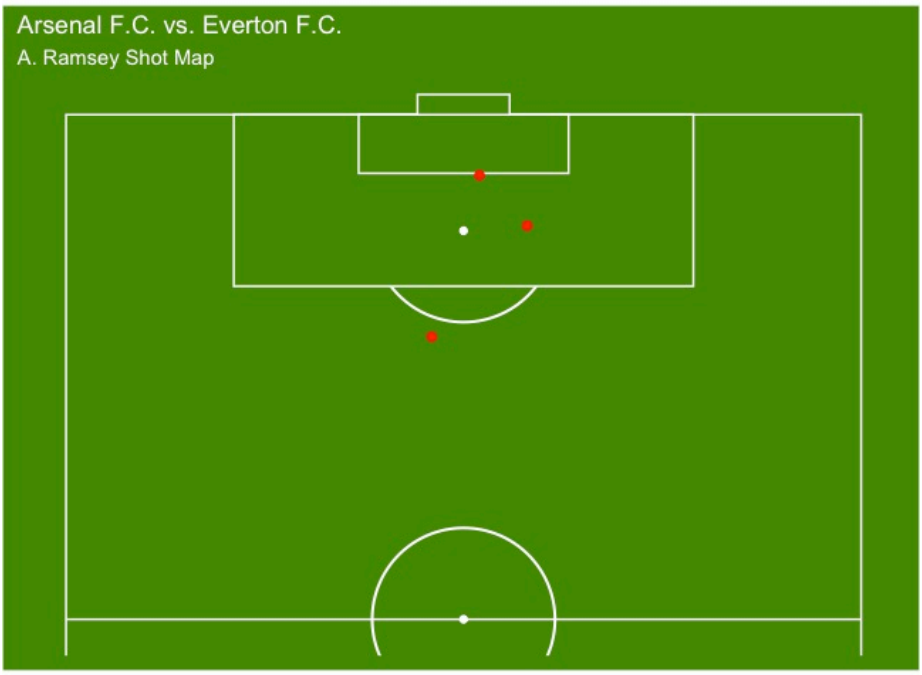


Figure 8: Aaron Ramsey (CM) Shot Map for Both Halves Versus Everton F.C.

Continuing on from the knowledge that was gained in the previous section, where it was observed that the visualisations would be more useful if they could be refined to display shots made by an individual player. Within 8, the Arsenal squad shotmap as seen in 3, has been refined in order to showcase the shots made by Aaron Ramsey in the same game.

Circumstantially, in this game, Aaron Ramsey scored a hat-trick, which given that he only made 3 shots throughout the game, describes a 100% shot accuracy. In terms of tactical analysis of the match, this visualisation can be used to determine how wasteful and individual player is in terms of shots made. For example, if a player has made 10 shots, while only scoring 1, it could be sug-

gested that the player is perhaps getting into the right positions to score, but is struggling with the finishing aspect. This can then be utilised to suggest training measures - whether this is to increase training sessions for that player, or to try alternative activities to improve their quality of shots.

7.3.3 Entire First XI Shotmaps in 10 minute Increments

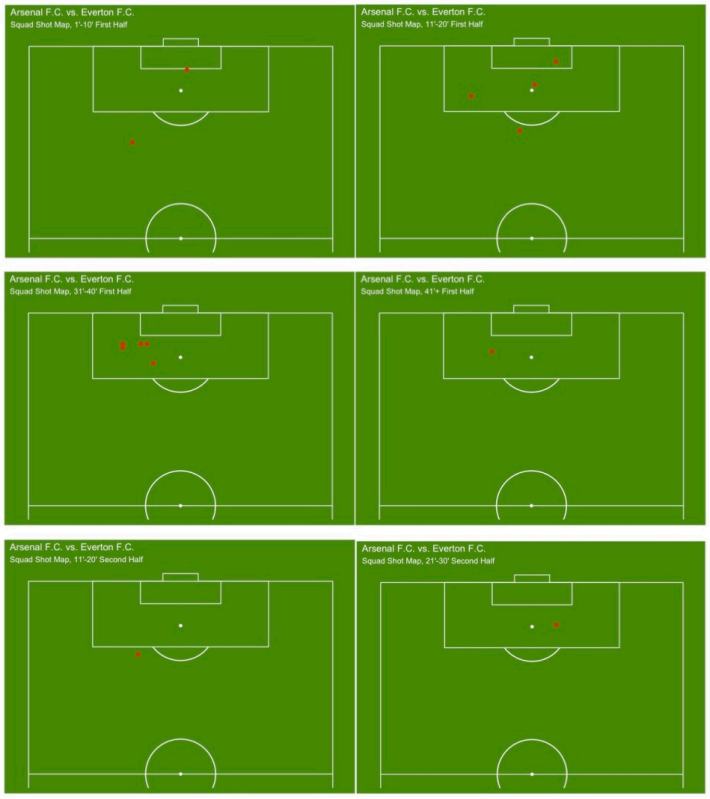


Figure 9: Entire Playing Squad Shotmap for Both Halves in 10 Minute Increments Versus Everton F.C.

As discussed in both the Expert Judgement section and the previous Passmaps section, splitting the match into 10 minute increments provide significant insight into the what has been involved in particular phases of play. The aforementioned Arsenal versus Everton match has been split into 10 minute increments, with

only those phases of play which includes at least 1 shot made being displayed - as can be seen in 9.

As can be seen in 9, it showcases shots made by the entire Arsenal squad as opposed to individual players, as 6 displays. In correlation to shotmaps, however, a team averages only 10 shots per game, with an average of 5 of the shots being made, on target (Barreira et al. 2016). This knowledge has been utilised in order to correctly choose the form of shotmaps created.

Within 9, it can be observed that there are only 6 phases of the game being analysed as opposed to 10 phases which can be seen in 6. This is to only display the phases of the game which will be of interest to the coach or manager. The phases of the game which are not shown here, will be subject to other forms of analyses by the coach or manager - perhaps analysing if these phases consisted of more defensive action, and how to improve or alter these phases.

8 Case Study - Exemplar Match Report

The aim of this section will be to provide the client with an exemplar match report - including both analyses and interpretation of said analyses. This will aim to aid the client in terms of creating analyses of match-day data and provide structure and insight as to what would be expected or most helpful in terms of tactical analyses for the coach or first-team managers and so on.

The match to be analysed as an exemplar, will be Arsenal F.C. versus Huddersfield Town A.F.C., which resulted in a 5-0 victory for Arsenal, on the 29th November 2017.

