

The Football Project

Nahim Kamruzzaman
The Football Project
University of Maryland
Baltimore County

Abstract—A key problem in the domain of soccer player analytics, specifically in the fantasy sports category, has been the lack of resources and tools that allow fantasy (and in some cases real-life) coaches to effectively summarize and compare player data. Most of the current tools shove raw numerical statistics in the face of the coaches. In this document, we describe how our visualizations, which include radar charts, coxcomb charts, and heatmaps, function to address the issues at hand and allow FPL players and coaches to effectively compare soccer players in a manner that grabs their attention, allowing them to remember the best and worst qualities of their players, and make quick decisions on which players to add to their starting teams.

Keywords—compare, summarize, heatmaps, radar charts, coxcomb charts

I. INTRODUCTION

For this project, we've created several visual representations of soccer players' data. One of these uses data on players' game statistics such as their goals, shots, shots on goal, passes, dribbles, etc. Figure 1 is a radar chart created using python and the mplsoccer library [5, 10, 13]. Our radar chart has up to 8 statistics at each vertex which the user may choose and change through the use of drop-down menus. The user is able to choose two players at once using drop-down menus and have their radar charts overlay on top of each other so that they can compare 8 statistics between the two players. Rather than having a list of numbers side-by-side, a radar chart, that uses color hue to categorize the two players and radial spatial position to compare who is better in which metric, grabs the user's attention and allows them to see an upfront visual representation of the comparison between two players immediately. This allows a coach to give attention to a player's best and worst qualities and analyze them so they can decide how to set up their line-ups and tactics later. Furthermore, this survey [15] taken in 2022

reveals that 66% of fantasy players are employed full-time. Thus, the majority of fantasy players (our main target users) don't have the time to sort through hundreds of data sheets on hundreds of players. A much more helpful method would be the ones implemented by video games like FIFA and AIMLAB which split the data into categories and implement it with and summarize it into an eye-catching visualization like a radar chart. Superimposing two radar charts of two players allows FPL players to easily compare summaries of player statistics and decide between two players they'd like to add to their fantasy team. Our second visualization uses data on where and how many goals and assists a player had on the pitch. Using spatial data, we create a heat map of the soccer field that lets the user compare which areas players have scored or assisted the most from. These heat maps grant coaches and managers the ability to easily find trends and outliers regarding where a player is most likely to score on the pitch. They can then create a corresponding game plan to feed the ball to those areas. We partitioned the field into 2 halves on a single pitch and used color saturation as the channel to identify which areas that player scored more goals or assists from. This type of visualization allows users to easily identify clusters on the soccer pitch where a player tends to be most active. The areas with more goals and/or assists are much darker in color making it easy for coaches to analyze hundreds of shot/pass event data and understand it at a glance. Furthermore, we added density plots on the x and y axes of the pitch to display the frequency of shots taken at different lengths of the pitch so that managers have a different encoding of where a player likes to play most with the same data but focusing on vertical vs horizontal position on the pitch.

II. RELATED WORKS

A. Infotimation: Visualize Soccer Player Statistics

In this report [2], the author goes through the visualization created with CSS, HTML, JavaScript, and the D3.js library. Like our first idiom, this visualization includes a radar chart, which displays the top 5 and worst 5 attributes of a soccer player, compared to the average of all players. It also includes an interactive image of a silhouette of a player which displays statistics on how many goals a player has scored with that body part when hovered over. The final visualization is a pitch that uses soccer balls as marks to represent where a player scored and points to represent where a player shot from. Our visualization will include a radar chart similar to the one featured here, however, the user will be able to interactively choose which attributes to display on the chart, rather than the top 5 and worst 5. The user will also be able to see the radar of two players at once so that they can compare and contrast them. Finally, while we both aim to represent goal data on the pitch, our idiom will be using color saturation to represent the density of goals and assists of a player in a certain area, whereas this idiom shows only the location of goals on the pitch using one mark that has no color saturation.

