FIFA Scouting Dashboard

An interactive scouting tool for the most popular sports video game in the world

INF552 - Data Visualization - 16.12.2022

Project website

Thomas Wimmer
TUM, IP Paris
thomas.wimmer@ip-paris.fr

Haileleul Haile
IP Paris
haileleul.haile@ip-paris.fr

1. Introduction

The football simulation FIFA is the most popular sports video game in the world[1]. While still being a simulation, to ensure an experience as realistic as possible, the players in the game have a number of attributes that are based on their performance in real-life. Through its popularity and detailed attributes that are well matched to reality, the video game often also serves as an indicator of which players are currently the best in the world.

While the game itself offers a very limited ability to compare two players, it only offers the ability to compare the current year's attribute values (the game is re-released each year with new player values). The dataset used in this project, on the other hand, contains data from the last eight years of the game and thus brings in a new dimension, namely the evolution over time. Further details on the dataset can be found in Section 2.

Career mode is an interesting feature of the FIFA franchise which allows users to become either a player or a manager over the course of the game. As a manager, the user is able to build teams and compete against other teams. An essential part of career mode is finding the right players for your team. That includes filtering by playing position, finding the best player for that position and making a trade-off between market value and player attributes. This can be a challenging task with more than 17,000 players.

The goal of this project was to develop an interactive dashboard that would allow for a more comprehensive analysis of player attributes. Furthermore, it gives users the ability to perform the task of a football scout, enabling them to find the best players for a specific area on the field. For this purpose, several different types of visualizations were developed and integrated into the website, available at this link[2].

2. Dataset

The dataset titled *FIFA 22 complete player dataset* is provided by Stefano Leone and was taken from Kaggle. It can be found at this link.

The dataset includes players' data from the career mode of FIFA 15 to FIFA 22. This dataset was scrapped from sofifa.com, which assigns a unique ID to the players throughout the different iterations of the game. It provides information about players such as their portrait, height, weight, club logo, and country flag which can be used to easily identify the player we are describing. Furthermore, it provides 6 general and 35 specific attributes which give an all-round information about the performance of the player. One of the most important features of the dataset is the feature `player_positions`; it shows the mainly played position of the respective player.

Players are generally ranked based on their overall score. We too used this score to rank players which will be shown later. The dataset is rich in that it also contains a potential feature which shows the potential overall score of a player throughout a career mode.

[1] According to https://www.rookieroad.com/sports/top-10-sports-video-games/.

[2] The code for the project is also provided in the submission and can be run locally

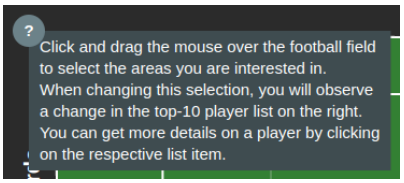


Figure 1. To help inexperienced users, we have included small buttons that display an explanatory text of the corresponding visualization when the user hovers the mouse cursor over it.

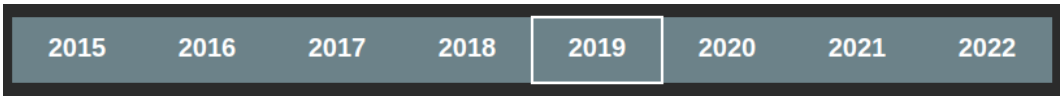


Figure 2. Selector bar for the year the user is interested in.

3. Visualization

The dashboard is embedded on a website and is made using D3.js. D3.js is a JavaScript library for manipulating documents based on data. It is a useful tool for bringing data to life using HTML, CSS and mainly SVG. We chose to work with D3.js because it provided the flexibility to realize our designs while enabling interactive visualizations which are an essential part of this project.

The dashboard is generally partitioned into three sections. The first partition consists of the Football Field Selection (see Section 3.1) and a Player Detail View (see 3.2). Below is a Player Comparison view (see 3.3) followed by a General Statistics partition in the bottom third (see 3.4). Each of these visualizations, their implementations and design choices made in the creation process are described in the following sections. At some places, we included small explanatory buttons on the website that can help inexperienced users with the navigation through the visualizations (as shown in Fig. 1).

