DATA VISUALIZATION (COM-480)

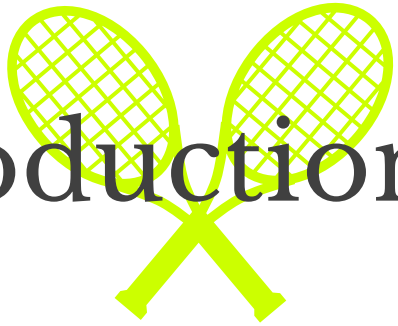
PLOT TWIST

Tennis data story

Process book

Félicie Giraud-Sauveur, Assia Ouanaya,
Alexandre Variengien

Introduction



Our Goal: A global picture of the tennis world, beyond superstars

Since its creation in the nineteenth century, **tennis** has kept increasing in popularity to become one of the **most** played sports, and one of the most mediatized. You cannot ignore the names of the best figures such as Rafael Nadal or Novak Djokovic that compete in the **unmissable** Grand Slam tournaments: from Roland Garros to Wellington. Nonetheless, the usual picture of tennis shared by the media tends to focus on a tiny subset of both the events and the players. With this visualization, we want to give a **holistic** image of the structure of the **community** of professional tennis players.

Data

We chose to work on a dataset of 36'120 matches from ATP tournaments, the most prestigious tennis tour for men players.
The database lists all ATP matches that have occurred from 2008 to 2022, and details the characteristics of each match such as location, tournament, court type, sets, etc. It also gives various information for each player such as weight, height, nationality, etc...

Peer assessment

- Assia made the two bar chart races and designed the website.
- Félicie made the graph visualization and helped with the design of the website.
- Alexandre made the tournament map visualization and helped with the design of the website.

