Visualization Playground for EURO 2020

By

Mahmoud Said

Max stieber

Simon HAMON

COM 480 Data visualization

Context

Problematic

Can we combine video game data with real events to have a new point of view ? Will such a connection convince gamers and sports fans to have a statistical approach ?

Motivation

The EURO competition is considered by many the second biggest football competition (the most popular sport) in the world. Fans passionately follow the competition every 4 years, but UEFA EURO 2020 has been delayed to June 11th, 2021, due to the pandemic. The fans normally tend to discuss matches with friends beforehand, and some place bets on certain teams(friendly or with booking agencies). For these reasons, we are trying to create a visualization playground that people can use to compare two teams head to head before their matches, to see how they stack up against each other.

- Offer general comparisons between teams (offense/defense)
- Being able to compare team players with each other
- Show stats that correlate with some simple strategies to expose teams' weaknesses and strengths
- The visualization can be easily adapted to future football sporting events like the World Cup or the Champions League

Dataset

The dataset we rely on is the FIFA 2021 football game. The dataset provides a large amount of information on every professional player. Each player is rated on every aspect of its game. The dataset is under a CC0 Public Domain, which allows us to use it without copyright issues.

Fifa dataset is widely used and accessible on Kaggle, thus the data is of great quality. It normally does not require much preprocessing nor cleaning.

Additional information

However to answer our problem we have to enter European teams that are not directly available in the dataset. We have to combine the main dataset with data from a publicly accessible Website. We scrape this website to find the different team compositions. We have updated the teams to the official team selections. In order to merge both datasets, we need to use fuzzywuzzy, a python library that is able to compare two strings and match them when they correspond.

Exploratory data Analysis

As said in the previous section, the dataset has a good quality. After creating the Euro teams we can explore the data.

Teams Composition

The first interestings statistics is to see how many players are competing to be selected for the Euro. Indeed we started in April, the final composition for national teams hadn't been released at the time. Therefore teams can have more than the 23 official players.

Here is a chart that shows us how many players are in each team:
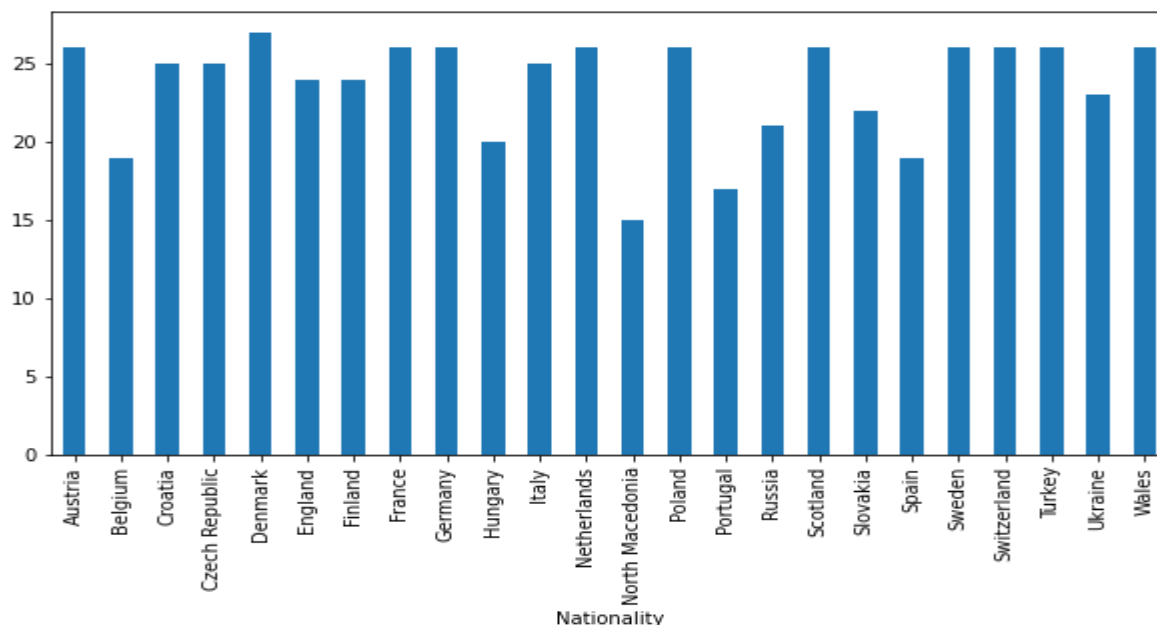
Unfortunately there are some players that were not available in the Fifa Dataset, among them many players from the Northern Macedonian national team. We can not really do anything to address this issue, except for the players to be discovered by major teams during this European Cup. That way, they will be available in the next Fifa dataset. The missing players in the Portugese and the Spanish team are explained by their different names in the two datasets. In one dataset they use the short form of their name in contrast to the other dataset. This leads to the fact that the threshold we set for fuzzy matching is not met.