3.1. Football Field Selection

The football field selector shown on Figure 4 is an important interactive part of the dashboard. It follows the idea of allowing users to scout players based on their playing positions. Often users are interested in players who can play in a certain area of the field, not just a specific position (as positions can slightly be adapted in the career mode through in-game training sessions). While the game itself does not allow direct comparison across different positions, this is possible in our dashboard.

While other parts of the dashboard directly use the time dimension of the dataset through showing the evolution of attribute values over time, the first section contains an animated year selector that can be used to switch between different years. This enables users to select the game version they are interested in and to navigate through the best players for this specific year. The selector is shown in Figure 2.

3.1.1 Design Choices and Implementation

Including the Goalkeeper, there are 27 possible playing positions that are available in FIFA. These positions are represented by two to three letters, which may be ambiguous to users. Hence, we mapped the playing positions to areas on a standard football pitch (with 3:2 height-to-width ratio, the standardized ratio for professional football fields).

After the initial mapping of the positions, the football field was modelled using d3.js. By using the native HTML object classes instead of a preloaded image, it is easy to adjust the colour, size and other properties of the pitch.

This is particularly handy when using d3's brush function. By clicking and dragging the mouse over the football field, the user can select specific areas of interest. The pre-calculated mapping of each pixel on the pitch to the defined positions allows a selection of all players playing in the selected area.

After some early drafts of this function and first user tests, we decided to show more clearly which positions on the field belong to the currently selected area. Therefore, we added dashed lines to indicate the different positions and also highlighted all selected positions in green when brushing. This functionality is shown in Figure 3.

Another design decision made in a later iteration was to display three text indicators to the left of the football pitch to show the orientation of the pitch. While this orientation may be intuitive for football fans, the text can help inexperienced

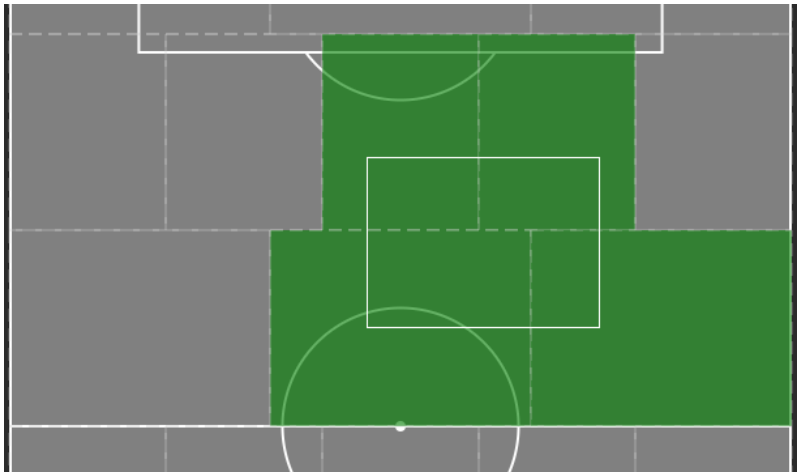


Figure 3. The green highlighted areas show the selected positions when brushing with the mouse (brush area indicated through white rectangle).

users navigate.

Next to the football field, the users can see the best ten players playing in the selected area of the pitch (according to their `overall` attribute value). The list updates on brush changes and shows the name, the overall value and an image for each player.

In Section 2, we mentioned that there are 6 general attributes describing each player. It is interesting for the user to see the general trend of these attributes for a selected area of playing positions on the field. When moving the brushed area over the football field, the change in this visualization also gives a better intuition about which attributes are more and which less needed for the different playing positions.

We decided to visualize these six key attributes using radar plots. While all attribute values are displayed in one graph, this visualization also fits the game-related look of the dashboard and allows for a nice visual change when the brush area is moved across the football field. The attribute values form the corner points of a hexagon that stretches across the entire chart. Thus, when changing the selection and thereby the respective data selection, we can observe a change in the area of the hexagon, which allows for an intuitive understanding of the key attributes of the players in the selected area. This type of visualisation is an element that reappears in various places in the dashboard.