Graph

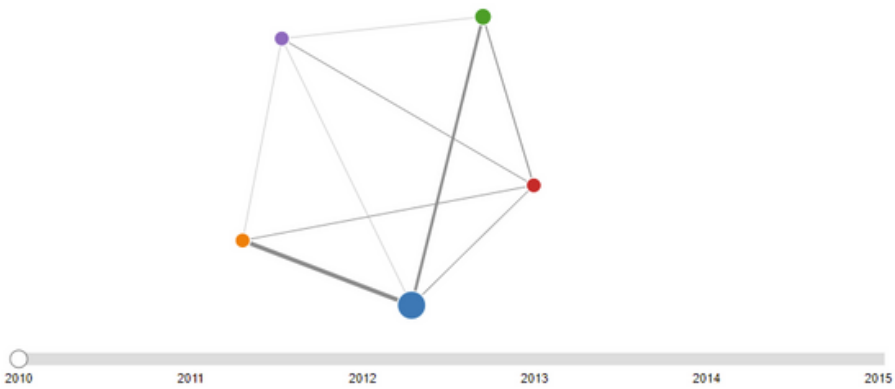
Interaction between players over time

Goal of the visualization

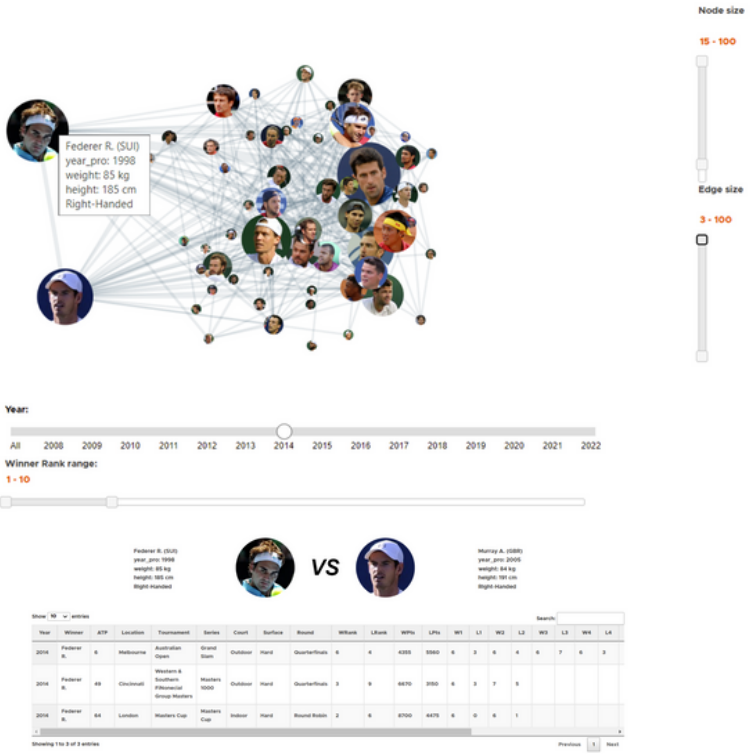
The purpose of this visualization is to analyse the interactions between players as well as the matches of a particular player. This analysis can be done according to the year and the rank of the winner. It focuses only on matches where both players have a rank below 40. It can answer questions like: Which players play the most? With whom does a player play the most? How does this change over time? What are the characteristics of the matches between two players? What are the characteristics of a given player?

Design evolution

The basic design of the visualization was to construct a graph linking players by year, with the size of the nodes and links proportional to the number of matches. Hovering the mouse over a node provided information about the given player. Due to the large amount of data, we then implemented a slider to select matches with the rank of the winner within a certain range. We also allowed to fix and defix the position of a node to be able to study node by node without them getting lost in the graph. These two elements have greatly facilitated the exploration of the graph. Then we wanted to improve the design. For this we replaced the coloured circles of the nodes by the portrait of each player. We also added the possibility to change the minimum and maximum size of nodes and links. Finally, we wanted to be able to access the information contained in the link between two players in more detail. We have therefore added the functionality where it is possible to click on a link to display a table with the list of matches and their characteristics (location, tournaments, type of course, sets, etc...) between the two players for the chosen year.



First design of the visualization



Final design of the visualization

Challenges & Design choices

Large data

The main challenge of this visualization was the handling of huge amounts of data. Initially, all matches were used to build the visualisation which resulted in a very slow loading time and a very confusing mass of data to visualise. We therefore made the choice to reduce the dataset considerably by selecting only matches with two players with a rank lower than 40. We considered that the choice of this rank already offered a very interesting exploration.

Interactivity

More generally, a challenge has been to add all the interactive elements to the visualization, the aim being to offer a complete and easy exploration.

Integration into the site

Each visualization was done separately and then integrated into the site. The graph was more complicated to integrate because it was not in one block but in several separate blocks. The use of Bootstrap was particularly useful here.

Key insights from the visualisation

Unsurprisingly, the players who have played the most ATP matches are the best-known players. We can also see how the community evolved through time: the force-based spatialization placed the old players on the left and the new on the right. And there is a lot more to discover by playing with the graph!

Tournament map

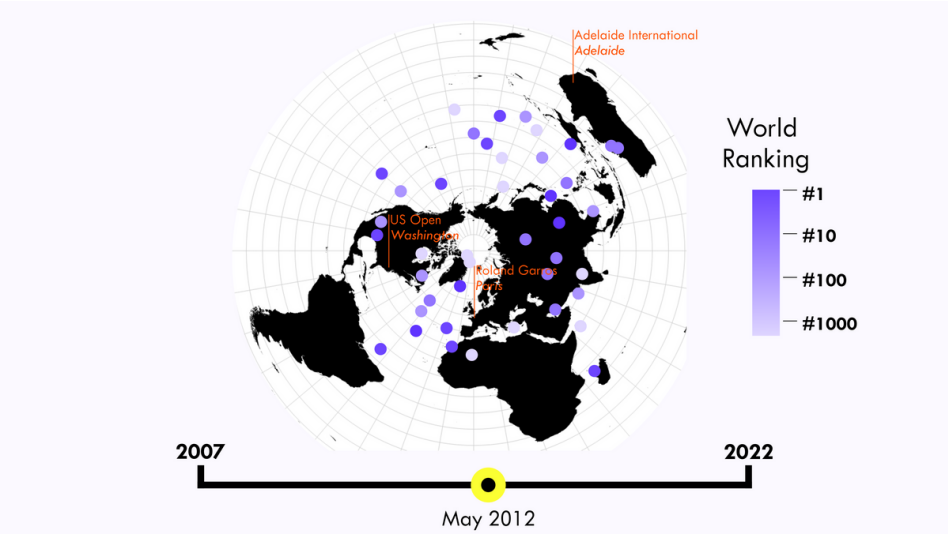
The Geographical Choreography of Tennis Players