Then thanks to Fifa players ratings we can calculate the average rating for each team. Here is the chart of teams average ratings:
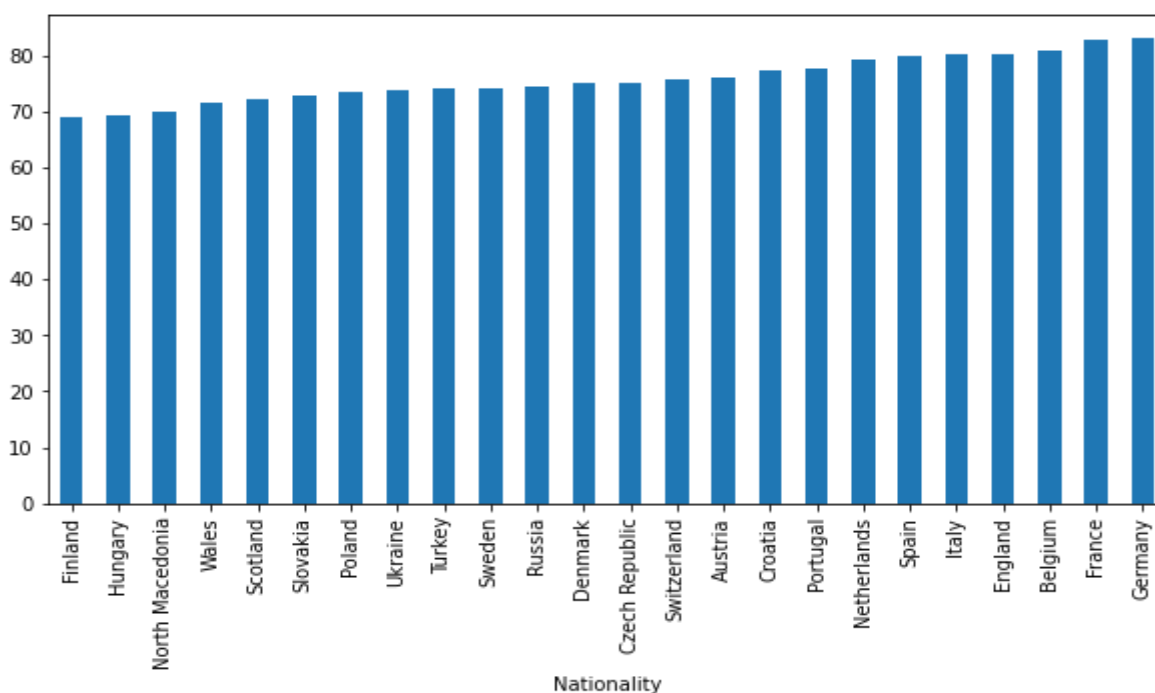
This statistics is trustable thanks to the work put in by EA games to have the most precise grade for each player in each aspect of his game. The ratings of players are conducted by hundreds of scouts. Their ratings translate into a rating for many categories between 0 and 99. Fifa ratings also take in account the level where a player evolves, as the competition increases.
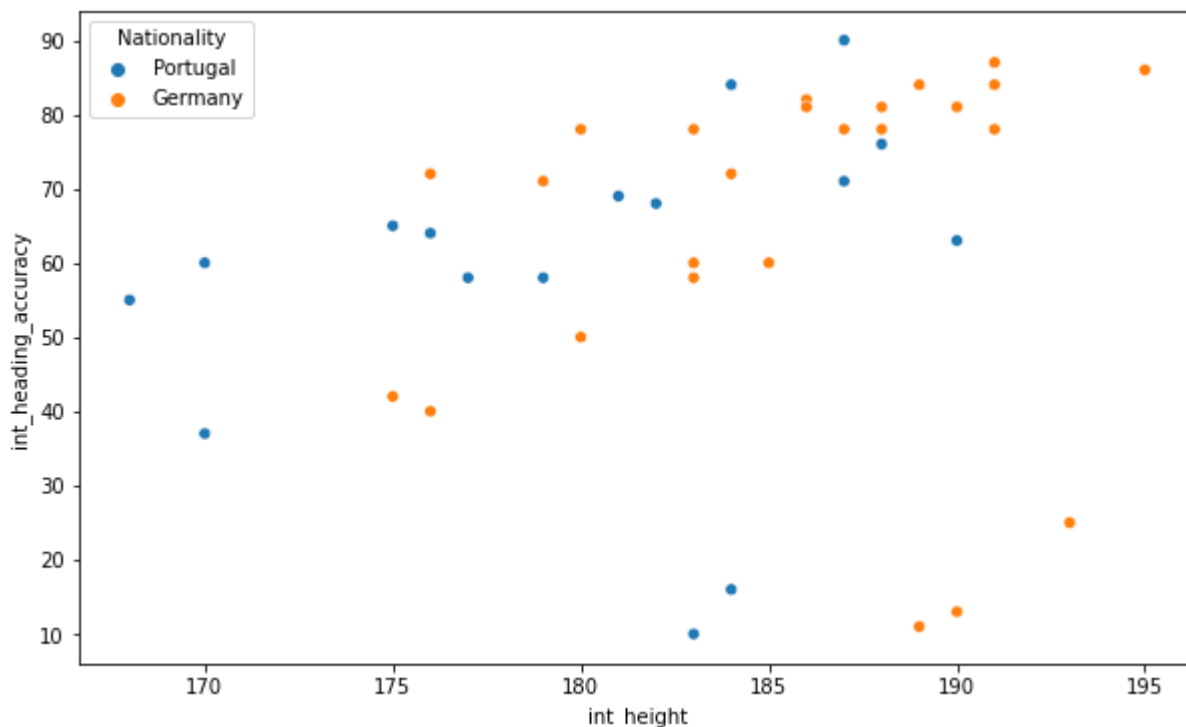
Comparisons that we want to show have an objective to compare team abilities in certain aspects. For example here is a comparison between Portugal and Germany, in terms of heading abilities of players:

