

Infofotation: Visualize Soccer Player Statistics

Yann Dubois

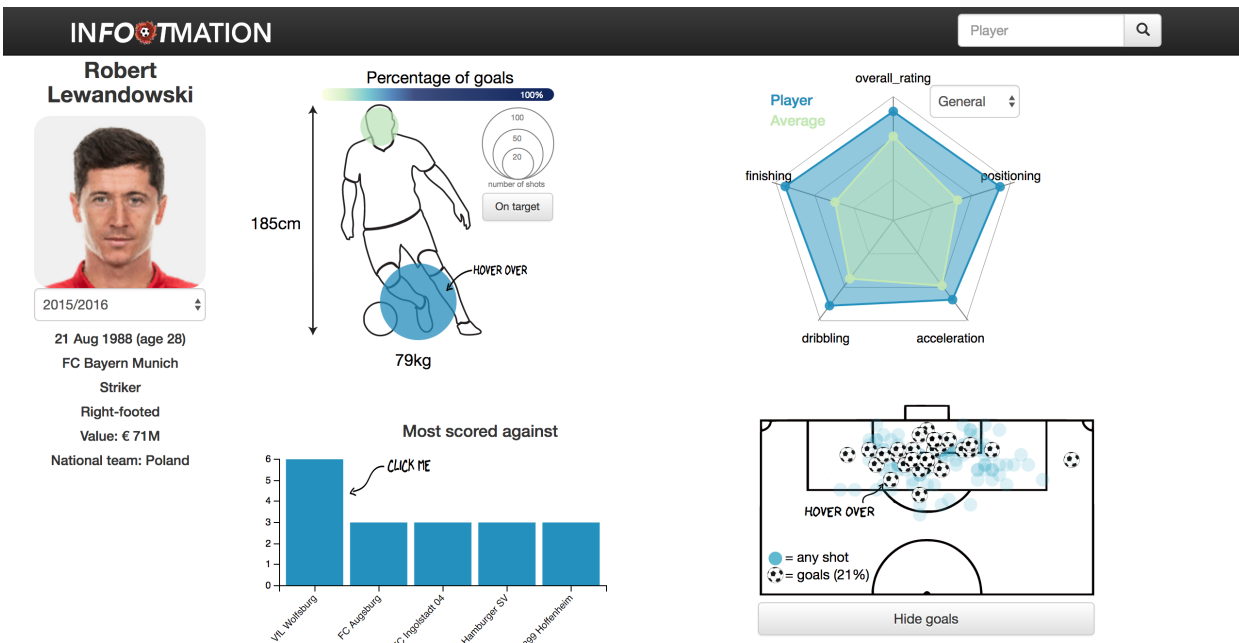


Fig. 1. Overview of the Infofotation website. Users can find a brief summary of the selected player statistics (left), a chart showing the goals/shot ratio depending on the part of the body used (upper middle), a radar chart comparing some of the players skills with the average player (upper right), a bar chart showing the teams against whom the player scored the most (lower middle), and a field showing the position of the shots and the goals made for the selected season (lower right).

Abstract—Soccer amateurs often like to learn at the statistics of their favorite players and to see how a player compares with the average. There already exists many tools to provide the statistics to fans. However, these tools rarely focus on the interactive visualization of this data. This is important in order to tame the spatial dimension of this sport, as well as to let users discover more advanced statistics in an enjoyable and easy way. Infofotation tries to fill in this lack of interactive visualizations in current tools. Infofotation supports comparing the players skills with the average, but it mainly focuses on all the details of the shots taken by the player. These details include the position of the player at the time of the shot, the ratio of goals per shots, the teams against whom the player scored the most, and the part of the body with which the player shot.