8.1 Arsenal F.C. Versus Huddersfield Town A.F.C. 29th November 2017

Within the upcoming sections, a comprehensive, exemplar match analyses will be produced. These analyses will consist of entire squad passmaps for both the home and opposition teams, with these being split into 10 minute incremental phases - as was seen previously. The analyses will also aim to create a valid and insightful group of shotmaps, focussing on the forward thinking players and the goalscorers within the match.

The aim is to also provide a personal interpretation of the match, using said analyses, to present an understandable and typical match-day analyses that intends to aid the client in producing their own analyses and interpretations.

Figure 10: Arsenal F.C. Versus Huddersfield Town A.F.C. Starting Team Line-ups, 29th November 2017

Within 10, the starting lineups of both Arsenal F.C. and Huddersfield Town A.F.C. can be observed. As can be seen, both teams were initially lined up to play in a 3-4-2-1 formation - this meaning, 3 central defenders, 2 central midfielders with 2 wing-backs, and with 2 wingers and a striker as the forward thinking players. This formation has become more popular in recent times, as it allows for a quick transition for teams between the attacking and defending phases of the game - this being reliant of the roles of the wing-backs (Tierney et al. 2016).

In the attacking phases of the game, both the right and left wing-backs tend to push forward leaving only the 3 central defenders to hold the back line. These wing-backs can push forward in various ways, with the most common being to overlap the left and right wingers, to provide support, and to allow for more players to come inside - either by making runs into the box for the incoming cross, or to make space on the outside of the box for a supporting pass.

Another style of wing-back that is becoming significantly more common is the "inverted" wing-back. This role has been adopted by numerous teams in modern football and has somewhat been trademarked as the invention of Pep Guardiola - the Manchester City manager. In the defensive phases of the game, the wing-back comes back on defence and plays in a typical full-back position to provide additional men in the back line, which alters the formation from a 3 at the back to a 5 at the back - this allows for a more solidified structure of defenders, restricting the amount of space the opposition has to play through balls or dribble through gaps in between said defenders. One of the most important roles of the inverted wing-back, comes in the attacking phase (Vilar et al. 2013).

In the attacking phase of the game, the wing-backs instead of overlapping the wingers, come central and play as a central midfielder. By doing this, it allows the actual central midfielders to venture further forward to provide more attacking support. Arsenal are known to do this, however, this is opposition dependent.

Within the case of Arsenal F.C. versus Huddersfield Town A.F.C., these formations can be presumed to have been tactically chosen, in order to unravel the respective opposition. Arsenal are a team who average a finish in the "Top 6" - from 1st to 6th place - as opposed to Huddersfield who most are most commonly battling to avoid relegation. With this knowledge, it could be presumed that Arsenal will most likely be the driving force of the game, with Huddersfield likely to be on the back foot for the majority of the game. The 3-4-2-1 formation is an obvious choice for Huddersfield as this will allow for a quick transition from attack to defence and vice-versa, with the majority of their play-style being based on counter-attacking football, where they will allow Arsenal to have the majority of possession, and take advantage of set-pieces or any mistakes made by Arsenal (Gonzalez-Rodenas et al. 2015).

The reason that Arsenal, however, have chosen this formation, is perhaps not to be considered as a tactical exploitation, but more of a strategy to accommodate for their defensive frailties. Considering that Arsenal would be the favourites to emerge victorious, they must also consider the fact that Huddersfield may outplay them and consider themselves favourites. Normally, looking at the gulf between the two teams in terms of ability and odds of victory, Arsenal would likely have played a 4 at the back system with only 2 central defenders, as opposed to a 3 at the back, with 3 central defenders. Now, of course, this formation allows for the quick transition between attacking and defending phases, but it also allows for an extra defender, which in the case of teams with defensive frailties, is a huge asset to have.

8.1.1 Arsenal F.C. Squad Passmaps

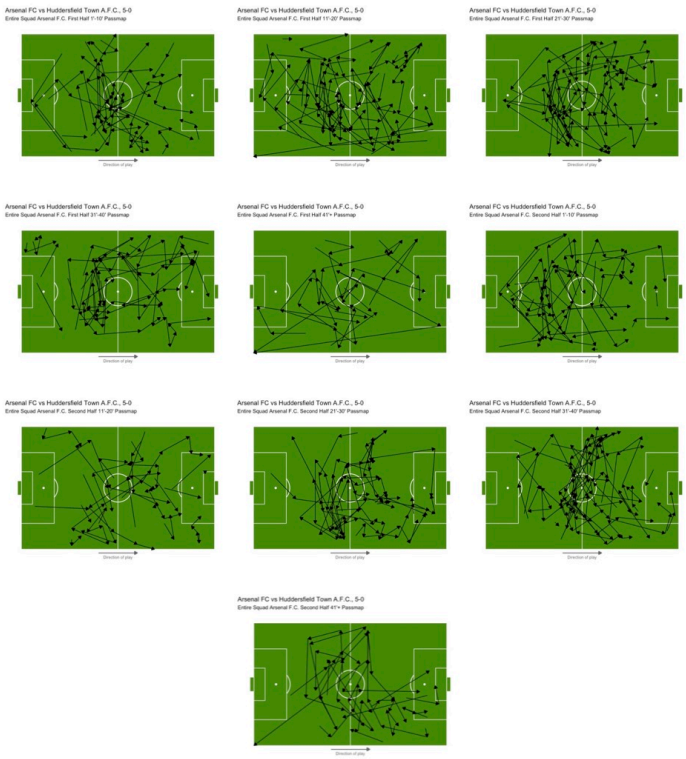


Figure 11: Entire Arsenal F.C. Playing Squad Passmap Versus Huddersfield Town A.F.C.

In 11, the passmaps of the entire Arsenal F.C. squad - of which 813 passes were made - can be observed in 10 minute incremental phases. The first and most prominent observation, is that the passes mostly follow the typical outline of the 3-4-2-1 formation as was seen in **??**. By examining this, it can be derived that the Arsenal players are more strict and conservative in terms of deviating from their tactical instructions, as it can be seen in every phase of the game, that

the 3-4-2-1 outline is almost always followed - with, of course, some deviation, incorporating for inverted wing-backs as mentioned previously.

On further inspection, the fact that the majority of these passes takes place in both the middle and forward thirds of that pitch can be seen. Of course, within some phases, namely 11'-20' in the first half and 1'-10' in the second half, it can be observed to show that there is a higher cluster of passes within the defensive third. This can be explained by a variety of reasons.

In the phase of 11'-20' in the first half, the reason that the majority of phases took place in Arsenal's defensive third, is perhaps down to Huddersfield growing more comfortable in the game. It is a well known prospect, that some teams can start the game "all guns blazing", while others take a period of time to settle in and compose themselves - this more so being the weaker of the two teams, as the stronger side will likely dominate possession from the first minute of the game.

