

PROCESS BOOK

Alexandre Reynaud - 258402
Bouraux Léopold - 257368
Rinaldi Vincent - 239759

INTERNATIONAL FOOTBALL DATA

A brief chronological history and data visualisations of international football from its beginnings to the present days.



Uruguay 1930



France 2020

INTRODUCTION

Our Project aims to present an historical and statistical view of every football match between nations from 1872, year of the very first official international match, to 2020, just before the COVID-19 pandemic stopped every activity around this sport.

The main motivations for the realisation of this website was to make available for the community a complete database about nations of the most popular sport in the world. Many football enthusiasts are often looking for football statistics for plenty of reasons : learn the history of the football team of their country of origin, retrieve data for a newspaper article, future interviews, or live presentation on TV to entertain and captivate the public on live TV studios, or even guess the future winner of a given football matches and beat bookmakers in the sport betting domain.

We found out that it is really difficult to find such a complete database tracing every football match since the creation of this sport, we thought that it was a great opportunity to finally make available such a thing at a single place on the web, and that is not importantly spread between dozens of websites that even have incomplete or inaccurate information.

The realisation of this project has been made possible thanks to the complete football matches dataset between nations we found on *Kaggle*, that has been made available by *Mart Jürisoo*, and containing more than 40,000 football matches with their date, facing teams, final score, the place were they have been organised, the related competition, and if the match was played in a neutral stadium or in one of the two teams. We preprocessed the dataset to keep only countries recognized by the FIFA, and still keeps more than 35,000 fixtures. This dataset can be accessed at the following link <https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>.

Our website is designed in 3 parts. Nous allons d'abord présenter l'History Tab, then we will discuss on how we made a visualization through a map and finally we will explain how we designed our last tab called details.

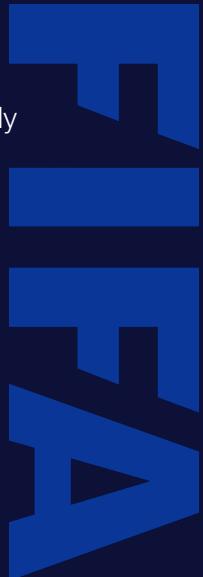
HISTORY TAB



Our initial idea at the beginning of the development of this website was to target two general types of users: those fundamentally interested in football statistics and looking for advanced visualizations, such as a sports journalist, or curious users, interested in football in general. The resulting idea in the design was to provide an original homepage to complement data visualizations. Since football is a fairly old sport and is of interest to all generations, we wanted to present the history of international football in a narrative way.

On the one hand, to teach young people how football became increasingly democratic and then imposed itself as a major sport, and on the other hand, to bring back memories of past moments. However, in order not to bother our regular users, a shortcut to the visualisations is present on the home image of the site and avoids having to go through the whole story before viewing the data.

To achieve our goals, we wanted this tab to be user-friendly, sober, and the way of reading it to be similar to a linear reading of history. So we chose to display the information by vertical scrolling, as if we were crossing the thread of chronological history. We have intentionally separated the matches from the first international in 1872 to the last at the beginning of 2020, by symbolic periods, displaying only one at a time to make the reading more enjoyable.



For the design of this part, we were inspired (of course in a simpler way) by the visualizations used by Apple to present their products and their stories. The ScrollMagic advanced layout library helped us a lot in this way although its use was complicated to understand at times.

As mentioned above, we wanted to come up with a brief summary of the history of international football, so we have focused on the key events, matches and famous players in history, in particular those which happened in FIFA World Cups. Information on this subject is available on a multitude of websites, but we wanted to centralise it here. More over, in order to make this story more attractive and to make the evolution of football understandable, we have added pictures whose sources are referenced in an Image Source tab in the bottom of the page. This footer also contains a link to our GitHub Repository as well as a link to a page which briefly describe ourselves. We also added home-made statistics to the history which are highlighted in texts

But we always need more statistics !!! We then introduced a timeline, showing general statistics specific to each period described. This first comparison of the data allows us to highlight the differences and the evolution of football through these periods. This drastic evolution over time can also be visualized through the two gifs images created from our map view and the <https://gimaker.me> website. The one here displays the number of goals scored by each country from 1872 in intervals of 5 years.



We also faced the problem of loading each image that was lagging the tab. To counter this problem we compressed the images as much as possible, trying to lose as little detail and color as possible.

MAP TAB

While thinking about which visualizations to use, we realised that, for most options, a crucial information was lost, the geographical position of the countries, so we designed the Map Tab.