Index Terms—Soccer, Goals, Player statistics

1 INTRODUCTION

With an estimated 3.3-3.5 billion fans over the world [18], soccer is the most popular sport of our time. The Beautiful Game, as they call it, is being watched and played on every continent. As in other sports, fans enjoy discussing about and comparing player statistics and skills. To achieve this, they currently have the choice between a variety of different websites, which mostly focus on showing summary statistics in static tables. The wave of interactive visualizations does not yet seem to have reached this sport, at least not for amateurs. Indeed many visualizations were developed [15, 19, 20, 24], but mostly costly softwares which focus on professional needs such as for medias or coaches. These softwares were deployed to make precise analysis of games and not to get information about players. The reason of this lack of free interactive visualization solutions is probably due to the price

of data. The amateur community, which are often the ones building such websites, thus do not have easy access to the data. Infofotation is therefore a pioneer in the domain, by taking advantage of the recently published soccer dataset on Kaggle [12].

The main goal of Infofotation is to present details about each shot taken by a certain player, which is achieved by providing multiple views showing the position of the player at the time of the shot, the ratio of goals per shots, the teams against whom the player scored the most, and the part of the body with which the player shot. These views are interactively linked with each other, which enables the user to easily get access to a level of details that could not be achieved by static visualizations.

While providing amateurs with easy and simple access to detailed shots information, Infofotation also aims to replace static soccer player websites by summarizing player statistics and comparing his skills with the average player.

There are two main target users of Infofotation, both are soccer fans but their use of the website would be different, the first type of user is looking for a specific information about a player but cannot find it on static websites which do not show detailed enough information.

- Yann Dubois is with the University of British Columbia. E-mail: y.dubois@epfl.ch.
- Live demo of Infofotation: <https://yanndubs.github.io/Infofotation/>

The second type of user simply wants to enjoy playing around with an interactive visualization, while possibly discovering interesting patterns or information that he never thought of. Infofootmation thus focuses on all three parts of data consumption: discovery, presenting, and enjoying [14].

2 RELATED WORK

Infofootmation goal is to bring interactive visualizations to the world of amateur football, the idea is novel but the technique is not, it mostly tries to transfer method used in other sports to soccer in order improve . The following sections explore the related work, by detailing and describing the contributions made compared to: the different available interactive visualizations in soccer, the current solutions being used for summarizing players statistics, and the solutions to the same problem in other sports which has been a great source of inspiration.

2.1 Soccer visualization

Soccer visualization is a relatively new field, indeed it has been noted [4] that professionals such as coaches have always been more reticent to use data than in other sports. These professionals are starting to change mentalities and have an increasing trust in sports analyst, which has motivated some pioneers to make the data more understandable by interactive visualizations. In the last few years we have thus seen an increasing number of visualization softwares, that mostly focus on supporting analyst in exploring soccer game data. Indeed, previously, clubs typically hired video analyst to detect vulnerabilities and improving their own team's behavior, according to a German video analyst this took up to three working days [10]. Additionally, video analysts manually inspected recent games of opponent to prepare for upcoming games. There are no fully automated alternatives yet, but softwares have been deployed in order to facilitate and improve the analysis of a game by sports analysts.

One of the first visualisation method which effectively showed and summarized a game by showing the statistics, the context, the position and the game phase, is SoccerStories which was presented in 2013 [15]. Other papers focused on this task while trying to classify each of the actions of a game based on the extraction of meaningful features [20, 24]. In 2016, Gyarmati et al. went one step further to represent the physical performance of players by concentrating solely on analyzing and summarizing the movement of players in a game [11]. The output of their algorithm are movement characteristics of the player, which can be used to identify uniqueness and consistency of players as well as similarities between players. They use the uniqueness and consistency to plot players in a scatter-plot, this plot would nicely blend with Infofootmation and although it has not been implemented due to the lack of required data it definitely should be looked more into for future work. It should be noted that many other algorithms have been implemented to quantify the performance of individual players, but they did not have a visualization component [8, 9, 13, 25].

On the web, we also can find some interactive soccer visualization but most of them aim to visualise specific competitions. For example, many visualizations were specifically implemented for the World Cup in 2014. Each of these method focuses on a different aspect of the competition. The New York Times made a visualisation to show how to make fairer world cup groups [16]. FiveThirtyEight [17], Bloomberg [5] and Yahoo [23] each made a very different visualisation to understand and see the predicted winner at each phase of the competition.

There are also some mobile applications that focus on showing single game data. One of the most known one, is Stats Zone, which effectively shows exact player movement and statistics both during a game and afterwards [6].