Within the 1'-10' phase of the second half, the reason could perhaps differ. At half-time, Arsenal lead the game 1 goal to nil, which is a difficult deficit to overcome for teams battling relegation, however, though it may be implausible; it is not impossible. Half-time would be an ideal time to provide motivation, alternative game plans and so on, in an attempt to revive the outcome of the game. The likelihood is that something of this sort happened in the Huddersfield dressing room at half-time, forcing Arsenal to be more defensive in the 1'-10' phase of the second half, until they had in fact settled back into the routine of their game plan - this can be seen by the fact that the majority of their passes again take place in the middle and forward thirds of the pitch in the 11'-20' phase of the second half.

Looking at the game as a whole through these analysed passmaps, it would be wise to say that Arsenal had a strong game in terms of tactical instruction and in terms of passing. This would be wise to say as these passmaps have

been examined, and in all but two of them, it has been seen that Arsenal have the majority of passes in both the middle and final thirds of the pitch. The two weaker phases of the game, can still draw positives however, as it can be observed that although, Arsenal make a significant amount of passes in their own final third, they are in fact still in control of both the passing and possession elements of the game.

8.1.2 Huddersfield Town A.F.C. Squad Passmaps

Figure 12: Entire Huddersfield Town A.F.C. Playing Squad Passmap Versus Arsenal F.C.

Within 12, the passmaps made by Huddersfield Town A.F.C. - of which there are 337 - in 10 minute incremental phases can be observed and examined. As opposed to what was seen in 11, and what was discussed in the previous section, Huddersfield Town did not have the strict tactical nature that Arsenal followed. It is known from **??**, that both teams began the game in the 3-4-2-1 formation, however, this is not particularly able to be seen from the Huddersfield passmaps. As mentioned, Huddersfield made only 337 passes throughout the full game, just

short of a third of the passes that Arsenal made, so the lack of tactical dexterity comes without questions.

It can be established from these passmaps that Huddersfield Town executed a game-plan reliant on nullifying the attacking threat of Arsenal, with the majority of passes taking place in primarily the defensive third, with of course some outliers taking place in both the middle and forward thirds. The best example to show the extent of their defensive tactics can be observed in the passmap which denotes the 21'-30' phase of the second half. Within this phase of the game, Arsenal scored 3 goals - in the 68' (24'), 69' (25') and 72' (28') minutes respectively. As observed within the passmap, Huddersfield Town display a team shape with at least 9 players - excluding the goalkeeper - behind the ball. The only player that appears to stay forward, is the number 9, which can be seen from the passes backwards from the centre circle. On the other hand, this could be argued to be the passes which are taken as part of kick off procedures after the Arsenal goals were scored, but without accurate information and access to the video tagging, this cannot be assumed.

Again, looking at the 21'-30' phase, the defensive structure of Huddersfield Town can be further scrutinised. Tactical discipline or "keeping the shape" is a hugely significant aspect in the defensive aspects of football, and within this passmap, it can be observed that Huddersfield do not particularly follow the defensive shape associated with their original 3-4-2-1 formation. In this phase, it looks likely that Huddersfield were extremely stressed and pressured with Arsenal's attacking plan, forcing them to retreat and reform their formation to more of a formation that would be used in ultra-defensive game-plans. This sort of style can work, however, it would most likely be used in the final few minutes of a game, where a team is trying to hold on to a narrow lead, with the opposition throwing everything at them to get another goal. The assumption, however, is that this style of play does not work for a continual amount of time, as can be established in this phase, where Huddersfield Town conceded three goals in a

short period of time.

The most attacking phase of the game for Huddersfield Town can be examined in the 1'-10' phase of the second half. Within this phase, it can be ascertained that Huddersfield Town were more attacking and pressured Arsenal to retreat into the middle and their defensive thirds. Huddersfield Town went in to the dressing room at half-time losing 1-0, which in totality, is not the most difficult result to overcome in the second half. Indubitably, this would have been a significant talking point in that Huddersfield dressing room at half-time. This type of scoreline at half-time can be used to motivate players as they are still in the game, and a turnaround in result is not to be laughed at. Also, from the expert judgement which was received, a discussion point arose, which denoted that players reach their highest heart-rate levels in both the 1'-10' and 41' plus phases in both halves. This can be used as an explanation as to how Huddersfield were able to pressure Arsenal back into their own half and into a more defensive phase, as they were given both motivation from the scoreline and from the players pushing themselves to their limit regarding heart-rate levels.

Some of the attacking input also comes in the 31'-40' phase of the first half. It can be seen that Huddersfield also have attacking passes in other phases of the game but this particular phase can be said to be the most promising attacking phase as they keep a stronger tactical dexterity. Although not alike their starting team formation of 3-4-2-1, in this phase, it looks like the coach has changed their team shape to a 4 at the back system, most likely a 4-2-3-1, as can be seen in the 31'-40' passmap. The coach has perhaps shouted in-game instructions to their team directing them to alter their formation as the original 3-4-2-1 has clearly not been as successful as they thought. This, of course, was to no avail, as Huddersfield ended the game without a goal, however, within this phase they looked more comfortable as a unit in the 4-2-3-1 shape.

Examining the game in it's totality via these passmaps, would suggest that Huddersfield Town were immensely distressed with their tactical disciplinary

instructions and were uncomfortable playing as a unit - this is exhibited from their low amount of passes throughout the entirety of the game. Analysing the game from a coaches point of view would be a fairly easy job, as more often than not, scrutinizing the negatives and the mistakes from a particular match, is more useful than only examining the positives from that match. It provides copious amounts of benefits as these analyses will be able to show the coaches and players a like, what not to do. These passmaps shown in 12 showcase very few positives, but they are positives nonetheless, and these are something to build a foundation upon. Looking at the passmap from the 31'-40' phase in the first half may advise the coach to alter the starting team formation to a 4-2-3-1 as opposed to a 3-4-2-1, as the players are clearly shown to be significantly more comfortable playing as a unit in this formation.