B. NBA Shot Visualization

This visualization [4] looks at NBA data and uses color saturation to represent where on the court a player takes the most shots. Similarly, our second idiom is looking to use color saturation to represent where on a soccer field a player scores the most goals or makes the most assists. This visualization however also uses size to represent how many of the shots made it into the basket. The higher the percentage, the larger the square in area. Along with the court map, several line graphs compare the individual players' statistics with that of the league average and the average of all players that play that position. We hope to create these types of simpler visualizations, which are easy to understand and more commonly used, on the front page of our website so that the users can get a general idea of the top players in the league right now. This visualization does now

allow you to compare with other players, which our idiom intends to do.

III. IMPLEMENTATION

A. Coxcomb Chart Visualization of Percentile Data Between Two Players

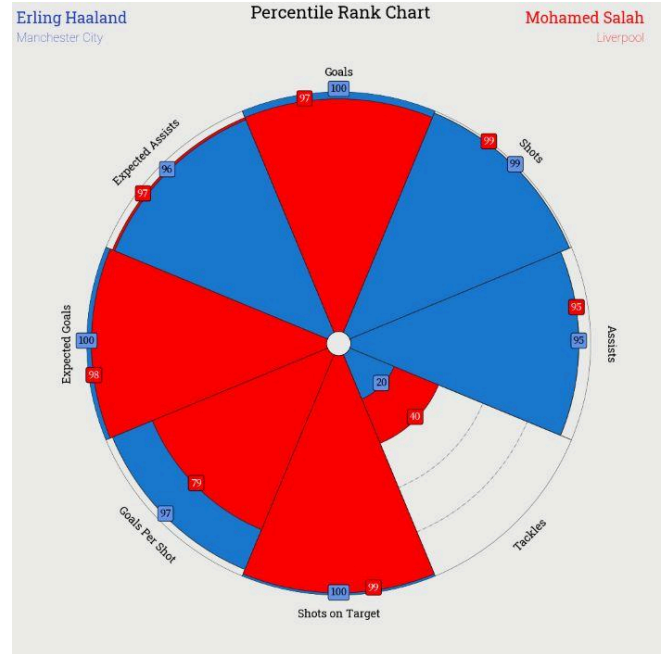


Figure 2. A comparison of percentile data of goal and shot attributes between two of the most highly rated forwards in the English Premier League. Erling Haaland of Manchester City (Blue), who has been on fire this season, and Mohamed Salah of Liverpool, who has had a slow start compared to other years.

Figure 2 is a coxcomb chart that allows you to compare up to 18 metrics between two English Premier League Players. Similar to the radar chart, color hue is used as a channel to categorize the two players. The marks for this visualization are the 2D shapes for each player that fill up a section of the circle which in this case would be the area. Unlike the radar chart, the shapes are not transparent so it is clear to the user who had the greater metric in a certain section since the greater player's shape will protrude out further away from the radial center. The exact percentile (rounded down to the nearest whole number) is also given by a text number in a box with that player's color.

For this visualization, we had to transform the stats for each player from “per 90” and “total” stats to percentile stats among all players in the league. This means instead of comparing

something like the number of goals or number of shots, we'd be comparing what percentile these players are in, amongst all players in the leagues, in terms of goals and shots. Then we use this percentile data and area (2D size of each shape in each section of the circle) as a channel to summarize and compare the percentile data of the metrics between the two players. In Figure 2, we can see that Erling Haaland is in the 99th percentile in the Premier League in terms of shots taken while Mohamed Salah is in the 97th percentile. However, Erling Haaland is in the 62nd percentile in terms of shots on target/shots while Mohamed Salah is in the 37th percentile for shots on target/shots. Thus we can conclude that Erling Haaland is much more accurate in his shots than Mohamed Salah.

Other than changing the raw data into percentile data we also changed the column names of our data set from things like "xG" to "Expected Goals" to make it clearer for our users to understand which metrics they are looking at.

Currently, this idiom allows the user to choose two players and choose the metrics they want to look at, however, we hope also to allow the user to choose to change the percentile data from being compared to all players to just being compared to players in that positions, i.e. defenders, midfielders, attackers.