We decide to represent heading rates in function of player height as it is an important consideration. Those charts' purpose is to represent how we can handle data to compare teams, and thanks to the Fifa dataset we can do comparisons in every aspect of the game, such as finishing of strikers, defense strength. We still have to consider that there are some parameters that we cannot control or implement, such as team configuration on field, coach's orders or players' confidence.

Design

The design should be both beautiful and functional and often it is a trade-off between the two attempts. By carefully planning the project, we try to avoid making too many compromises between the two. Therefore we have crafted graph sketches that will be shown in the following.

Sketches

Basic comparison

The competition teams are divided in groups for the pool phase. As we want to compare 2 teams we want to provide an easy selection, such as a drag and drop of national teams, based on their flags :
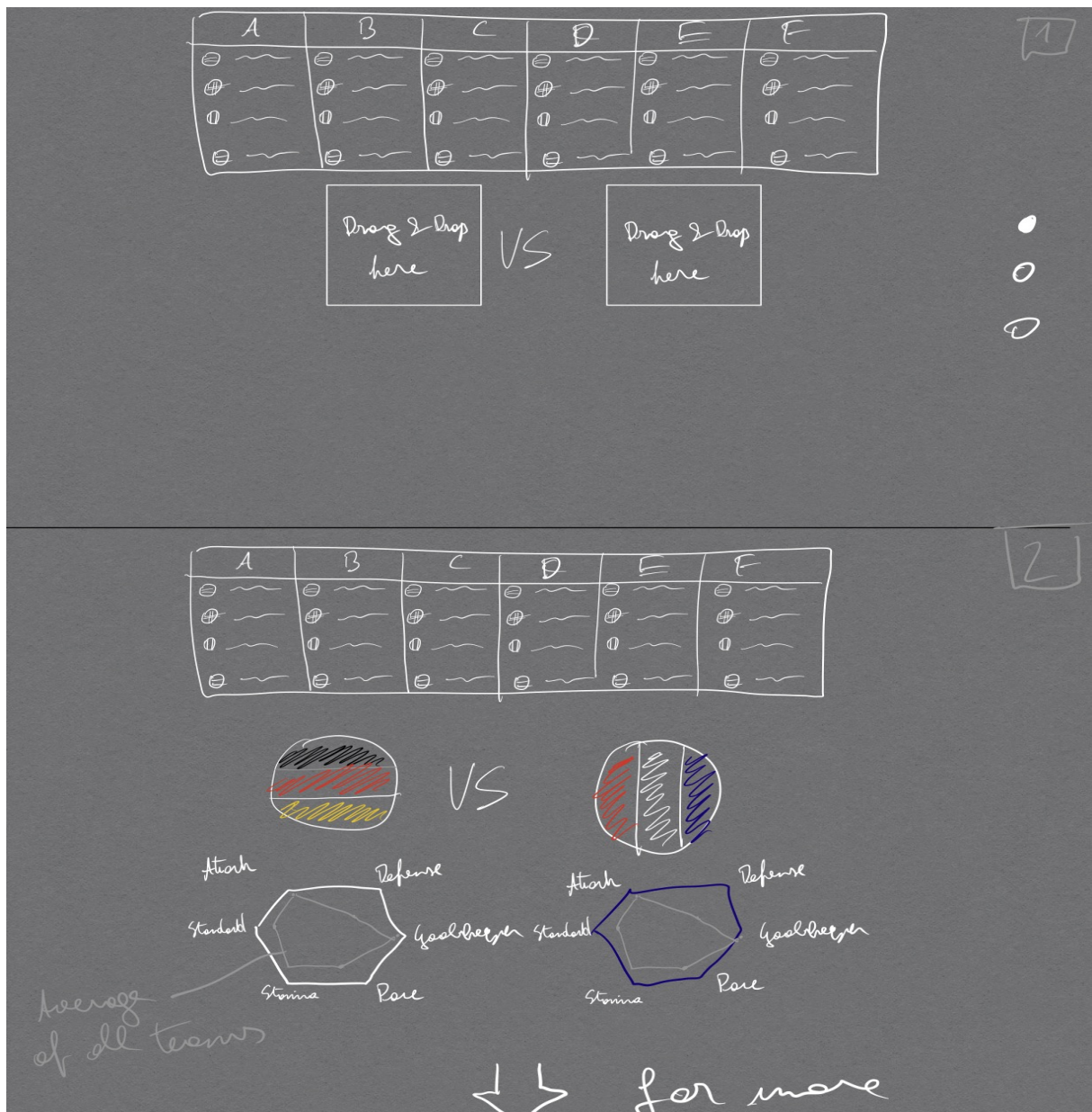
When 2 teams are selected basic statistics are provided. We select radar graphs, with a background part in gray of the average statistics, thanks to their intuitive representation.