8.1.3 Arsenal F.C. Forward Players Shotmaps

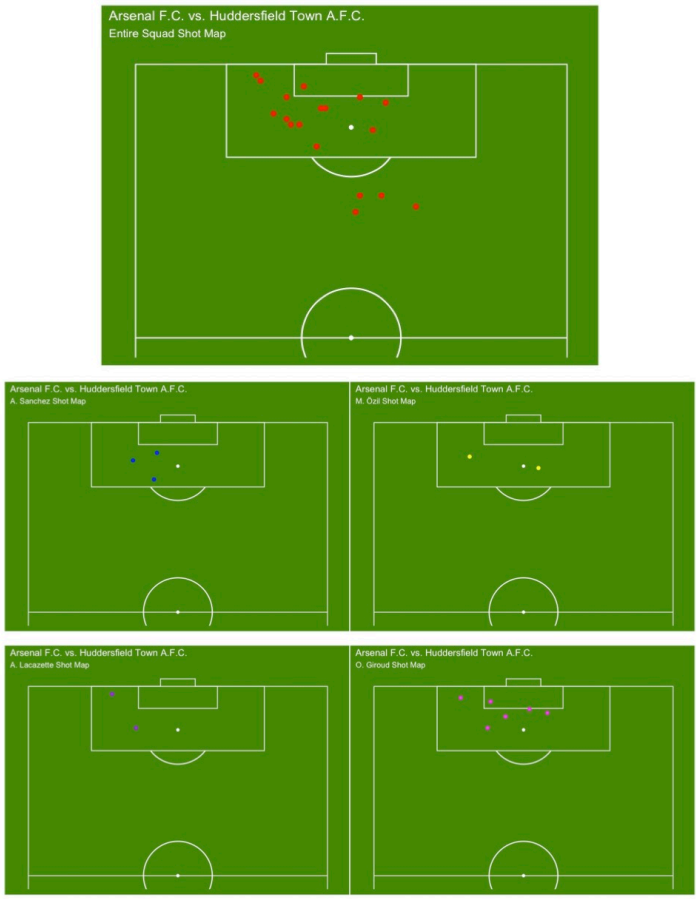


Figure 13: Entire Arsenal F.C. Playing Squad and Forward Players and Goalscorers Passmaps Versus Huddersfield Town A.F.C.

In 13, the shotmaps for the entirety of the Arsenal F.C. squad map and the forward players over the course of the entire Arsenal versus Huddersfield Town match can be seen. As previously observed, shotmaps can be extremely useful

in terms of displaying the stronger attacking minded players in a match and the positions of shots where perhaps the opposition players were struggling to contain the attacking nature of the team.

Looking at the shotmap of the entire game, including of the entire squad within 13, showcases that all bar 4 of Arsenal's shots take place within the 18 yard box. This is an important statistic, as the further away from the goal that a shot is taken, the less likely that it will hit the target - this is down to potential of blockage, deflection, ball speed slowing and so on. As a coach, establishing that 78% of the shots taken are taken within the 18 yard box. This, overall is a remarkably promising statistic.

Moving on to the shotmap of Alexis Sanchez seen in 13, it can be observed that Sanchez took 3 shots, with 33% of those resulting in a goal. Purely from the observation of the shotmap, the shots takne by Sanchez all emit from the left hand side of the 18 yard box, which is concurrent with the positional instructions given to Sanchez - performing in the left-winger role. This shotmap shows a promising aspect, which is that Sanchez - who is a confident finisher of the ball - can be given sufficient space within the 18 yard box to take a goal-bound shot as part of the 3-4-2-1 formation. This is more of a justification that the formation chosen has worked to one of the desired effects.

Analyses of the shotmap of Mesut Özil showcases that Özil tends to come more central as opposed to sticking to his assigned position. Comparing the Sanchez and Özil positions initially, both are playing as inverted wingers - with this being that they cut in onto their favoured right and left foots respectively. The Özil shotmap as well as the Sanchez shotmap is a display of this, as with traditonal wingers, it would be more likely to see crosses from the wing as opposed to shots made by the wingers. Özil is displayed as having 2 shots with a 50% goal ratio, however, Özil tends to deploy himself as an assister as opposed to a finisher and was played out of position in this match, deployed as right-winger instead of his preferred central attacking midfielder - which from this, would impress the

coach, as Özil has done his job regardless of his preference.

Examining the shotmap of Alexandre Lacazette, it can be observed that Lacazette took a total of 2 shots. For a typical number 9 - which is the role assigned to Lacazette - this is a relatively low value of shots to have, especially against a side of the calibre correlated to Huddersfield Town. The positive aspect of this is that Lacazette scored 1 goal in this game, which amounts to a 50% goal ratio. However, in terms of a coaches perspective, this is not particularly good enough, and perhaps, not through any fault of Lacazette, his desired impact has not been established. This may be due to the game-plan, the formation, or even the opposition's defence, but often in these circumstances a substitute would be advised.

Olivier Giroud was substituted on at half-time, most primarily due to Lacazette's failure to provide sufficient impact within the game. The shotmap of shots made by Olivier Giroud is shown in 13. A brief examination of this shotmap would showcase that Giroud was able to emit 6 shots ranging in positions inside the 18 yard box. Giroud was able to produce 3 times the amount of shots in 45 minutes that Lacazette was able to produce in the first 45 minutes - albeit, only resulting in a 33% goal ratio. Although Giroud corroborated to a lower goal ratio than the ratio that Lacazette provided, he was a bigger goal threat in that final third, providing 6 shots. One major difference between having Giroud on the pitch and Lacazette on the pitch is the respective results of the first and second half. In the first half, Arsenal won 1-0 as opposed to the second half where they won 4-0. Indubitably, this could be explained by the potential capitulation of Huddersfield Town in the second half, but could also be explained from the difference of strikers. Giroud could provide more of a hold-up style of play allowing both Sanchez and Özil to venture further into the box and take their chances, as opposed to Lacazette who tends to get in behind the oppositions defence - which for this game, perhaps, was not the best style of play to apply.

Looking at the game as a whole in terms of shots, Arsenal had an abundantly productive game, scoring 5 goals out of 18 shots. The major fallacy of this is that Arsenal only converted 28% of their total shots. Of course, a 5-0 scoreline is not a result to be overlooked, as with the passmaps, Arsenal wholly dominated Huddersfield Town. Another impressive aspect is the clean sheet that was kept. Often when a team pushes forward so far, it tends to leave itself open to counter attacks and are more likely to concede, so to score 5 goals and concede 0 is a hugely impressive feat. From these analyses, the likely outcomes would show that this 3-4-2-1 formation chosen has delivered the desired result, but perhaps a change in personnel would be more likely to result in the team working to it's potential ability as a unit, playing in this formation.

8.1.4 Match Interpretation

One of the main purposes of the analyses formed, is to enable the coach to comprehend the outcome of the game - more in-depth than they already will have seen. These analyses provide sufficient ground to cover both for positives and negatives and to identify key areas needing improvement in order for the team to work to their best ability and meet their potential.