The radar plot shows three different aggregations of the player data for the selected positions: the median, the lower (first) and the upper (third) quartile. By including the quartile information, users can also get a quick overview of the underlying distribution of the data: While the median values alone would not be sufficient to represent the diversity of the different position profiles, the information gained through the quartile charts allows for a better understanding of the data.

Experiments with using the minimum and maximum values for the six attributes instead of the quartile information showed that these values give a false indication of the underlying data distribution, as the maximum/minimum values for different attributes can come from different players and due to the diversity of players the resulting areas show a more extreme picture that gives a false intuition about the actual data distribution. The inclusion of quartiles and extreme values resulted in an overloaded visualization.

3.2. Player Detail View

The player detail view is an in-depth visualization that users can find in the first section of the dashboard. It is an extension of the football field selection, as it displays detail attributes of a player selected from the list indicating the top-10 players in the selected area. The idea behind this visualization follows the main objective of giving the user an all-round information about a player. Hence, once a user has searched for a player using their playing position and their general attribute using the football field and radar plot, they may select individual players to see detailed information about the player.

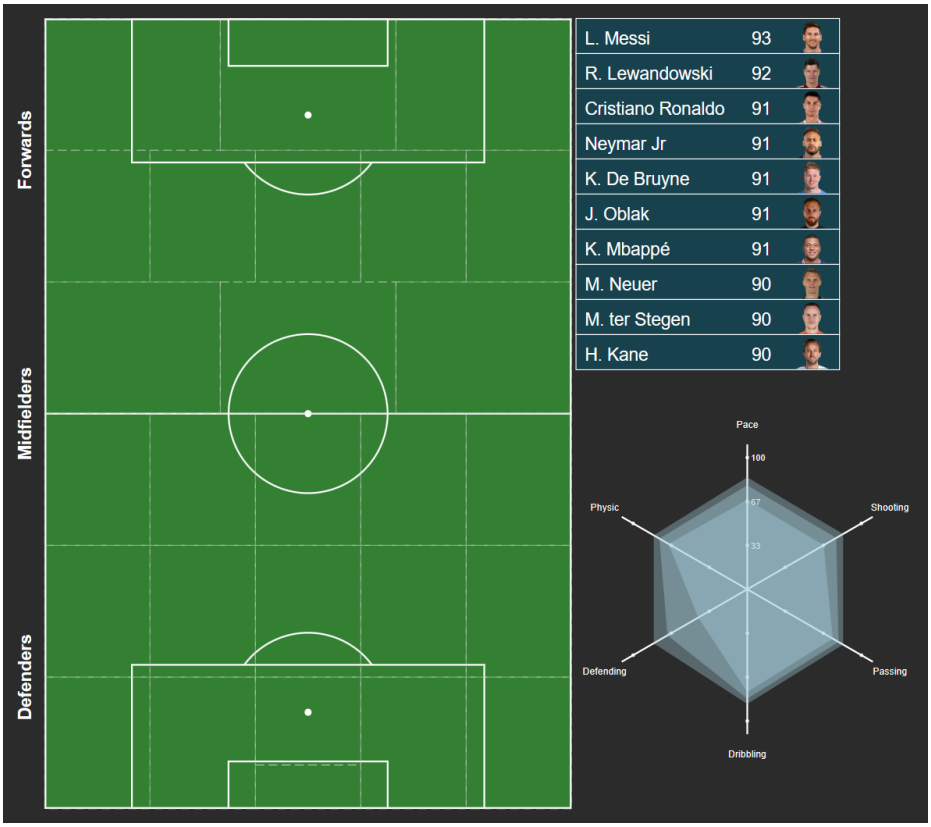


Figure 4. Football field selection (left), Selected position player's List (top right), and General attributes radar plot (bottom right).

3.2.1 Design Choices and Implementation

The dataset contains a lot of details on players, allowing us to display a lot of information. We are, however, constrained by both area and the need for simplicity.