The original idea was to display a choropleth map in order to compare statistics up to a global level without losing crucial positional information, maximizing the data density in a simple yet very effective manner. When building the corresponding tab, the Map Tab, we first used the stock image of a map as background in order to experiment with the layout, then realized, what if the user has very limited information about the countries they wish to compare ?

As a response to this preoccupation, we decided to implement various ways to identify the countries of interest:

- A Flag Slider, was put at the bottom-centre of the graph, displayed the flags of all land masses whose team is recognised by the FIFA, and was supposed to allow the user to select the desired country by clicking on its flag.
- A Flag Search Bar, built in conjunction with the Flag Slider, had for purpose to allow a user that knows no more than the name of the country to find it anyways by filtering the flags displayed by the slider according to the written input. We decided to locate it bottom-right of the tab, right on top of the “Flag Slider”.

The first version of “Flag Slider” was made of dense circular buttons and the sliding was done through other squared buttons that needed to be clicked. This original, cloggy version of the “Flag Slider” was soon replaced by an instance with smoother design and scroll-sensitivity:



Now that countries were simple to identify, we needed to introduce criterions that would determine which statistic to display and how to filter the dataset, we came up with the following features: The first one, the “Measure Box”, which strictly contained radio buttons and was intended to display a variety of measures that quantify performance, to use as the statistic from which comparisons are drawn. It was placed top-left of the tab.

The second, the “Competition Box” which was supposed to serve as a filter that would allow to choose which competitions to consider while computing the statistics directed by the “Measure Box”. It was positioned right below the “Measure Box”

The third and last, the double-ended flag slider, that was designed as a second filtration channel that would allow to constrain the time period of study. It was located right on top of the Flag Slider.

At this point in the implementation, we had a prototype of the layout of our Map Tab, it could be interacted with but the logic had yet to be implemented.

Here below, you can see the appearance of the “Map Tab” as it was published for Milestone 2.



Now came the real challenge, we had to implement the actual functionalities, possibly add new ones, revamp the whole aesthetic of the page and add beautiful yet light visuals and animations.

We first started by replacing the stand-in map by a leaflet frame, added a legend, a real-time display which prints the value assigned to the hovered country if any and chose a color scale considerate of colorblind users and ideal for sequential information:



The true difficulty was in finding a geojson dataset that fit our needs, it had to include all lands recognized by the FIFA, meaning countries but also sometimes states and islands. Since finding such a dataset was unrealistic, we decided to find datasets which separated land at different scales (countries, states, islands) and merged them in a way that would cover all needed entities at the optimal scale (Scotland, Ireland, North Ireland, Wales and England instead of the United Kingdom; United States instead of California, Georgia, etc). Finally, we arrived to an almost complete map. Only 4 major countries are not represented: Czechoslovakia, German Democratic Republic, Saarland, and Yugoslavia. These are former countries that we could not display because of possible time intervals where these countries would have disappeared between the beginning and the end of the range.

Then, for the criterions, we simply implemented their logic so they would interact in the desired manner with our dataset at the moment to compute and assign statistics to the map but also added some restrictions on their use to prevent abuses:

- For the “Measure Box”, we made it so it would disable the “Friendly” checkbox from “Competition Box” when selecting a measure that only regards official tournaments. Similarly, we disable all checkboxes but the “Friendly” one when only unofficial matches qualify.

- For the “Competition Box”, we included an “All” option which disables all other checkboxes when used to prevent senseless redundancy.



Regarding the alternative search mediums, we implemented most of the logic without many issues, resulting in a “Flag Slider” perfectly in sync with our “Choropleth Map”,

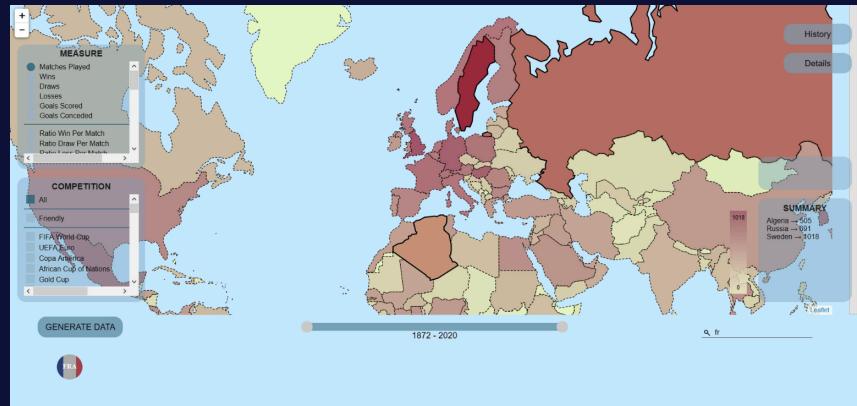
Implementing this wasn't done without trouble as the leaflet's interface requires to implement advanced modules called "Controls" to access internal information such as individual country tiles, the only exception being at initialization time. We dealt with this complication by assigning listeners when initializing the map; listeners on the map tiles to make the corresponding flags reactive when using them and listeners on the flags, to make them able to affect the corresponding tiles when interacted with.