Starting with Arsenal F.C., it can be easily observed that they were the driving force within this game. From 11, the tactical discipline of Arsenal can be examined, with the majority of the passmaps showcasing their ability to hold true to their starting formation of 3-4-2-1. As mentioned previously about their play-style, Arsenal tend to allow their left and right wing-backs to overlap the left and right wingers respectively, allowing the wingers to come short to look for a cutting pass or to drop into the box to provide more of a goal-threat.

The issue from this match in particular, is that the defensive structure of Arsenal is not easily examinable. In most games, their tends to be both an attacking and defensive phase of gameplay, within this match however, it looks likely

that Huddersfield were happy to defend their goal for the entirety of the match. This can be seen in 12 where the passmaps of passes made by Huddersfield Town can be seen to show passes majorly in the defensive third with some outliers venturing into the middle and forward thirds. Although a serious amount of positives and insights can be drawn from the attacking phases, it would be convenient to also be able to analyse a defensive phase in order to keep the team improving in both attacking and defensive aspects of the match.

Again looking at 12, it can be drawn that Arsenal were able to apply constant pressure on Huddersfield Town as can be established from the depth of the passes made by Huddersfield Town. This is an exuberant positive, as applying constant pressure requires a significant amount of stamina and lack of fatigue and injuries - this would be a reference of credit to the physio team in their ability to rest and recover the players. The downside to this, however, is that applying constant pressure will also increase the risk of injury to players and increase the risk of significant injury. For a team like Arsenal, who are most commonly also playing other competitions - domestic cups and European cups - the likelihood is that they will have matches every 3 or 4 days. Knowing this, and if they are aiming to play this play-style every game - which the likelihood is not as each game must be prepared for differently - then the most likely outcome is that copious amounts of players will be either injured or fatigued for long periods of time. Of course, players can be replaced and substituted, however, within the English Premier League there is a squad limit of 25 players, with 18 players being used in a match-day squad.

In terms of shots, Arsenal made 18 shots against Huddersfield Town in this game, 5 of those 18 resulting in goals. This amounts to a 28% goal ratio, which may seem like a poor return for the game, but when looking at the shots of the individual goal-scorers, the goal ratio appears a lot more promising. The main interpretation to take away from 13, is that Lacazette is perhaps not the best mould of player to be utilised within the 3-4-2-1 high pressing system that

Arsenal implemented, as opposed to Giroud who can be observed to have a significant impact, offering 6 shots in 45 minutes whereas Lacazette was only able to provide 2 shots in 45 minutes. Of course, this only applies if Arsenal are to play the same 10 other players in those same positions and roles. If, for example, Arsenal were to opt for a natural winger as opposed to an inverted winger like those deployed in the match versus Huddersfield, then Lacazette might offer a more suitable playing role, as natural wingers rely on forward runs behind the defence into the 18 yard box, rather than a player to hold up the ball and play in the wingers.

The game as a whole seemed to show an excellent attacking display for Arsenal. The defensive aspects of the game could not be easily examined due to Huddersfield Town failing to provide any legible attacking threat against Arsenal. This would be a game, which analyses would be utilised to indicate any mishaps or improvements to be made in the attacking phases of the game as opposed to the defensive phases - which perhaps rely on a team which will provide genuine attacking threat, and force Arsenal to defend.

9 Discussion

Within this section, the aim is to discuss the main aspects of the project, providing an overview in terms of the visualisations created and the knowledge which has been gained. The section will delve into the passmaps and shotmaps created, as well as looking in to the feasibility of recreating an analyses model.

9.1 Passmaps

Within these passmap visualisations created, it can be observed that the most useful version can be considered to be what is seen in 6. Referring back to the Expert Judgement section, splitting the game up into 10 minute increments to

view different phases of the match being analysed.

With respect to tactical analysis of a particular match, one could argue that analysis individual players proves to be significantly more useful than analysing the squad as a whole - albeit not to underestimate the analyses of the entire squad (Shaw & Glickman 2019). This can be suggested to be due to the fact that, analysing and individual player allows for the focus to be on one underlying factor, and not several as analyses of the whole team would suggest.

The analyses which took place around individual players, showcased passes made by Lacazette, Xhaka and Özil, in varying forms - with these being a striker, deep lying play-maker and a central attacking midfielder, respectively. Referring back to the expert judgement provided, these players were chosen in order to display the differences in both influence and impact in a match, depending on the position of the player analysed.

9.2 Shotmaps

From observations of these shotmap visuals, it is clear that the visualisation showcase in 9, seems to be the most useful for tactical analysis from a coach or manager. Again, referring back to the expert judgement received, these visualisations have been created to reflect the method of analyses that will most aid the coach or manager in the decision making element.

In terms of tactical analyses of a singular match, shotmaps can be utilised in order to perform analyses of both individual players and the entire squad of a particular match (Barreira et al. 2016). As mentioned in the Data Testing Shotmaps sections, the number of shots in a game is significantly far off the number of passes within a game. This knowledge allows for the use of analyses of the shots made by the entire team throughout the game - most preferably split into 10 minute increments, as mentioned previously.

9.3 Feasibility of Recreating an Analysis Model

Forming an insightful and corroborative analyses model requires a significant amount of resources. The data-set which was utilised in forming these analyses, utilised the method of video-tagging to gather the data. In terms of gathering this data for usage, higher ranked teams with significant assets will be able to employ this technique without any real issues being met (Baboota & Kaur 2019).

In terms of lesser funded teams, these forms of analyses will of course run in to some difficulty. In order to form these analyses, video-tagging is a significant asset but, indubitably, this will not always be possible for some teams to do or be able to afford. There are ways around this, for example, to implement an employee to be at the match and manually gather the match events, but this may result in missed or wrong data, and of course, takes a significantly longer time to do. Using a video-tagging software is not without error, however, the likelihood of missing data or an error will be much lower than that of an employee manually tagging match events (Thani & Heenan 2017).

10 Limitations

Within this project, several forms of passmap and shotmap visualisations have been created for analyses. As discussed in previous sections of this report, there are many positives which can be derived from these, however, they are not without their limitations. This section aims to delve into the limitations of this project, the visualisations created and the analyses which can be correlated to the visualisations.

10.1 The Data-Set

The data-set used within this project was an open-sourced data-set and not unlike a typical events based football data-set, however, the one main issue with this was the transparency in insight as to how this data was gathered. It was made known that a video-tagging software was used to tag events but was not easily examined as to how this data could be utilised - something that would be fairly transparent in a typical coach-analyst setting (Barton & Look 2001).

Another major issue that was met in terms of the data-sets, was that there was no given knowledge as to the outline of the clients data-sets. There was discussions around the fact that they would include such factors as heart-rates and positional data, but no exact factors were given. Again, in a typical coach-analyst setting, this type of limitation would not be faced.