2.2 Summarize Soccer Player Statistics

Although visualizations softwares have been deployed to support sports analyst and coaches analyze whole games, there are still only few solutions for visualizing players statistics . In the literature, there are only two solutions for visualising player statistics that I am aware of. First, in 2010, Soccer Scoop was one of the first softwares that tries to extensively use data to support choices of professional in the soccer

domain, it aims to easily measure the performance of a player and compare him to other players [19]. Although the idea is very similar to Infofootmation, its potential user was different. They were indeed concentrating on soccer team managers that would use it before signing players to a contract. The visualization was therefore very focused on numbers and not at all on aesthetics and easiness of use.

The current way amateurs get access to the summary statistics of their favorite player is through static websites. There is a big choice of these static websites considering that most sports website have one, but the go-to website is Transfermarkt [3] which shows in a very clear and effective way, the statistics of the player as well as news about him. the only interactivity is a simple tooltip of a line chart that shows the value of the player with respect to time. There seems to be currently only two websites that interactively present the data of multiple players, Pointafter [1] and Squawka [2]. The former is relatively complete but only shows player of the british premier league. It also focuses on statistics either from the whole career of a player or on a single game but not on the season data. The visualizations used are only tables or bar charts which takes a lot os space and doesn't make it easy to understand. Squawka on the other hand is much more comparable to what Infofootmation tries to achieve. The data is available by season and each major European league. Different interesting charts are being used to convey the data. One of the major issues with the solution proposed, is that they only show a single chart at a time. This is not as optimal as comparing juxtaposed facets due to memory load needed to remember the previous facet [14]. The link between plots is thus not as smooth as it could be for users to really enjoy interacting with it, in addition to that the page is very text heavy which makes it hard to focus on the plot and can be disturbing. The charts used are for their part very interesting, Infofootmation thus tries to build from Squawka by trying to make a better use of interactivity as well as facilitating usability.

2.3 Other sports

As previously stated, the soccer community has historically been more reticent or unable to use as many analytic visualisation methods as in other sports. On the other end of the spectrum, the basketball community has always been very data driven. This is partly due to the will of the National Basketball Association (NBA) to distribute freely part of its data. Infofootmation thus inspires itself from some basketball visualizations. One visualization method which is extensively used in basketball, is the position of the player when he shoots and the percentage of the player at that position. This is very often encoded with a discrete heat map on a field [21]. This method is not often used in soccer as most softwares focus on a single game which makes this visualization not adapted due to the low number of shots per game compared to basketball. As Infofootmation shows data from an entire season, it uses a comparable method. An interesting and novel paper that tries to show the impact of a player on his team is the Plus-Minus Plot [22]. In basketball, the Plus-Minus statistics is a widely used "advanced statics" that aims to represents how well the team plays while a certain player is on the field. Being able to derive and present such a statistic in Infofootmation would be very useful, but would require data that I could not get. This paper not only describes effective plots to show this statistic, it also reminds us of the importance of simple plots such as bar charts.

One web visualization which really inspired Infofootmation due to its simplicity and effectiveness is Buckets [7]. The visualization is close to what Infofootmation tries to achieve, indeed, it shows a single page per player with a summary of his statistics as well as multiple different charts that are linked and allow the user to find interesting information with interactivity.

3 TASKS

There are two main target users of Infofootmation, both are soccer fan but their use of the website would be different, the first type of user is looking for a specific information about a player but cannot find it on static websites which do not show detailed enough information. The second type of user simply wants to enjoy playing around with an interactive visualization, while possibly discovering interesting patterns

or information that he never thought of. In both cases information is intended to satisfy the curiosity of football fans that are eager to have an in-depth view of their favorite or most hated players the major difference is how much they are willing to interact with the visualization. Although the general tasks are the same for both, the solution has to be two fold: quick/effective but at the same time giving the possibility of going into more details. Information thus focuses on these tasks:

- Discover part of the body with which each player shoots the best and/or the most .
- Compare player skills to the average skill of other players and discover his strength and weaknesses.
- Summarize basic player statistics
- Discover the position of the player when he scored each goal.
- Compare the players performance depending on the opposing team.
- Enjoy football visualization and statistics as a fan.