This idiom has a limit of 18 metrics so that the data doesn't become too clustered and it's easy to differentiate between the metrics.

B. Heatmap Visualization of Player Shots Between Two Players

Figure 3 is a heat map that displays shot data on a soccer pitch with a density plot on each axis to inform the frequency of shots taken at that position. The visualization uses color hue as a channel to categorize the two players and uses color saturation to encode quantitative value on the map as a channel to convey which areas have a higher volume of shots for a player. This type of visual representation allows users to easily find clusters that help identify where players like taking shots from the most. The density plots on the axes allow the user to compare distributions of shots and find trends in which areas a player likes to shoot from the most. These density plots are a different encoding of the heatmap using the same

data. However, with the density plots, the user can compare the vertical and/or horizontal positions of actions specifically. The density plots take the data from the heatmap and separate the vertical and horizontal data. The density plots also use color hue to categorize the players and the hue is redundant throughout all three visualizations (blue for player 1 and red for player 2) so that the user may associate the same players in 3 different views. and use position on a common scale (horizontal for x and vertical for y) to convey the frequency of shots at an x or y position on the pitch.

The marks on this idiom are the areas on the pitch which represent the positions on the pitch where the players took shots from. The marks for the density plot are also areas that vary in height depending on the position on the x or y-axis.

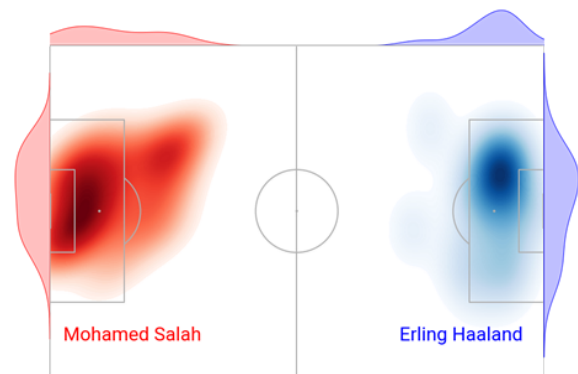


Figure 3. A comparison of the location of shots taken on a pitch between two of the most highly-rated forwards in the English Premier League is represented as a heatmap. Erling Haaland of Manchester City (Blue), who has been on fire this season, and Mohamed Salah of Liverpool, who has had a slow start compared to other years. **The dataset for this visualization is simulated** since event data of Premier League soccer games are extremely expensive.

In Figure 3 we can see that Erling Haaland, the star striker of Manchester City, trends towards shooting from inside the penalty area slightly toward the left side of the goal. Mohamed Salah, the star winger from Liverpool also trends towards mostly taking shots from inside the penalty box, however, we can see that he is also comfortable taking shots from outside the box on the right side of the goal. By comparing these two players, we know that Haaland is a close-range finisher and thus should have balls passed to him inside the penalty box area, whereas Salah should have balls both inside and on the outer right side of the penalty box given to him to provide for the

most comfortable shot. The distributions of the density plot further convey this idea since we can see that the graph on the x-axis spikes much higher inside the box for Haaland than it does for Salah.

For this visualization, we used event data to identify where shots were made on the x and y axis of a pitch. We then transformed that data which was given on a 60x80 x-y axis to one that was 80x100 so it can fit the field and on a larger pitch visualization. Other than that, the x and y coordinates were given and the library we created was able to use the individual shot locations to create a heatmap of the densest locations on the pitch in terms of shots.

This visualization currently uses simulated data which only contains data of all actions but we hope to filter the data to specific events like shots or passes in the future.

C. Radar Chart Visualization to Compare Raw Statistics Between Two Players in the EPL



Figure 1. A statistical comparison of goal and shot attributes between two of the most prominent up-and-coming rookies in the English Premier League. Erling Haaland of Manchester City (Blue), who has been on fire this season, and Darwin Núñez of Liverpool, whose performance has been lackluster in the eyes of his fans.