10.2 Passmaps

Within previous sections, passmaps have been observed to be an exceptionally powerful form of analyses that can be utilised as part of tactical analyses. Passmaps have provided very important insights into terms that is comprehensible to the coach or manager, for a long period of time, but are becoming a much more utilised aspect of data analysis in the world of football in recent times.

Passmaps are an extremely useful tool in order to provide insight into how well a team is performing in terms of passing. However, passing does not always reflect on the outcome of the game - in similar terms as greater possession doesn't always mean that a team will emerge victorious. Talking about the EPL specifically, it contains 20 teams - some will be challenging for first, some will be battling relegation. One cannot expect a team, no matter how strong they are, to dominate passing in every game - for example, a team battling

relegation will likely struggle to make 100 passes in a game against a team fighting for first, so other forms of tactical analyses would need to be used to identify a correct game plan to tackle this game.

10.3 Shotmaps

As has been discussed in previous sections, shotmaps are an extremely significant visualisation that can be used to gain insight into tactical aspects of the game of interest being analysed. However, shotmap visualisations are not without their limitations.

Shotmaps are extremely useful in providing the position of a shot taken, however they do not always display the nature of the shot - for example, deflection, set piece, on or off target, and so on. These can, of course, be additional factors in to the analyses carried out, but require more events to be tagged and further specification into such events.

Similar to what was mentioned in the previous section, shots alike, do not necessarily correlate to the outcome of the game. For example, if a team who at the bottom end of the table face against a team a the top end of the table, the likelihood is that the team closer to the top will dominate the game in terms of shots made. However, it is difficult to determine as to whether they are dominating the game without video, as there are some tactical play-types - counter-attacking football - which relies on soaking up pressure, allowing the opposition to play their style of football and then seize an opportunity - whether that be a lapse in concentration or a mistake made by the opposition.

In simple terms, analyses referring to both section 9.1 and 9.2, are very important for tactical decision making. However, a major aspect needed is context. It is not enough to simply know the positions of passes and shots - they must be correlated with some form of context in order for them to provide true value

to the coach or manager.

10.4 Recreating an Analysis Model

One of the main limitations in the recreation of an analysis model is the cost of setting up an effective analysis team. In order to utilise data analyses to the highest potential, a team will require a team of data analysts, ranging in speciality - whether this be physiological or tactical analysis, to name but a few. This team of analysts, of course, comes with its downfalls. Employees will require a decent salary, with benefits and the more knowledgeable the analyst, the higher the price a team must pay to employ them - as with the majority of sectors. Indubitably, a team could utilise a trainee analyst, perhaps a student, on a volunteer basis, however, this could potentially impact the significance of the analyses formed due to the inexperience of the analyst.

Another significant limitation is the aspect of video-tagging within a match. Video-tagging is one of the most important assets that can be utilised in the world of football analysis, as it provides the analyst with sufficient detail to make a comprehensive and insightful analyses for the coach to make sense of. This style of software is extremely costly, with limitations of usage applying to lesser-funded teams. There would also be a significant cost to training employees in how to use this software to ensure that the best forms of detailed analyses can be established.

11 Recommendations

Carrying out this form of project, and the analyses that it consists of, is extremely time consuming, both in data gathering and performing the analyses. One such recommendation would be to invest into arranging a group to focus on these forms of analyses - for example, individuals focusing on event tagging,

individuals focusing on pass analyses and so on.

Another deterrent of data analyses in the footballing world, is the money involved, in order to get a high-level of sophisticated visualisations and analyses. In reality, money will be a huge factor in the output of data analyses, but it is not the pinnacle of factors. For example, simple investments such as GPS trackers provide a huge amount of benefit in terms of event tagging. Of course, an event tagging system will work in tandem with this GPS system, in order to identify and specify events which have taken place. These systems would be another such recommendation.

A significant recommendation for the client would be to invest in a video-tagging software. This would be an extremely important investment, both in terms of saving time performing analyses and also cutting out any errors or missed data. This software will of course come at a high cost, but will be utilised for a long period of time, and will eventually pay for itself in regards to the improvements of the squads and the potential for the team to achieve higher placed finishes in their respective leagues or cups.

12 Future Work

In regards of future work within this area, it would be significant to perform analyses which allow for varying statistical features to be determined. Once a software is in place to determine the positional data of players, the events data can be used in tandem with this to create a player passing network model - which can be continually updated after each match, when further detail becomes available.

Further work, should include the production of visualisations showcasing such features as xG, xA, shot probability, shot on and off target, touch maps, key passes and average positional maps or heatmaps, to name but a few. Production

of these visualised features, allow for easier decision making for the coach or manager. Just as importantly, it allows for the players to visualise areas of the pitch to exploit, or areas where a goal is more likely to be scored and so on.

13 Conclusion

In conclusion, it can be observed that visualisations of both passes and shots made by a specific team are an extremely significant tool to have for a coach or manager. They provide a great insight to aid in the decision making aspect of the tactical aspect of football.

Again, as mentioned previously, these forms of visualisations are not without their limitations. One major point which has been raised is the context of these events. One must know specific detail about both teams involved in the match to be analysed, including details of the players and areas of exploitation.

In order to round the report to a successful conclusion, one must review the main purpose of the project. This, of course, being to provide a comprehensive set of tools to give the client a skill-set to best utilise the data they have gathered. This report provides an expansive starting point for a beginner in terms of creating a multitude of data analyses which can be seen to be hugely significant for the client, as it allows them to utilise their data in a variety of ways that can improve them in numerous tactical aspects. Indubitably, these facets of data analyses are not the only way that they can be utilised - shotmaps could be altered for a map of key passes, passmaps can be further refined to show only passes made in the air and so on. This provides the client with a useful and understandable variety of tools that can be seen to met their initial request.