4 DATA

Soccer data is normally very expensive and not open to public. This is a major barrier to visualisation solutions for amateurs, which are often made by simple fans. In June 2016, the first large scale free dataset was publicly released on Kaggle, the European Soccer Database was obtained by web scrapping by Hugo Mathien [12].

Most of the data used comes from the "European Soccer Database" and was simply downloaded as a .sqlite database. The dataset contains details about every game played in the eleven major European leagues from 2008 to 2016. There are thus 25979 matches and 11060 players. As Information focuses on players, some of the data could not be used. The useful data included the players: name, birthday, weight, height, potential, overall rating, preferred foot, FIFA EA games rating of multiple skills (passing, shooting, attacking rate, shot power), and details about each shot taken by the player (position, type, opponent team, time, data). The data set is unfortunately not regularly updated.

Although the data already contained many details, some very basic information about the players was not present. I therefore had to do some additional web scrapping with R. The additional information was taken from Sofifa.com and includes the players picture, market value, salary, team and national team, and position.

The database was separated in 7 different tables (Country, League, Match, Player, Player attribute, Team, and Team attribute). All the preprocessing was done in R, besides joining tables, filtering the interesting information and cleaning the data, a large part of the work was aggregation.

5 SOLUTION

5.1 Player Summary

The player summary Fig.2 gives a general summary of the selected player. It shows in order the players picture, his age , his current team, his position in the team, his preferred shooting foot, his current value on the market, and the national team for which he plays (if he does). Note That in the middle the user can also select the season as well as select all seasons.

5.2 Silhouette

The silhouette graph Fig.3 is a novel chart that intuitively shows ratio of goals made per total shot taken. It distinguishes between headers and shots with a foot. The ratio is color-encoded and the total number of shots is size encoded, not that the radius is proportional to the square root of the number of goals in order to keep the area linear proportional to it. The exact numbers can be found by hovering over one of the bubbles Fig.4 . Note that the user additionally choose to show the ratio of on target shot Fig.4. The height and weight of the player can also be seen.

Robert Lewandowski



2015/2016

21 Aug 1988 (age 28)

FC Bayern Munich

Striker

Right-footed

Value: € 71M

National team: Poland

Fig. 2. The summary of striker Robert Lewandowski

5.3 Skills chart

The skill chart Fig.5 is a radar chart which shows the general skills of the player as given on the video game FIFA16. Note that the user can chose between the general five skills, but also the best and worst 5 skills of the player compared to the average player in the database Fig.6. Hovering on the nodes gives the exact quantitative score Fig.6.

5.4 Opponent team chart

The opponent team chart Fig.7 is a simple bar chart which shows the three teams against whom the player scored the most. Note that if the player scored the same amount of goals against multiple team, the chart will display more than three teams.

5.5 Shots position chart

The shots position chart Fig. 8 shows the 2-dimensional distribution of all the shots taken by the player (in light blue) as well the goals scored (shape encoded). The percentage of goals is also color encoded using the color scale of the silhouette chart on Fig. 3. By hovering over a ball, the user would get information of the team against which the goal was scored, the time in the game, the date as well as additional information on the shot (distance shot, bicycle kick, penalty ...). The user can also chose to hide the goals in order to better see all shots taken. This chart is linked to the the bar chart as we can see in Fig. 9. This enables the player to filter the shots by team against which it was scored.

6 IMPLEMENTATION

In addition to the default web development languages (CSS,HTML, and JavaScript) as well as R for the preprocessing and web scrapping, the additional libraries have been used:

- **D3.js:** This library was used for all the interaction between graphs, it was the main library used. I only used D3 to implement the silhouette chart as well as the field chart. Additionally I modified the radar chart and implemented the interaction with the bar chart using D3.