This visualization [10] is a radar chart that allows you to compare up to 8 metrics between two English Premier League Players. In this chart, we used the color hue as a channel to categorize the two players. The color for each player is based on the primary colors of their respective teams, thus Erling Haaland's data is sky blue and Darwin Núñez has red data. The areas for both are transparent so that you can see where both data overlap and differ.

The marks for this visualization are the 2D shape for each player, which in this case would be considered an area.

Here we used both vertical and horizontal positions on a common scale to order and compare which metrics a player was superior in. The position on each scale of the 8 metrics is given by the tip of the mark (or perimeter of the mark) for each player on that scale. For example, you can see the tip of Haaland's mark on the "Goals" scale is at around 15 while the point of the mark for Núñez is at somewhere around 2. Thus, we know that Haaland has scored 15 goals while Núñez has scored 2. Then we can look at a different metric on the radar like Shots Per 90 and do the same thing, but now using the scale for "Shots Per 90". We can see the tip of Haaland's mark (2D shape/area) is at around 4.5 while Darwin Núñez is much higher at around 6.45. Thus, we can compare the two players and order Haaland higher in goals but order Núñez higher in Shots Per 90. Using the position of the marks on this scale we now know that Darwin Núñez has many more shots than Haaland per 90 minutes of game time, however, Haaland converts many more of his shots since he has so many more goals than Núñez.

Most of the data displayed were given by our dataset, thus not much data transformation was required. We plan to add columns to our data set by dividing some of the data by the minute the player has played so that we can get more "Per 90" statistics since that is the average of the performance in ratio to the number of minutes that player has played which is much more useful to our task users (FPL players) than just total stats. To make the scales and range for the metrics of the radar chart, we took the maximum and minimum of both players in each metric and increased and decreased them by 25%

respectively to make sure all the data fit on the chart instead of going outside of it. This also ended up creating the scale on each line of each metric.

This idiom also has interactive features to allow the user to be able to choose the attributes they want to be displayed rather than hardcoding it in. The user may also choose the players they want to compare using a dropdown or search bar.

IV. RESULTS

For our alpha release, we created a program that takes in data on all Premier League soccer players and creates a radar chart that summarizes different metrics for two players. Now the user can choose any two players and compare up to 15 different metrics between them including shots, shots on target, goals, assists, etc. as well as these statistics per 90 minutes. Thus we added interactivity to our visualization which we wanted to do after the alpha release. We realized that while this allows users to compare raw data between two players it doesn't help in seeing how both these players perform when compared to other players in the league.

Thus we decided to add a second visualization in the form of a coxcomb chart which takes the raw player data and converts it into percentile data compared against all players in the league. Originally, we planned for this percentile data to be in the form of a radar chart as well however with the number of metrics we were trying to compare at once a radar chart had very low discriminability and was very clustered causing it to become difficult to interpret. A coxcomb chart seemed to be the perfect alternative since we were comparing percentile data, thus each metric was on a scale out of a hundred, and comparing between two players would be simple and effective since the larger percentile would be further out on the radial circle. Other than this additional visualization, our project has maintained its scope and timeline as planned.

The final visualization we added was our heat maps of shots on the soccer pitch. For this idiom, we originally created two separate pitches, one for each player. Not only did this make it harder for the user to compare the two players since there was a spatial distance between the two visualizations but it also led to a longer

computation time since we drew the pitch line by line. Thus it made more sense for us to put the shot data on one visual soccer pitch with one player on each side of the pitch. Since most players don't make shots from their own half of the pitch, you could easily have the shot data of each player on each half of the field and even if they did shoot from their own half, it would be easy to tell that this shot was an outlier since we are creating heatmaps which are generally catered toward finding clusters and outliers. We also realized that a heat map alone does not convey the frequency of shots taken on a position on the field, thus we added a density plot on the x and y-axis of the pitch for each player that conveys the frequency of shots taken at that position on the field. This is especially useful for coaches to understand how deep a player likes to play and shoot from on the field. Before the heatmap, we tried to create a cluster map on the pitch using points as marks to represent the exact position shots were taken. However, there simply wasn't enough shot by an individual player to create clusters, thus a heat map was much more useful to convey the general location on a pitch that players tend to shoot from.