14 Appendix

14.1 R Packages for Analysis

14.2 R Code for Data Loading

```r
## Events ##

events_EPL <- fromJSON(file = "events_England.json")
# Loading Events dataset from JSON

events_df <- data.frame(do.call(rbind, events_EPL), stringAsFactors = FALSE)
# Making list into dataframe
```

14.3 R Code for Refined Data Loading

```r
## Initial Arsenal Data Loading ##

events_df_subset_1609 <- subset(events_df, events_df$teamId == 1609 | events_df$eventName == "Pass")
# Refining events dataframe to only show passes for team 1609 - Arsenal
# This will be key for filtering through the data, as teams can be changed, matchID can be added, etc

players_df_subset_1609 <- subset(players_df, players_df$currentTeamId == "1609")
# This refines the players dataframe to only Arsenal players
# Allows us to easily find out playerID for pass maps of individual players

matches_df_arsenal_win <- subset(matches_df, matches_df$winner == "1609")
# This refines the matches dataframe to only those matches which Arsenal have won
# Allows us to easily find out matchID and results of matches

# matchID matches wyID in matches dataframe, so will use this to filter further
```

14.4 R Code for Exploratory Analyses

```
## Exploratory Analysis ##

afc_free_kick_all <- subset(events_df_subset_1609, events_df_subset_1609$subEventName == "Free Kick")
# Creating subset of amount of free kicks Arsenal have had to work out average free kicks per game

afc_clearance_all <- subset(events_df_subset_1609, events_df_subset_1609$subEventName == "Clearance")
# Creating subset of amount of clearances Arsenal have had to work out average clearances per game

afc_throw_ins_all <- subset(events_df_subset_1609, events_df_subset_1609$subEventName == "Throw in")
# Creating subset of amount of throw-ins Arsenal have had to work out average throw-ins per game

afc_fouls_all <- subset(events_df_subset_1609, events_df_subset_1609$subEventName == "Foul")
# Creating subset of amount of fouls Arsenal have had to work out average fouls per game

afc_touch_all <- subset(events_df_subset_1609, events_df_subset_1609$subEventName == "Touch")
afc_touch_all_bellerin <- subset(events_df_subset_1609, (events_df_subset_1609$subEventName == "Touch" & events_df_subset_1609$playerId == "167145"))
# Creating subset of amount of touches Arsenal and Bellerin have had to work out average touches per game

afc_bellerin_touch_WHU <- subset(events_df_subset_1609, (events_df_subset_1609$matchId == "2499760" &
                                                           events_df_subset_1609$playerId == "167145" &
                                                           events_df_subset_1609$subEventName == "Touch"))
# Creating subset of Bellerings touchs v WHU

afc_bellerin_touch_positions <- data.frame(matrix(unlist(afc_bellerin_touch_WHU$positions), ncol=4, byrow=FALSE), stringsAsFactors = FALSE)
# Unlisting positions column in dataframe to give a matrix of dimensions, 10x4
# This allows for use of individual columns in the plotting of the touch map

names(afc_bellerin_touch_positions)[1] <- "Y Start"
names(afc_bellerin_touch_positions)[2] <- "X Start"
# Renaming columns for ease and comprehension

afc_bellerin_touch_y1 <- as.vector(afc_bellerin_touch_positions[[1]], mode="integer")
afc_bellerin_touch_x1 <- as.vector(afc_bellerin_touch_positions[[2]], mode="integer")
# Defining variables to be used for x and y co-ordinates in pass map

ggplot(afc_bellerin_touch_WHU) +
  annotate_pitch(col = "white", fill = "chartreuse4", limits = FALSE) +
  geom_point(aes(x=afc_bellerin_touch_x1, y=afc_bellerin_touch_y1), col = "blue", size = 3) +
  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("Arsenal F.C. vs. West Ham United F.C.", "H. Bellerin Touch Map")
# Plotting Bellerin touch map for one game
```

14.5 R Code for Passmaps

```
## AFC v WHU Entire Team Passmap ##

afc_v_whu <- subset(events_df_subset_1609, events_df_subset_1609$matchId == "2500060")
# Creating subset of events dataset to only show those referring to matchID of AFC v WHU

afc_v_whu_pass <- subset(afc_v_whu, afc_v_whu$eventName == "Pass")
# Creating subset of only passes made by Arsenal in AFC v WHU

afc_v_whu_pass_positions <- data.frame(matrix(unlist(afc_v_whu$positions), ncol=4, byrow=FALSE), stringsAsFactors = FALSE)
# Unlisting positions column in dataframe to give a matrix of dimensions 1
# This allows for use of individual columns in the plotting of the passmap

names(afc_v_whu_pass_positions)[1] <- "Y Start"
names(afc_v_whu_pass_positions)[2] <- "X Start"
names(afc_v_whu_pass_positions)[3] <- "Y End"
names(afc_v_whu_pass_positions)[4] <- "X End"
# Renaming columns for ease and comprehension

afc_v_whu_y1 <- as.vector(afc_v_whu_pass_positions[[1]], mode="integer")
afc_v_whu_x1 <- as.vector(afc_v_whu_pass_positions[[2]], mode="integer")
afc_v_whu_y2 <- as.vector(afc_v_whu_pass_positions[[3]], mode="integer")
afc_v_whu_x2 <- as.vector(afc_v_whu_pass_positions[[4]], mode="integer")
# Defining variables to be used for x and y co-ordinates in pass map

ggplot() + annotate_pitch() + theme_pitch()

ggplot(afc_v_whu_pass_positions) +
  annotate_pitch(col = "white", fill = "chartreuse4", limits = TRUE, dimensions = pitch_opta) +
  geom_segment(aes(x=afc_v_whu_x1, y=afc_v_whu_y1, xend=afc_v_whu_x2, yend=afc_v_whu_y2), arrow = arrow(length = unit(0.25, "cm"), type="closed")) +
  theme_pitch() +
  direction_label() +
  ggtitle("Arsenal FC vs West Ham United, 4-1", "Arsenal Passmap")
# Creating passmap using ggplot/ggsoccer
```

14.6 R Code for Shotmaps

```
## Shotmaps ##


## AFC v EFC Squad Shotmap ##

afc_v_efc <- subset(events_df_subset_1609, events_df_subset_1609$matchId == "2499970")
# Creating subset of events data for AFC v EFC

afc_v_efc_shot <- subset(afc_v_efc, afc_v_efc$eventName == "Shot")
# Creating subset of shots made in AFC v EFC (Both Halves)

afc_v_efc_shot_positions <- data.frame(matrix(unlist(afc_v_efc_shot$positions), nrow=14, byrow=T), stringsAsFactors = FALSE)
# Creating data frame of positions of shots

names(afc_v_efc_shot_positions)[1] <- "Y Start"
names(afc_v_efc_shot_positions)[2] <- "X Start"
names(afc_v_efc_shot_positions)[3] <- "Y End"
names(afc_v_efc_shot_positions)[4] <- "X End"
# Renaming columns for ease and comprehension

afc_v_efc_shot_time <- data.frame(matrix(unlist(afc_v_efc_shot$eventSec), nrow=14, byrow=T), stringsAsFactors = FALSE)
# Creating data frame of shot timings
# Have to do this to determine if goal or not, as dataset does not include this information

afc_v_efc_shot_minute <- afc_v_efc_shot_time/60
# Converting seconds to minutes

afc_v_efc_y1 <- as.vector(afc_v_efc_shot_positions[[1]], mode="integer")
afc_v_efc_x1 <- as.vector(afc_v_efc_shot_positions[[2]], mode="integer")
# Defining variables to be used for x and y co-ordinates in pass map

ggplot(afc_v_efc_shot) +
  annotate_pitch(col = "white", fill = "chartreuse4", limits = FALSE) +
  geom_point(aes(x=afc_v_efc_x1, y=afc_v_efc_y1), col = "red", size = 2) +
  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("Arsenal F.C. vs. Everton F.C.", "Squad Shot Map")
# Creating shotmap using ggplot/ggsoccer
```

References

(n.d.*a*).