Goal of the visualization

Many names of tournaments and places are associated with tennis: Wimbledon, Paris, Roland Garros, Melbourne... Nonetheless, their geographical position is not always clear.
We wanted to share what does a tennis player's path looks like on a map, what are the other tournament besides the big names, and what can we learn by observing the spatial dynamics of the community of tennis players?

First design

In the first maquette, each player corresponds to a circle, and its position is the city where he played matches. In between two matches, his location is interpolated on the Earth sphere. We also wanted to include names of tournaments as they unfold, a cursor to select a precise date, and visual information linked to the rank of the players.



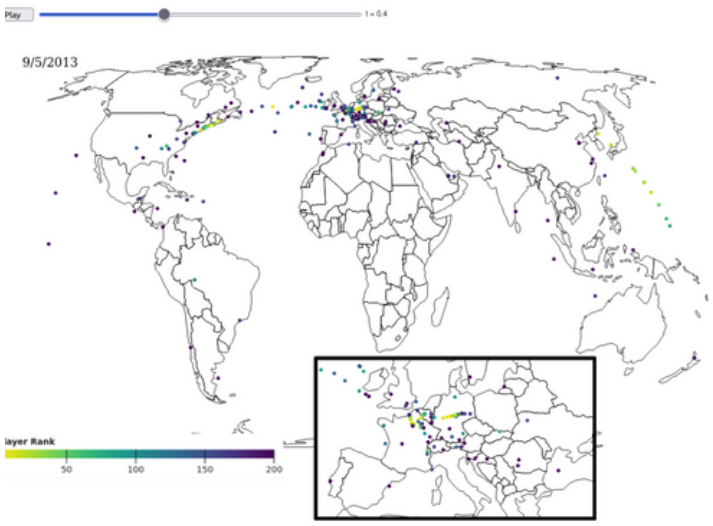
First maquette of the visualization

Challenges & Design choices

Data heterogeneity

The first challenge we faced while implementing the maquette is the heterogeneity of the location of the tournaments. They are concentrated in a few spots: Europe, south of Australia, and the US. My first solution to tackle this problem was to make a close-up of the Europe region to make more room for the tournaments happening here.

However, this required switching from looking at the world view to the close-up when a player enters Europe. This broke the continuity of movement that is the main interest of the visualization.



First solution: close-up on Europe

Minimalism by interactivity

We then decided to use a zoom behavior to be able to inspect regions of interest. Even if the global picture at the world scale is crowded, it is possible to follow the path by panning and zooming.
One of the technical challenges was that the zooming behavior cannot be implemented naturally, as the circles must keep the same size while the map has to grow.

Choice of the projection

In the beginning, We planned to use a polar azimuthal equidistant projection to obtain smooth movement by avoiding teleportation at the border of the map. We finally chose the Winkel tripel presented in the course. It is more familiar, and it did not distort the surfaces. This was one of the goals of the visualization: showing that developing countries that cover most of the world do not host any of the major tournaments.

Key insights from the visualisation

- The color scale enables easy identification of the group of the higher-ranked player. We can observe that they remain tightly together and move really fast as they play the most matches.
- We can identify the four tournaments of the Grand Slam (Australian Open, French Open, Wimbledon, and US Open), they gather the whole community of tennis players and are visible as the biggest circles.
- After March 2020, we observe that all the players begin to move slowly and disappear from the map, leaving it empty. This is the effect of Covid-19 restrictions worldwide that stopped tennis competition for several months.