Our users for this project were Fantasy Premier League players and our task for them was to allow them to compare different players to find the players who are currently performing best and are in-form and to allow them to add said players to their fantasy teams and earn the most points. We have accomplished this goal through our radar chart which allows the user to compare raw player statistics between two players in a variety of metrics, from goals to shots, to expected goals. Our coxcomb chart also takes into account these metrics and compares them to ALL players in the league so that the user can easily identify whether the two players comparing are in the top percentiles of certain metrics and which of the two is higher/better when compared to the entire league. Since in Fantasy Premier League, things such as goals, assists, saves, shots on target, tackles, etc. are what earn the players more points, the user can choose the specific metrics they want their players to have, then compare between two players to see if they rank high in these metrics amongst all players.

Our other users were managers and coaches and our task for them was to allow them to

compare players on their team and in the league to decide which players to include in their lineups and tactics to play around those players. The two idioms mentioned above aid in informing our coaches on which aspects of the game certain players are the best at. For example, when looking at Erling Haaland it becomes clear that he is a sharpshooter as he is in the 99 percentile of shots on target in the league and has more shots on target than even veterans like Mohamed Salah. Our heatmaps aid in helping managers choose their tactics since these notify where a player likes to shoot from the most. In the case of Erling Haaland, we can see that he excels in shooting from inside the penalty box slightly to the left of the goal. Thus a manager can get his midfielders and wingers to feed the ball into that area so that Erling Haaland can get more opportunities to shoot from that area and convert to goals. Managers can easily identify which area of the pitch has players shooting a higher frequency of shots and thus have the ball fed into those areas so that those players can make more shots they're comfortable with. And of course the more shots you take the higher likelihood of a goal.

To evaluate our algorithm, we analyzed each idiom/program's computational complexity. Our programs run at or below logarithmic time and space complexity, which is important since the input data size is quite large, thus an efficient algorithm was necessary so that the process does not take a huge amount of resources as the size of the input data increases. We measured the system time and memory for each idiom as we repeatedly tested pulling up visualizations on individual player statistics and this time was, on average, less than 4 seconds. We have also analyzed system time and memory used when pulling up data on two players since comparing between two players will be a functionality in our programs and this time was similar to pulling up information on a single player and, on average, less than 4 seconds. To evaluate the effectiveness of our visual encoding and the interactions of our idioms, we will survey a group of Fantasy Premier League players to understand how helpful our visualizations were in picking their weekly teams. These people are constantly looking and analyzing data on soccer players to be able to make the best Fantasy Team possible thus the survey will question what parts they liked about

the idioms, what parts they disliked, whether they believe there are more effective ways to visualize the same data, whether there should be different data and different idioms to accomplish this task, and how much their weekly score changed using our idioms vs. previous. These questions will clarify whether our specific idioms help pick players for teams and whether there are more effective idioms or different data the task user is looking for. To evaluate our data/task abstraction we will pull up a visualization for random players from the premier league and analyze whether our visualizations are portraying data that corresponds with the data set. We will also ask whether the data/statistics of our idioms are effective in helping FPL players choose their starting lineups in our survey. We can do a follow-up survey after some time to observe if the FPL players are still using our idioms. If the FPL players are still using our idioms, then we know that it can keep their attention and aid them in decision-making over a period of time. To evaluate whether we addressed the domain situation, we can also ask our surveyors how our tools compare to other tools, such as those provided in the FPL app itself. If our tool proves to be more useful, we have addressed the problem, which was to effectively visualize player statistics.

FUTURE DIRECTION

Our goal for the future of this project is to publish it on a live website where it will update automatically with little admin input by pulling data after every match week from our data source. Furthermore, we hope to publish this and make it available to the general public and FPL players so that they can use the feature of our program and so perhaps FPL may adopt similar visualizations themselves for their application.