The top section of the detail view contains basic information about a player such as name, age, weight and height as well as playing position. The player's portrait is displayed along with the club logo and country flag, making it easy to learn about the player. The top section also contains information about a player's market value and expected salary, which can be helpful for users who want to make a trade in career mode.

Next to the player's portrait, we show the player's overall rating in the game. When buying a player for a team in FIFA, it is important to know his potential for improvement. Since this is a potential overall rating, we decided to display it when the user moves the mouse cursor over the overall rating.

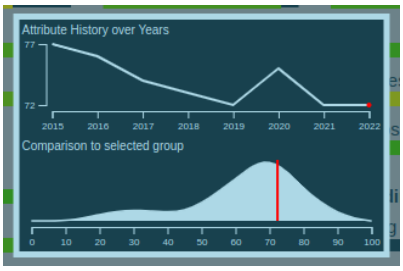
To the right of the player's portrait, the radar plot as introduced in the previous section is displayed showing the six key attribute values for the selected player.

In the lower area, all detailed attributes of the player are displayed. For better structuring, the attributes are grouped into several superordinate areas, each containing 3 to 6 attributes. While the colour-coded attributes already give a first indication of the player's strengths and weaknesses, they do not provide detailed information on how these values compare to other players in the game and to the values for the same attribute and player in another year.

To allow a more detailed comparison with other years and other players, we have added a small pop-up window that appears when hovering over a specific attribute (as shown in Figure 5b. The graphs show the evolution of the attribute value for that player over time (with the current year shown as a red dot), as well as the comparison with all players in the selected area of the football field for that specific attribute (with the player's value shown as a vertical red line).

It is important to note that we have explicitly designed the system to allow the user to change the selected area on the football field and therefore the group for comparison. In this way, a player can be compared not only with other players in his position, but also with a wider range of players, up to and including all athletes of the selected year.

(a) Full player details view



(b) Plots shown on hover over an attribute

Figure 5. The player details view allows an in-depth analysis of the different attributes of a player. When moving the mouse pointer over one of the attributes, a comparison with its value in different years and with the values of other players is possible.

It should also be mentioned here that we experimented with displaying the comparison with other players directly in the overall view of the player details, but this made the display look cluttered. A possible solution would have been to remove some of the attributes from the view, but this contradicts the original aim of allowing an in-depth analysis of a particular player.

3.3. Player Comparison

Player comparisons are an important part of world football. We see that such comparisons shape football culture. You often hear and see people comparing Lionel Messi and Cristiano Ronaldo and asking who is the greatest footballer of all time. Our platform provides a platform for this and other similar discussions to make a data-driven decision.

In the context of the game, player comparison is an essential part of the career mode. A user may be undecided whether to select one player for his team or another. This may depend on various factors, such as the player's overall score, the score of individual attributes, salary, market value, etc. For this reason, we have dedicated the second section of the dashboard to the player comparison between two players selected by the user.

3.3.1 Design Choices and Implementation

Before diving deep into the actual comparison panels, we have to consider again that there are more than 17,000 players in the dataset and it is difficult for users to scroll through all the players to find who they want to compare. To simplify this task, we first sorted the players in the selection list by their overall scores. This way the most relevant players will be displayed