Some additional problems also arose and new components were added to solve or mitigate them: We realised that computations were too time consuming for the map to adapt in a fully dynamic fashion, we therefore decided to create a new component, the "Generate Data Button" which considers the criterions selected at the moment and updates the map accordingly.

Our visualization was now able to compare countries at a global scale without losing sight of their positional information. Nevertheless, it wasn't able to compare countries in a less whole englobing manner and we thought this was a matter worth addressing. In order to do away with said matter, we added a display box middle-right of the tab and the option to select multiple countries and used these to support the following

behavior: when clicking on a country, in addition to being highlighted, its name and current value are saved in the display box, resulting in a selection of countries for which position name and value are all neatly put in evidence:

Now that functionality was addressed, the last issue to deal with was style quality and coherence, so a revamp was in order:



Given our color scale is reddish and the land takes a good chunk of both our map and our tab, we decided to offset this with a soft celestial blue background which ended-up being both aesthetically pleasing and appropriate given this background represents the sea.

Also, we added transparency to most of our components (ex: "Measure Box") in order to limit interference with the visuals of the map, results are visually pleasing and usability isn't decreased.

DETAILS TAB

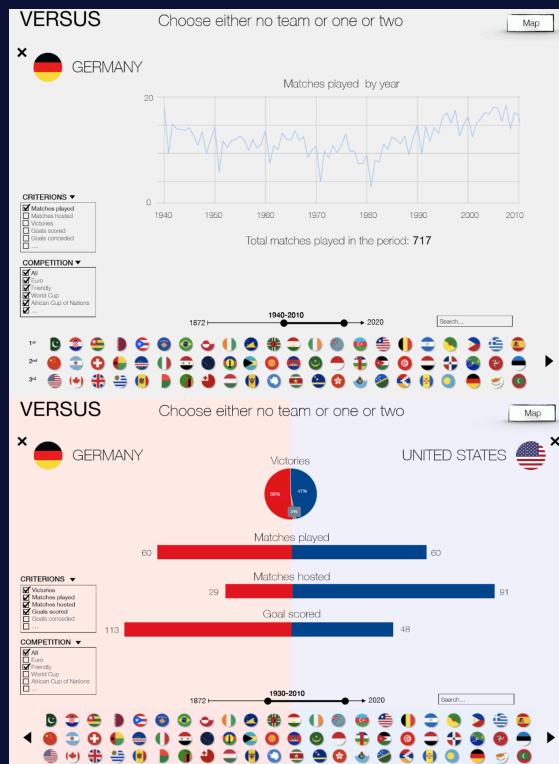
We didn't want to stop here and wanted to bring another type of visualization to our website. The map is a pretty attractive and very interactive way to obtain a nice overview of world-wide statistics, but it has a weakness : this is not the best way to visualize detailed information on a single entity, like a country. To make our web site more complete, we then wanted to add a tab which would make us be able to visualize statistics about football national teams via a more compact and simplified way to fit all kinds of user preferences. This led to the creation of the last tab which is the Details tab.

We wanted to keep things simple for the user, with an easy handling of the different ways to display detailed information about a country, and visualize it on the tab, but also produce an implementation that is sufficiently complex to bring our personal touch and greatly distinguish our web site from the others about football statistics at the same time, since we are trying to produce a database so complete, at a point that it is really tough to find a publicly available one on the web, that gathered so much information about football nations, and especially on the largest possible time period since we are presenting statistics since the very first official football match of the history of football.

We wanted to divide our Details tab in three different part : statistics over time between every countries in the world following a chosen criterion ; the capture of information about a single country performance where it is obviously possible for the user to select the country he wants, but also the

different criterions, among those we would make available, he would possibly like to retrieve and display numerical data about ; and a VS mode to make the user able to show statistics of not one but two different countries to obtain comparative statistics and extend even more the capacity of our website to fit any sport fan or journalists demand if they need historical statistical information about two nations before an important match for example.