Bar Chart Race

The ATP Ranking of Players and Countries

Goal of the visualization

Through this visualization, we aim to show the ranking of the best players and countries in the tennis field. In addition, we can also follow the evolution of the ranking per year of the **top 12 players** and **top 12 countries**.

First design

In the first maquette, each bar corresponds to a player or a country and its length indicates the number of matches won by this player or country. The number of wins is normalized by dividing by the number of matches played each year. The bar chart race does not include any timeline or way to stop or replay the bar chart race.

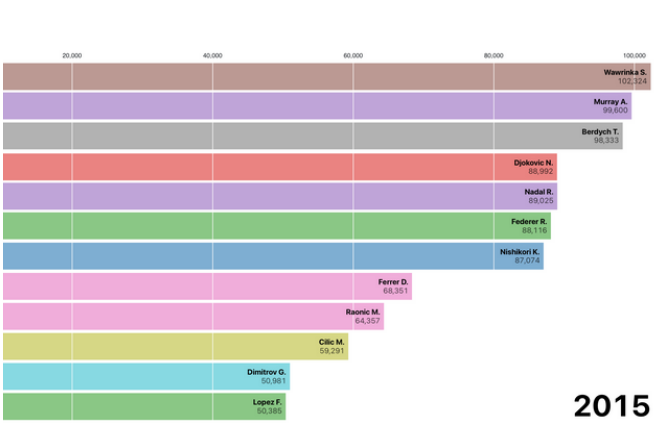
Challenges & Design choices

Data normalisation

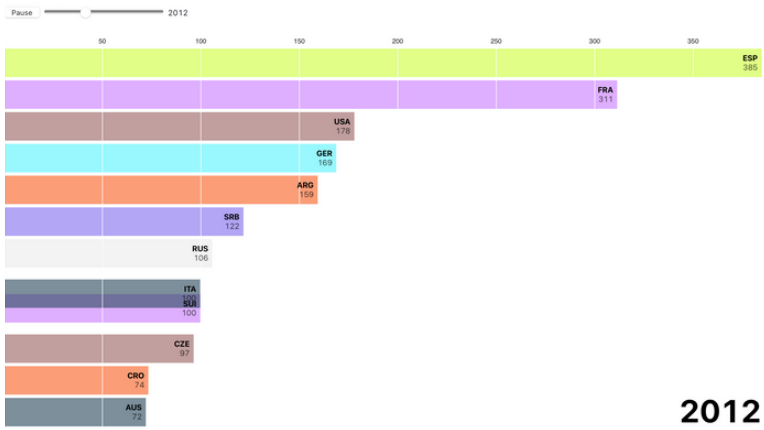
The first challenge was how to normalize the data set, we chose to normalize by dividing by the number of matches played each year

Slider

Since we had no interactivity nor control over the visualisation, we wished to add a slider to be able to choose the year of interest. We also added a pause feature to allow the user to stop whenever they want in the animation.



First maquette of the visualization



Final design

Key insights from the visualisation

- As expected, the big names of the tennis field such as Nadal, Federer and Djokovic are ranked high in the players bar chart race. But several other players appear in the bar chart race throughout the years and it is interesting to follow.
- We can identify Spain, France and USA as the highest winning countries in the tennis field!

Website

Goal

Through our website, we wish to create a simple and minimalistic canvas to show our visualization. We incorporated a color palette close to the tennis world by including vivid yellow for example or tan-orange. Small attentions have been given to details to make the website as lively as possible by incorporating gifs, fade-in effects and redirection links to our visualisations.

First design

The first design of the website was pretty simple with only the main page and tabs redirecting to our visualization, no special details were incorporated yet.

Challenges & Design choices

Incorporating the visualizations

Indeed, adding the different types of visualization and adapting them to the same web page template while keeping them web-responsive was a challenge. For this effect, we used bootstrap and our CSS knowledge from the course to help us. You can visit it here.