URL: *https://www.premierleague.com/stats/top/clubs/total*$_c$*learance?se* = 363

(n.d.*b*).

URL: *https://www.premierleague.com/stats/top/clubs/touches?se=363*

Baboota, R. & Kaur, H. (2019), 'Predictive analysis and modelling football results using machine learning approach for english premier league', *International Journal of Forecasting* **35**(2), 741–755.

Barreira, J., Vendite, C. & Vendite, L. L. (2016), 'Analysis of shots and passing sequence of a soccer team and its opponents during 2014 brazilian championship', *International Journal of Sports Science* **6**(4), 163–167.

Barton, J. M. & Look, H. (2001), 'Analog video tagging and encoding system'. US Patent 6,215,526.

BBC (2012), 'The on-field history of rangers football club 1873-2012'.

URL: *https://www.bbc.co.uk/news/uk-scotland-glasgow-west-15307021*

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S. & Matthews, I. (2014), Large-scale analysis of soccer matches using spatiotemporal tracking data, *in* '2014 IEEE International Conference on Data Mining', IEEE, pp. 725–730.

Buldú, J. M., Busquets, J., Martínez, J. H., Herrera-Diestra, J. L., Echegoyen, I., Galeano, J. & Luque, J. (2018), 'Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game', *Frontiers in psychology* **9**, 1900.

Cintia, P., Rinzivillo, S. & Pappalardo, L. (2015), A network-based approach to evaluate the performance of football teams, *in* 'Machine learning and data mining for sports analytics workshop, Porto, Portugal'.

Curry, G. (2020), 'Putting football on the academic map: Eric dunning and the study of association football', *Sport in Society* **23**(10), 1606–1614.

de Football Association, F. I. et al. (2011), 'Physical analysis of the fifa women's world cup germany 2011$^{\text{TM}}$', *Zürich: FIFA* .

De Silva, V., Caine, M., Skinner, J., Dogan, S., Kondoz, A., Peter, T., Axtell, E., Birnie, M. & Smith, B. (2018), 'Player tracking data analytics as a tool for physical performance management in football: A case study from chelsea football club academy', *Sports* **6**(4), 130.

Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jiménez, S. & Sampaio, J. (2017), 'Exploring team passing networks and player movement dynamics in youth association football', *PloS one* **12**(1), e0171156.

Gonzalez-Rodenas, J., Lopez-Bondia, I., Calabuig, F., James, N. & Aranda, R. (2015), 'Association between playing tactics and creating scoring opportunities in elite football. a case study in spanish football national team', *Journal of Human Sport and Exercise* **10**(1), 65–80.

Hill-Haas, S. V., Dawson, B., Impellizzeri, F. M. & Coutts, A. J. (2011), 'Physiology of small-sided games training in football', *Sports medicine* **41**(3), 199–220.

Khan, A., Lazzerini, B., Calabrese, G. & Serafini, L. (2018), Soccer event detection, *in* '4th International Conference on Image Processing and Pattern Recognition (IPPR 2018)', AIRCC Publishing Corporation, pp. 119–129.

López Peña, J. & Touchette, H. (2012), 'A network theory analysis of football strategies', *arXiv* pp. arXiv–1206.

Memmert, D. (2019), 'Data analytics in football: Positional data collection, modeling, and analysis', *Journal of Sport Management* **33**, 574.

Moura, F. A., Martins, L. E. B., Anido, R. D. O., De Barros, R. M. L. & Cunha, S. A. (2012), 'Quantitative analysis of brazilian football players' organisation on the pitch', *Sports biomechanics* **11**(1), 85–96.

Müller, O., Simons, A. & Weinmann, M. (2017), 'Beyond crowd judgments: Data-driven estimation of market value in association football', *European Journal of Operational Research* **263**(2), 611–624.

Ortega, J. I., Evangelio, C., Clemente, F. M., Martins, F. M. L. & González-Víllora, S. (2016), 'Analysis of physiological, technical, and tactical analysis during a friendly football match of elite u19', *Sports* **4**(2), 35.

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D. & Giannotti, F. (2019), 'Playerank: data-driven performance evaluation and player ranking in soccer via a machine learning approach', *ACM Transactions on Intelligent Systems and Technology (TIST)* **10**(5), 1–27.

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D. & Giannotti, F. (2019), 'A public data set of spatio-temporal match events in soccer competitions', *Scientific data* **6**(1), 1–15.

Rangers Womans Team – Rangers Football Club – Info and club news (2008). **URL:** *https://rangersnews.wordpress.com/tag/rangers-womans-team/*

Rein, R. & Memmert, D. (2016), 'Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science', *SpringerPlus* **5**(1), 1–13.

Sacha, D., Stein, M., Schreck, T., Keim, D. A., Deussen, O. et al. (2014), Feature-driven visual analytics of soccer data, *in* '2014 IEEE conference on visual analytics science and technology (VAST)', IEEE, pp. 13–22.

Shaw, L. & Glickman, M. (2019), 'Dynamic analysis of team strategy in professional football', *Barça Sports Analytics Summit* .

Tenga, A. & Larsen, Ø. (2003), 'Testing the validity of match analysis to describe playing styles in football', *International Journal of Performance Analysis in Sport* **3**(2), 90–102.

Thani, S. & Heenan, T. (2017), 'The ball may be round but football is becoming increasingly arabic: oil money and the rise of the new football order', *Soccer & Society* **18**(7), 1012–1026.

Tierney, P. J., Young, A., Clarke, N. D. & Duncan, M. J. (2016), 'Match play demands of 11 versus 11 professional football using global positioning system tracking: Variations across common playing formations', *Human movement science* **49**, 1–8.

UEFA.com (n.d.), 'Rangers fc: Member associations'.
URL: *https://www.uefa.com/memberassociations/teams/50121–rangers/*

Vilar, L., Araújo, D., Davids, K. & Bar-Yam, Y. (2013), 'Science of winning soccer: Emergent pattern-forming dynamics in association football', *Journal of systems science and complexity* **26**(1), 73–84.

Wickham, H. & Wickham, M. H. (2007), 'The ggplot package'.