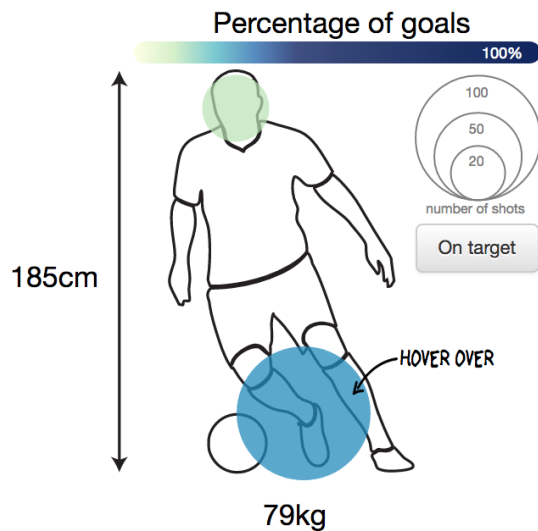


Fig. 3. Silhouette chart showing the distribution of shots by part of body.

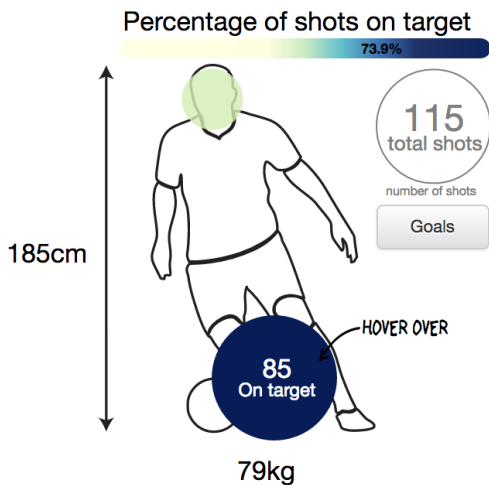


Fig. 4. Silhouette chart showing the interactive response to hovering as well as the possibility of showing changing from goals to on-target shots.

- **Crossfilter.js:** I used crossfilter to store all my data on the website. This enabled an easy quick filtering by: season, player, and the filtering done by the bar chart.
- **JC.js:** JC is a library that uses crossfilter and D3 to make simple charts, I used it to draw the bar chart.
- **Bootstrap:** I have used bootstrap to make the website responsive to the size of the screen and mobile friendly.
- **Radar-chart-d3:** Although not a library I used this example on github of a d3 radar chart as a working base for mine: <https://github.com/alangrafu/radar-chart-d3>.

7 SCENARIO WORK-THROUGH

The visualisation proposed, targets football fans that want to enjoy visualizing statistics of their favourite, most hated or other interesting players. It mainly tries to show all available statistics at the same place and in an enjoyable way. Below, I take every targeted task and I try to explain the proposed infovis solution as well as a scenario of use.

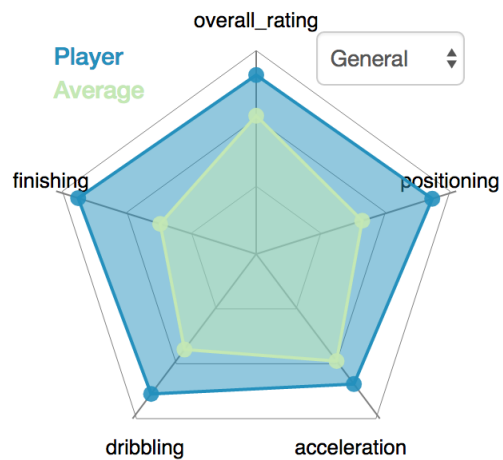


Fig. 5. Radar chart comparing the general skills of a player to the average.

For the scenario of use, let's suppose that Ben is a huge Bayern Munich fan. Let's also suppose that Sacha who hates Bayern's best player, Robert Lewandowski, goes to have a lunch with Ben. They start what seems to be a peaceful lunch but suddenly start to talk about football. Sacha states that Robert Lewandowski is far from being the current best forward player. In order to prove him wrong Ben decides to show him Lewandowski's infotainment page. They see a single page subdivided (using juxtaposed facets) into multiple visualisations that target specific tasks Fig. Fig. 1.

Sacha argues that Lewandowski uses too much his left foot. He thus shows Ben the silhouette chart on Fig. Fig. 3 that clearly shows that Lewandowski uses a lot his head although it has a bad goal percentage. To mark his point, he hovers over his head which gives the exact number of shots and goal percentage. Ben has to agree on the 2016 season, but he argues that this is only true in the selected season. He then selects all season to prove his point.