The previous part allows anyone to understand how it works. We will add guidelines to make it even clearer.

Then users are invited to scroll down to access advanced statistics, centered on players. Yet we don't know if we want to forbid scrolling before selecting teams, because statistics will be empty, so there is no point in terms of comparison, but meanwhile it could help users to understand the process of selection.

Advanced statistics

Advanced statistics are about the team's players. Our goal is to make a dynamic graph, with filtering selection made by the user, that represents the player on different statistics. We are highly inspired by a graph available in first diapositives :
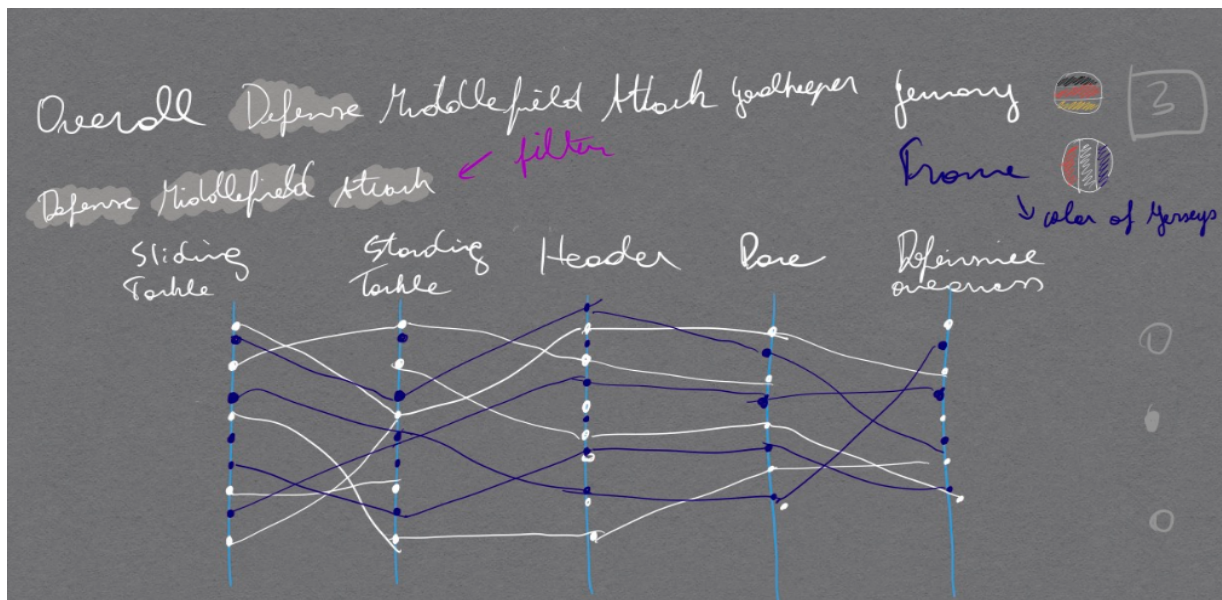
As represented with a light gray background, users can select their filters. They will be able to compare only attacking or defending players or they can just select a few detached players. Moreover user will be able to toggle between different types of attributes e.g. attributes we link to defensive qualities like standing tackles or anticipation.