CONCLUSION

Our goal for the future of this project is to publish it on a live website where it will update automatically with little admin input by pulling data after every match week from our data source. Furthermore, we hope to publish this and make it available to the general public and FPL players so that they can use the feature of our program and so perhaps FPL may adopt similar visualizations themselves for their application.

REFERENCES

- [1] Gozhulovskiy, A. (2022) "Visualizing Football Game Data", Medium. Towards Data Science. Available at: <https://towardsdatascience.com/visualizing-football-game-data-6a124fab911b> (Accessed: October 4, 2022).
- [2] Dubois, Y. (no date) "Infootmation: Visualize Soccer Player Statistics", cs.ubc.ca. Available at: <https://www.cs.ubc.ca/~tmm/courses/547-17/projects/yann/report.pdf> (Accessed: October 6, 2022).
- [3] Sleight, H. (2015) "What is stats zone?", fourfourtwo.com. FourFourTwo. Available at: <https://www.fourfourtwo.com/features/what-stats-zone> (Accessed: October 7, 2022).
- [4] Beshai, P. (2022) "NBA Shot Visualization", Buckets. Available at: <https://buckets.peterbeshai.com/app/> (Accessed: October 7, 2022).
- [5] Khalid, I.A. (2022) "How to visualize soccer data using Mplsoccer in python", Medium. Better Programming. Available at: <https://betterprogramming.pub/how-to-visualize-soccer-data-using-mplsoccer-in-python-39e2eddfb7d9> (Accessed: October 10, 2022).
- [6] Bayer, T. (2021) "Premier League Radar Charts", Tableau Software. Available at: https://public.tableau.com/views/PremierLeagueRadarCharts/Dashboard?%3Aembed=y&%3AshowVizHome=no&%3Adisplay_count=y&%3Adisplay_static_image=y&%3AbootsTrapWhenNotified=true&publish=yes (Accessed: October 8, 2022).
- [7] Varlamov, A. (2020) "W&M football #football #soccer #SportsVizSunday", Tableau Software. Available at: https://public.tableau.com/views/WMFootball/WMFootball?%3Aembed=y&%3Adisplay_count=yes&%3Aorigin=viz_share_link&%3AshowVizHome=no (Accessed: October 9, 2022).
- [8] EPL XG table and scorers for the 2022/2023 season (2017) EPL xG Table and Scorers for the 2022/2023 season | Understat.com. Available at: <https://understat.com/league/EPL> (Accessed: October 9, 2022).
- [9] Football statistics and history | fbref.com (2015). Available at: <https://fbref.com/en/> (Accessed: October 1, 2022).
- [10] Radar charts (2021) Radar Charts - mplsoccer 1.1.8 documentation. Available at: https://mplsoccer.readthedocs.io/en/latest/gallery/radar/plot_radar.html (Accessed: October 2, 2022).
- [11] Patel, D. (no date) A beautiful game, Football Data Visualization. Available at: <https://dhvanil.com/football> (Accessed: October 4, 2022).
- [12] Slothfulwave612 (2021) Soccerplots/radar_chart.MD at master · Slothfulwave612/Soccerplots, GitHub. Available at: https://github.com/Slothfulwave612/soccerplots/blob/master/docs/radar_chart.md (Accessed: October 12, 2022).
- [13] A. Durgapal and A. Rowlinson, "mplsoccer is a Python library for plotting soccer/football charts in Matplotlib and loading StatsBomb open-data.," Quick start - mplsoccer 1.1.8 documentation, 2021. [Online]. Available: <https://mplsoccer.readthedocs.io/en/latest/index.html>. (Accessed: 06-Nov-2022).
- [14] Premier League stats | fbref.com (2022) FBREF. Available at: <https://fbref.com/en/comps/9/Premier-League-Stats> (Accessed: October 12, 2022).
- [15] Industry demographics - fantasy sports & gaming association (2022) Fantasy Sports & Gaming Association - Connecting Fans to the Sports They Love. Available at: <https://thefsga.org/industry-demographics/> (Accessed: December 15, 2022).