Initially the radar plot Fig. 5 shows the 5 most general football skills depending on the player's position. As Lewandowski plays as a forward it shows his: speed, accuracy, dribbling skills, shot power and finishing score. Sacha quickly checks the worst checkbox which changes the shown skills to indicate Lewandowski's worst skills compared to other players of his position Fig. Fig. 6. Ben finds that unfair and thus clicks on the best checkbox that does the opposite. They thus choose to go back to the general one, as it is summarizing the important skills. The data doesn't prove that Lewandowski is the best forward player, but it clearly shows that he's well above average for every important skill.

After arguing for a long time, they still can't agree on whether Lewandowski is the best forward player of their time. At that point they see an interesting visualisation showing the position of Lewandowski when he scored his goals. Ben asks Sacha: Do you remember when he scored against Manchester United in 2010 from half court: he hovers over the ball near half court and sees that he's right: the shot was made in 2010 against Man-U at the UEFA Champions League (88th minute of the game). They start playing quizzes trying to remember against who / when was that shot made. Fig. Fig. 8.

As Ben and Sacha see that Lewandowski scored most of his goals against Milan, they are very happy. Indeed, both do agree on the fact that they hate Milan! They hover over the Milan bar to show only the goals in Fig. Fig. 9. that were scored against Milan.

8 EVALUATION

As soon as I had a decent implementation, I wanted to make it evaluated by potential users. This was not only important to evaluate the current work, but essential to make a good use of the remaining time

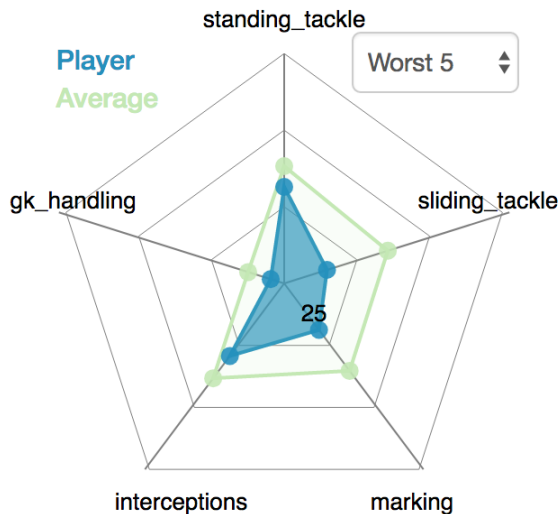


Fig. 6. Radar chart showing the worst skills of a player. This also show the possible interaction with the chart on hover.

at disposition. In this idea, I have asked 14 soccer fans, aged between 20 and 25 years old to say what did they like, what would they change, and what would they add.

8.1 What did the users liked

All of them were unanimous in saying that they enjoyed playing with it and did not understand why such softwares were not yet available. They complimented the clean design which looked professional and made them want to interact with it.

8.2 What changed

After the feedback I decided to add some features which are shown in yellow in Fig. Fig. 10. The users were indeed unanimously saying that such a website had to show a little table with summary details of the player, as in all of the static solutions. This forced me to do additional web scraping and to add a small summary of the player on the left. They also thought that the initial solution clearly lacked clarity, the silhouette chart was not easily understandable, they did not know about the interaction before I told them and finally they didn't understand the blue heat map on the field. I thus added some labels as well as little arrows on the default page showing how and with what users could interact. In Fig. Fig. 10 in red we can also see what I changed. This includes autofill for searching players as well as adding a description of the number of shots when the user hovers over the bubbles in the silhouette graph.

8.3 What should change

The users also gave many ideas of features to add or to change, but it was not implemented due to the time constraint. This includes showing assists on the field, comparing between players, add more colors and search by team.