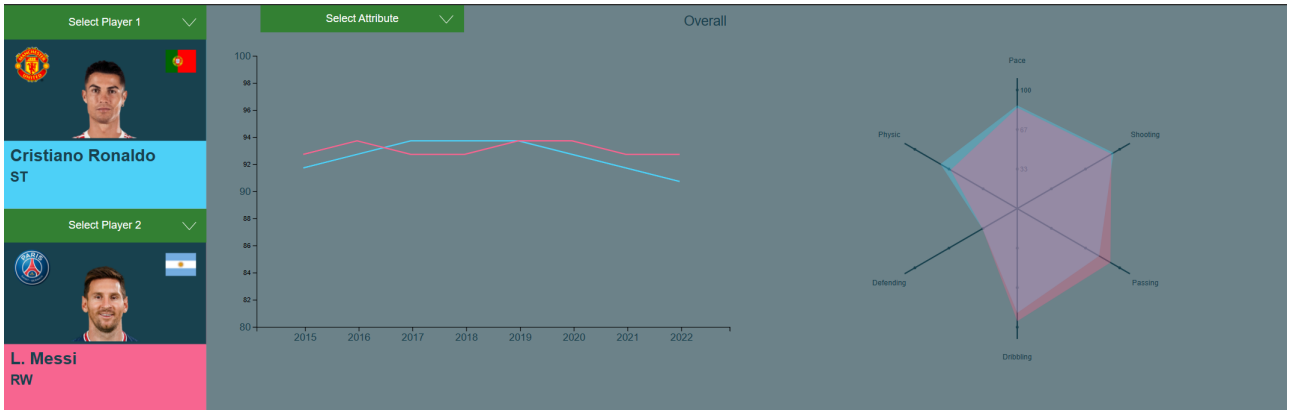


Figure 6. Players comparison section. Player selection menu (left), Line plot comparison (middle), and Radar plot comparison (right).

first. We also added a search bar for users to search a player using the player's name or a small sequence of characters they know in the player's name. The search is fast and simplifies the player selection.

The two selected players are distinguished by their colors, with player 1 having a light blue color and player 2 having a pink color. These colors were selected using Adobe's online color selection tool. They were chosen to be harmonious yet easy to distinguish from the background and from each other. The color of the line charts follows the colors assigned to the players on the Selection tab.

The first comparison display is a simple line graph with a static x-axis representing the 8 years in the data set. The user can use a drop-down menu to select any attribute used to compare the two players. The selection triggers a smooth-looking animation of the line chart.

An important design decision for this line plot comparison is how to handle the different magnitudes of the attribute values of the two players. This is even more pressing when the users have fairly similar attribute values which makes it very difficult to visualize the differences. To alleviate this problem, we made the scale of the y-axis dynamic ranging from a value slightly lower than the minimum of the two players' scores over the eight years (to give a good distance between x-axis and the bottom line chart) to the maximum value. To prevent users from not noticing that the y-axis scale changes when they change players or the comparison attribute, we added a smooth animation to the scale transition that shows how the scale increases or decreases.

On the right side of the player comparison section, we again use the radar chart. However, in this case, only the 6 key attribute values for the two selected players are displayed. This radar plot provides the user with a general comparison of the two players. It can also provide the user with information about the nuances of the differences between the two players. For example, two players who play the same position and have a similar total score can be distinguished using the radar plot based on the 6 attributes. In Figure 6 it can be seen that although the two players are quite close in overall score, Cristiano Ronaldo performs better in the areas of physicality and pace, while Lionel Messi performs better in dribbling and passing.

3.4. General Statistics

The final section of the dashboard is a general statistics section. The section consists of one plot that allows identifying correlations between different attributes.

3.4.1 Design Choices and Implementation

The general statistics section presents users with an easy-to-use scatter plot that is designed in a flexible way; the users can select the x-axis, y-axis, and hue attribute of the plot. The x and y axes selectors contain all attribute values, overall and potential scores, market values and physical metrics of players. The hue selector contains player positions, as well as information on leagues, the preferred foot, skill moves and weak foot attributes as possible choices for the hue. To prevent the hue from being indistinguishable because of two many individual colors, we decided to cluster positions in the general categories defense, midfield and attack, as well as only showing the 5 main leagues and grouping other leagues under "Other".

In order to have a clean plot that is not overloaded, we limited the players to the best 150 players of the year 2022. When hovering over a player point in the plot, more information on the player is shown.