A radar chart is also a fantastic tool to compare different attributes. It is well-suited if the range of the different attributes are the same. Fortunately, this is the case for the Fifa Data that is scaled between 1 and 100. We will therefore use radar charts to quickly compare different abilities of players.

The last will be about strength and weaknesses comparison. We aim to detect some statistical aspect of the game that can be judged as a weakness or strengths. We are mainly restricted to use team compositions because systems vary a lot and are not known in advance, as well as for titularized players. To counter that we can still show better players in each position.

Challenges

Project stakes

Throughout our project implementation the main concern we had was to create the most intuitive and informative data visualization.
The hardest part of data visualization is to be concise and to give as much information as needed but not too much otherwise we flood our users. As seen in class there are multiple ways to strewn information on a website. The technique we use the most is interactivity. We aim to give a user experience that does not require to give many explanations about how to use the tool. The user should be able to flow through the website.

Domain Specific Visualization

Some of our visualizations are domain specific to football (e.g pitch and comparison bars for strategies) meaning they do not come out of the box from D3 or other visualization frameworks. That meant we had to manually draw them using other components.

In addition, we construct complex visualizations that summarize information about the sport which added some overhead work this included:
1. Searching for heatmaps for different strategies to make the highlight areas.

2. Searching for attributes that correspond to the complex values displayed for each strategy by looking into multiple sources for football analysis.

Design Changes

Final Implementation

Our final implementation on our website begins with a visualization of every group for competition and areas for dragging teams are displayed to invite the user to drag two teams. If no team is selected no chart will be displayed. The drag and drop can be changed as many times as the user wishes.

Best Players

A first informative table pops up after dragging the team inside of the dropping area. If the user wishes to change the team at this stage, we can still do so.

Parallel Coordinates Chart

We wanted to create a chart that allows us to compare different sets of skills of teams. We therefore decided to implement a parallel coordinates chart which ticked all the boxes for our requirements. Moreover we wanted to make it interactive such that the user could filter out different positions because the user might want to evaluate a certain type of quality that is specific for a position. We were able to implement all of these features by getting some inspiration from the Popular Blocks - bl.ocks.org webpage. The chart is able to filter every parameter at the same time and explore the dataset in an intuitive manner. Moreover, the user can also compare a unique selection of players of the opposing teams. We give some filters that allow us to only look at players playing in the attack, middlefield, defense or the goal.

Radar Charts

The main aspect for radar charts is the comparison between two teams and therefore comparison between team and the average of all teams. Beside giving a color to every country we wanted radar charts to be interactive so users can play with them and select the players the user wants to compare.

On the bottom of the page there are buttons to add more labels to the current radar chart, it gives the user the possibility to model the graph the way he wants.

The scale of the chart adapts to the value printed, this allows us to visualize several categories in the meantime.

When a mouse is hovered above a point the exact value is given to the user, as complementary information.

Strengths and weaknesses comparison

The strengths and weaknesses comparison consists of comparing three different complex attributes (each is a combination of simple attributes in the dataset) by aggregating the values for team players of relevant position for each of the complex attributes. The chart itself is manually designed by creating a D3 rectangle with rounded corners. The ratio of colors depends on the values of both

teams for the corresponding value with a circle of the same color as the winning team to separate them.

The pitch is drawn manually using D3 rectangles, lines, markers, circles and arcs. The highlights for corresponding heatmap are created using the same way.

Peer Assessment

Max Stieber

- Has built the first page which has all the groups, the drag and drop functionality to choose the two teams to compare for the later stages
- Designed and implemented the chart for fine grained comparisons between all the players (Parallel coordinates)
- Designed the general outline for the website

Simon HAMON

- Has built the radar charts
- Designed and implemented the per team analysis pages
- Designed and implemented the second page of the website with the comparison radar chart

Mahmoud Said

- Design and implement the comparison of teams on different strategies by drawing a chart using svg which does not use any pre implemented charts
- Draw the canvas for the football pitch and the highlighting on it for different strategies
- Handle the background functionality of the website such as data loading, population and transitions