9 CONCLUSION

In conclusion, Infootmation tries to offer interactive visualizations to soccer fans, to open the road to new visualization methods for the public. It is above all a website to satisfy the curiosity of the more than three billion soccer fans that are always looking for new ways to enjoy their sport. It comes with four primary charts, a so called "silhouette chart" that show the distribution of shots by part of the body as well as the ratio of goals per shot. A radar chart which enables users to find the strengths and weaknesses of their favorite player. A soccer

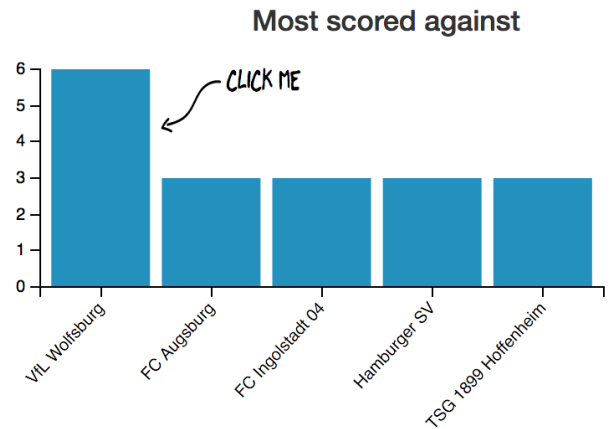


Fig. 7. Bar chart showing the three teams against whom the player scored the most. Note that here there are five due to an equal number of goals scored against four teams.

pitch that enables the user to have the precise position of each shot and goal made by a player. And finally a bar chart that shows the teams against whom the player scored the most and get additional details by filtering the dataset to show only these shots. The initial responses were very positive and I will be continuing to work on this project in order to finally make such an interactive visualisation available to soccer amateurs.

REFERENCES

- [1] Premier league players. Retrieved 28 April 2017 from <http://premier-league-players.pointafter.com/>.
- [2] Robert lewandowski. Retrieved 28 April 2017 from <http://www.squawka.com/players/>.
- [3] Robert lewandowski - spielerprofil 16/17. Retrieved 28 April 2017 from <http://www.transfermarkt.de/>.
- [4] Three at the back: Accelerating the pace of soccer analytics. Retrieved 28 April 2017 from <http://www.sloansportconference.com>.
- [5] World cup 2014 predictions and results - bloomberg. Retrieved 28 April 2017 from <https://www.bloomberg.com/graphics/2014-world-cup/#0,0,-1>.
- [6] What is stats zone?, Sep 2015. Retrieved 28 April 2017 from <https://www.fourfourtwo.com/features/what-stats-zone>.
- [7] P. Beshai. Buckets: Nba shot visualization, 2014. Retrieved 28 April 2017 from <http://buckets.peterbeshai.com/>.
- [8] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes. Using network metrics to investigate football team players' connections: A pilot study. *Motriz: Revista de Educaao Fısica*, 20(3):262–271, 2014.
- [9] J. Duch, J. S. Waitzman, and L. A. N. Amaral. Quantifying the performance of individual players in a team activity. *PloS one*, 5(6):e10937, 2010.
- [10] P. M. D. GmbH. Videoanalyse: Der 1. fc koeln ist das vorbild, 2016. Retrieved 28 April 2017 from <http://www.spox.com/de/sport/fussball/>.
- [11] L. Gyarmati and M. Hefeeda. Analyzing in-game movements of soccer players at scale. *arXiv preprint arXiv:1603.05583*, 2016.
- [12] H. Mathien. European soccer database — kaggle, 2016. Retrieved 28 April 2017 from <http://www.kaggle.com/hugomathien/soccer>.
- [13] I. G. McHale and Ł. Szczepański. A mixed effects model for identifying goal scoring ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(2):397–417, 2014.
- [14] T. Munzner. *Visualization analysis and design*. A K Peters Visualization Series, CRC Press, 2014.
- [15] C. Perin, R. Vuillemot, and J.-D. Fekete. Soccerstories: A kick-off for visual soccer analysis. *IEEE transactions on visualization and computer graphics*, 19(12):2506–2515, 2013.