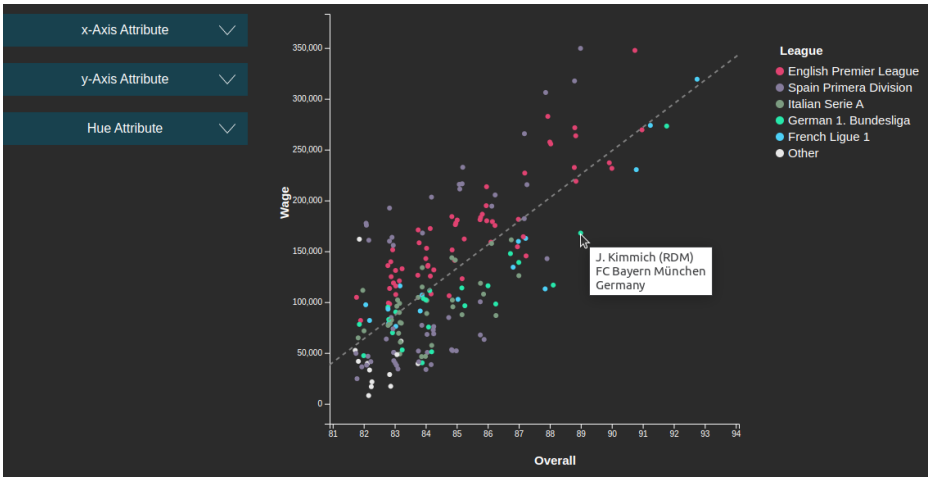


Figure 7. The general statistics scatter plot allows an analysis of general correlations between different attribute values. We can for example plot the wage against the overall attribute value. The information gained through the hue helps finding new insights (as e.g., that English and Spanish clubs overpay their players). When hovering over one of the points, we can get more information on the player.

We encountered a few challenges with this representation. One is similar to the challenge we faced in the player comparison, namely the different scales of the attributes in all attribute selections. We used the same solution as for the player comparison and designed dynamic axes with animated transitions (now both, x- and y-axis).

The second challenge we identified was that different hue attributes required different color schemes. While preferred foot is a Boolean attribute, the other attributes consist of multiple values. However, even attributes with multiple values have a difference in their value's nature. The Weak Foot and Skill Moves attributes have a linear change in values while Position and League have a nominal(categorical) nature.

While using a color palette that follows the flow of the entire dashboard, we used different solutions fitted to the different data categories. For nominal attributes, we selected colors which are harmonious, while in the case of the linear attributes, we set the blue and pinkish colors from the player comparison section as maximum and minimum values of a linear scale.

The third challenge is that while players have values between 0 and 100 for most attributes, the values are still integers. This means that in many cases the circles indicating a single player would completely overlap and hide many players with the same attribute combination for the x and y attribute. We solved this problem by adding a small amount of random noise to each point in both directions. The change is small enough that it does not alter the expressiveness of the graph and fixes the problem we encountered earlier. Using a heatmap-like plot would have been another possible solution to this problem but this would not have allowed to differ between single players.

A final, but nevertheless very important feature of the graph is the dashed line indicating the general relationship between the attributes on the axes. It was calculated using a linear regression. We are aware that a linear regression should normally be displayed along with an uncertainty band indicating the uncertainty of the predicted slope and axis values. We have implemented this behavior, but decided not to use it because for most attribute combinations, the uncertainty interval spans the entire graph and provides no additional information to users. However, we are aware that linear regression without any sort of uncertainty estimate can be misleading in some cases, which is why we used a dashed line and subtle gray for the linear regression output. Nevertheless, we believe this is a useful tool for identifying general trends.

4. Conclusion and Future Work

The developed dashboard enables users to explore various aspects of the FIFA players dataset. It is designed in way that both FIFA gamers and real-world football fans can use it to make data-driven decisions. Football fans may use this dashboard to learn more about the players, leagues and trends in attributes for their favorite players.

The dashboard is even more important for FIFA Career Mode users. The users of our dashboard can find players who can strengthen the team in certain positions, compare them to potential alternatives in similar positions and find the best compromise between the key attributes of the players and their market value.

Considering the extensive dataset available, we can think of a few ways to extend this dashboard and provide users with even more information and visualizations. To give more prominence to the spatial information available, one possible

extension would be a world map visualization showing summary statistics for different nationalities or leagues from different countries. This could give us a better sense of the different types of players in different countries. Such a visualization could, for example, help debunk the myth that South American and especially Brazilian players are more skilled in their dribbling.