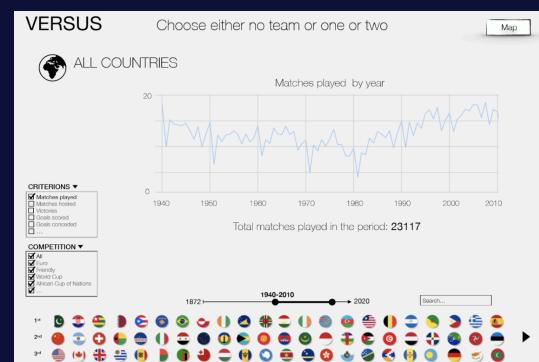
At the very beginning we thought about displaying a line graph for the world-wide statistics part, and the single country visualization. For the two-countries comparative part, we though, as an initial idea, to display colored bar plots going in two different directions, left and right, since it's greatly used in the world of sport to compare team and athletes statistics given a particular competition, and maybe other kind of visualizations that we can also sometimes find in the football domain, like a pie chart.



statistical comparison feature, since that, according to the course, and common data visualization practices, pie charts must be avoided since it's difficult to represent information and display them in an effective way for great audience comprehension. Also, concerning the world-wide statistics feature, we realized that a graph would not really be the best option to represent every country statistics at once since that we have a too huge amount of them, and the visualization would be unclear for most of the users with a line graph.

Finally, for the single country information feature, we thought that displaying only one measure at a time would be unfortunate, since that we should have more space to display greater advanced information for enhanced visibility when focusing only on a single data element.

We then decided to display an horizontal bar plot for every feature in the Details tab, since it seemed to be the best choice for this kind of visualization, in terms of efficiency but especially clarity and simplicity for the viewers. The plot would be displayed in descending value order from top to bottom in order to highlight the most valuable country for a chosen criterion in the world-wide feature, or the criterion that has the highest value for the selected country in the single country statistics feature, always according to the chosen time period. We also wanted to add the possibility for the user to select specifically which criterion to display in the single country feature so that he just can focus on the data he is really interested in, but also because it would bring us a more advanced website in terms of functionalities.



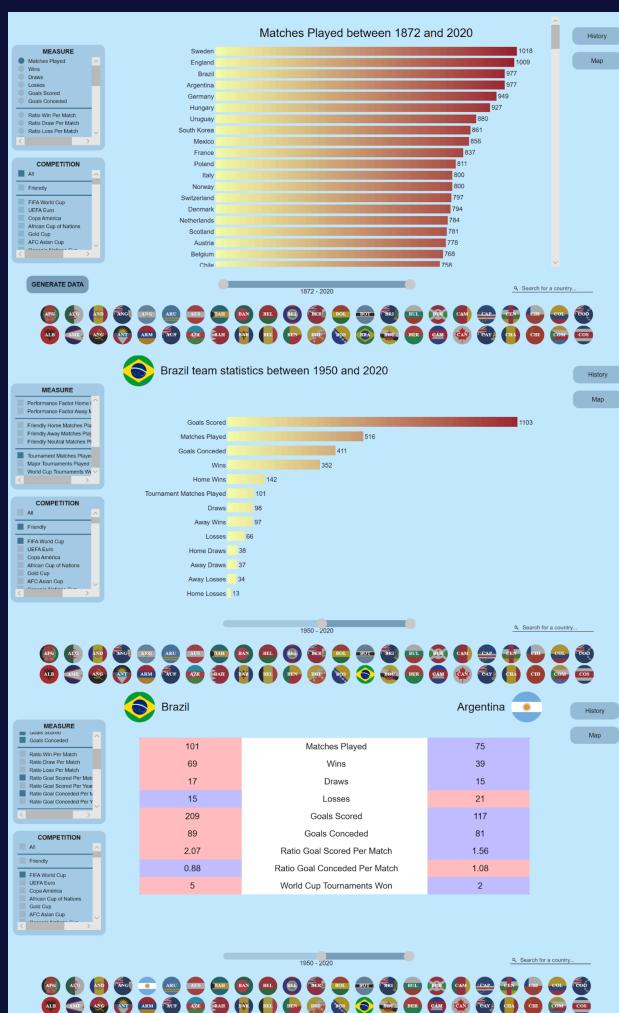
At the bottom side of the page, we would have, like in the Map tab, a flag slider, make us able to select the different countries, among 215 different ones, we want information about, even those that don't exist anymore, like Yugoslavia or German Democratic Republic. A Search bar would be also available to find more easily the country we want to analyse. To keep bringing the user in a total time travel, a time slider would be available like in the Map tab. The two criterions menus, one being the different measures to display, and the other one being a list gathering all competition and acting as a filter, would be available and displayed at the left side of the screen.