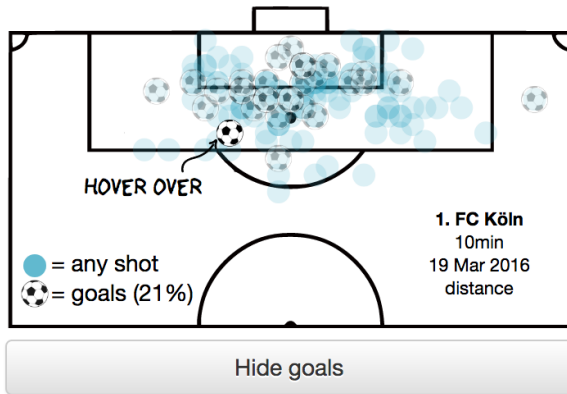


Fig. 8. Chart showing the 2-dimensional distribution of position of the player when he shot. Note that we also see additional details due to hovering.

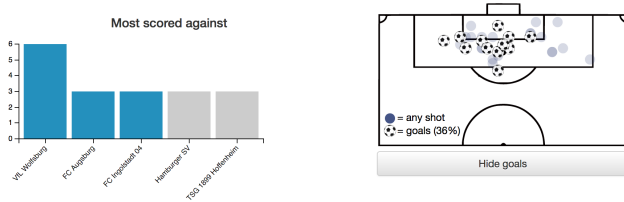


Fig. 9. An overview of the filtering process between the bar chart and the field chart.

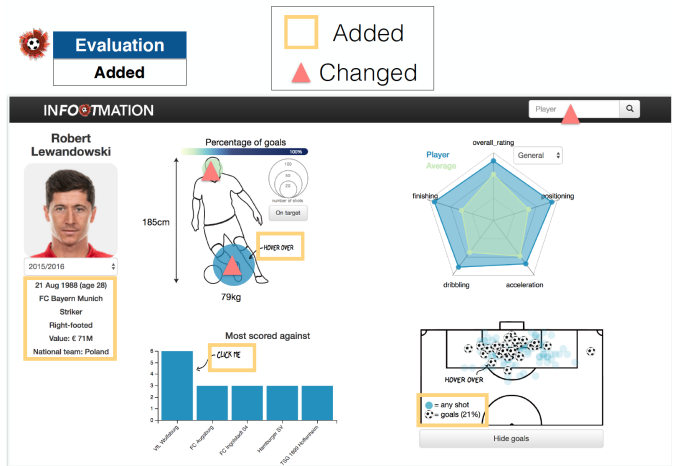


Fig. 10. An overview of what was changed and added after the evaluation

[16] G. A. Quealy and Kevin. A fairer world cup draw, Jun 2014. Retrieved 28 April 2017 from <https://www.nytimes.com/interactive/2014/06/03/upshot/world-cup-draw-simulation.html>.

[17] RitchieSKing. 2014 world cup predictions, 2016. Retrieved 28 April 2017 from <https://fivethirtyeight.com/interactives/world-cup/>.

[18] B. Russell. Top 10 most popular sports, Feb 2017. Retrieved 28 April 2017.

[19] A. Rusu, D. Stoica, E. Burns, B. Hample, K. McGarry, and R. Russell. Dynamic visualizations for soccer statistical analysis. In *Information Visualisation (IV), 2010 14th International Conference*, pp. 207–212. IEEE, 2010.

[20] D. Sacha, M. Stein, T. Schreck, D. A. Keim, O. Deussen, et al. Feature-driven visual analytics of soccer data. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pp. 13–22. IEEE, 2014.

[21] A. Shortridge, K. Goldsberry, and M. Adams. Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. *Journal of Quantitative Analysis in Sports*, 10(3):303–313, 2014.

[22] R. Sisneros and M. Van Moer. Expanding plus-minus for visual and statistical analysis of nba box-score data. In *The 1st Workshop on Sports Data Visualization. IEEE*, 2013.

[23] D. Skau. A look at the most interesting world cup visualizations and champion predictions, Jun 2014. Retrieved 28 April 2017 from <http://www.scribblelive.com/blog/2014/06/11/visualizing-predicting-world-cup/>.

[24] M. Stein, J. Häußler, D. Jäckle, H. Janetzko, T. Schreck, and D. A. Keim. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information*, 4(4):2159–2184, 2015.

[25] L. Verburgh, E. J. Scherder, P. A. van Lange, and J. Oosterlaan. Executive functioning in highly talented soccer players. *PLoS one*, 9(3):e91254, 2014.