With all those ideas gathered, we could produce the sketches representing the appearance of our Details tab at first sight. Here is how we thought to represent this page for our website at the beginning, with the three different kind of visualizations : world-wide statistics for one criterion, single country information, and the two-countries comparison feature.

However, when getting into work and starting to program the first elements of the website structure, we decided to reconsider some specificities of our initial choices. First of all, we all agreed to definitely get rid of the pie chart of the

A new decision has been made afterwards for the world-wide feature : the display of several different measures at the same would make the visualization too heavy in terms of displayed information, and would be difficult for the user to visualize and understand easily the full plot. We then decided for this part to force the user to only select one criterion at a time to display the data. Also, since we also wanted to dynamically recalculate immediately the bar plots at each single user interaction with the criterions menus, the time slider, or the flag slider, without the user having to press a button to replot the graph with the new parameters, we realised that generating the data at a world-wide scale was not performed immediately by the code, but needed to take a few seconds before displaying on the web page. Then for the world-wide feature, the user, after choosing its parameters, would have to press on a button with the label Generate data, so its experience is not harmed by huge and slow calculations that would freeze its webpage at inconvenient times. However, this specificity is kept in the two other features, so the user simply has to move the slider, or select a specific number of criterions, and the statistics update immediately in live time.

To build those graphs, the D3.js library has been used to create SVG canvas. We didn't want to simply plot bars with a single color and without animation. Also the graphs had to be displayed in an horizontal manner. Some advanced handling was then required with the library in order to animate the bar displaying from their base axis (y axis for an horizontal bar plot) and growing to the right until they reach their full width. Also, filling them with a gradient color to make a better comparison between the different results would provide a more advanced overall representation of the statistics and highlights better the countries having a value hugely different for a given criterion, so we can make clear distinctions.



Unfortunately we faced programmatic issues with those graphs when trying to generate a two-sided animated horizontal bar plot for the two-countries comparison feature. The graph seemed to difficulty fit on the screen due to some issue in managing the geometry of this complex figure, and we had huge difficulties to manage the position of the base axis of the two sides of the plot, and the labels in the middle. Due to a lack of time to fix those issues, we chose an alternative for the two-teams comparison feature, and we decided to implement a table that also updates dynamically in live which present the selected statistics for the two chosen countries, where each cells representing a numerical value is filled with a color depending on the fact if it is lower, higher, or equal to the value for the other country for the same measure.

The user can, at any time, switch back to the Map tab or History tab whenever he wants with the buttons available in the top right corner of the screen, as in the Map tab.

The three images represent the final obtained results for the Details tab, with the first image being the world-wide feature, the second being the single country statistics feature, and the third one being the two-countries comparison feature. The time slider is at the bottom of the screen over the flag slider and indicates the chosen time period. The flag slider highlights the countries that have been selected by the user to display data.

We can go even further and extend the Details tab in future work by adding the possibility to compare more than 2 teams at once, or improve the time slider capacity by not only choosing between years, but also months so we can retrieve information on an even more restricted and precise time period.

PEER ASSESSMENT

• **Alexandre Reynaud** - Alexandre was in charge of implementing the Map tab and brought an important contribution in the realisation of the choropleth aspect.

He was always paying attention to his two other teammates' suggestions and contributed productively to the team discussion and work. He kept a huge flexibility when a teammate was in disagreement with his ideas, and was always prepared and on time before a meeting.

• **Léopold Bouraux** - Léopold had the task to implement the History tab, and has been able to include very advanced visual effects to transform the main page of the website into a complete immersive story of the history of football since its creation.

Like Alexandre, he was always paying attention to his two other teammates' suggestions and contributed productively to the team discussion and work. He kept a huge flexibility when a teammate was in disagreement with his ideas, and was always prepared and on time before a meeting.

• **Vincent Rinaldi** - Vincent worked on the Details tab and managed to implement an user-friendly interactive page where switching between the different available features is fast and very easy, while at the same time numerous different criterions are available at selection in order to display advanced historical statistics on any teams to fit most of the user statistical demand.

Like Alexandre and Léopold, he was always paying attention to his two other teammates' suggestions and contributed productively to the team discussion and work. He kept a huge flexibility when a teammate was in disagreement with his ideas, and was always prepared and on time before a meeting.