Game, Set & Match

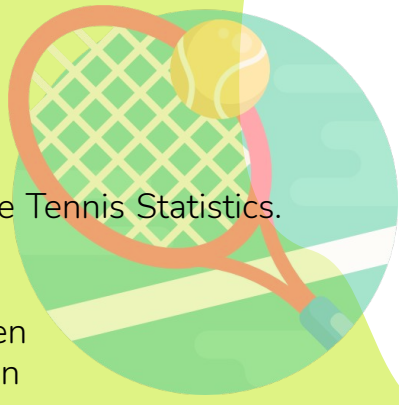
Marie Biolková
Marija Lazaroska
Ljupche Milosheski

Introduction

Tennis is one of the most popular individual sports world-wide. Professional tennis events receive a lot of attention, in particular the four major tournaments known as the Grand Slams. In 1968, Grand Slams opened their doors to amateurs to compete alongside professionals, marking the start of the so-called Open Era. Prize money was established for the tournaments, thereby enabling the players to turn tennis into a career. This dramatically increased the quality of tennis and, consequently, the interest in the sport. One of the reasons why many people enjoy following tennis events is the gripping matches between notorious rivals. The Open Era has seen many of these – arguably the greatest ones being Roger Federer vs Rafael Nadal, and Martina Navrátilová vs Chris Evert.

Goals

Our goal is to create interactive visualizations to highlight and **compare the individual performances** at Grand Slam tournaments and **identify frequent rivals** as well as **one-hit wonders**. This will allow the users to travel back in time and explore the careers of the best players in tennis history and their rivalries. The tool should satisfy anyone from the general public: **from sports newbies to tennis-savvy users**. A more practical use case would be **sport commentators** who would get **an intuitive tool** and **visualization to search for the players' statistics** instead of tedious way of searching through tables.



Dataset

We created our dataset by scraping data from Wikipedia and Ultimate Tennis Statistics. Our data consists of:

- The Grand Slam singles finals dataset, both for men and women
- Information about the players dataset, both for men and women
- Statistics on the final match dataset, only for men

Unfortunately we were not able to find a dataset or create one on our own for the statistics on the final match for the female players.

The obtained datasets including the information about the players have some missing data. For some of the players we managed to manually fill in the missing information and we left it empty for the rest, as this is not crucial for our visualization.

We did not end up using the match statistics dataset as initially planned.

Design process

Once we have agreed on the dataset and listed our goals, we brainstormed some ideas on how to achieve them. Having established all the details in the first meeting, we had a very clear idea about the form of the visualization from early on. The final implementation therefore practically coincides with our original plans (as seen from the sketches below).

Visualization

In order to show the individual successes, we implemented a **Sankey diagram** of winners and runner-ups, which will allow the user to see **the most successful players** in Grand Slams, since the width of the connections reflects the number of wins (or runner ups).

For the rivalries, we constructed a **network** of players who played against each other in a Grand Slam final. The edge width is proportional to the number of matches between a given pair. Consequently, the user will be able to see the **biggest rivals**, as well as **identify uncommon finalists**.

Finally, the above components are controlled by a **timeline**. Choosing a wide range allows users to see how the network and **rivalries evolved over time**, and they will be able to identify the most successful players in Grand Slam history in terms of number of wins or runner ups.

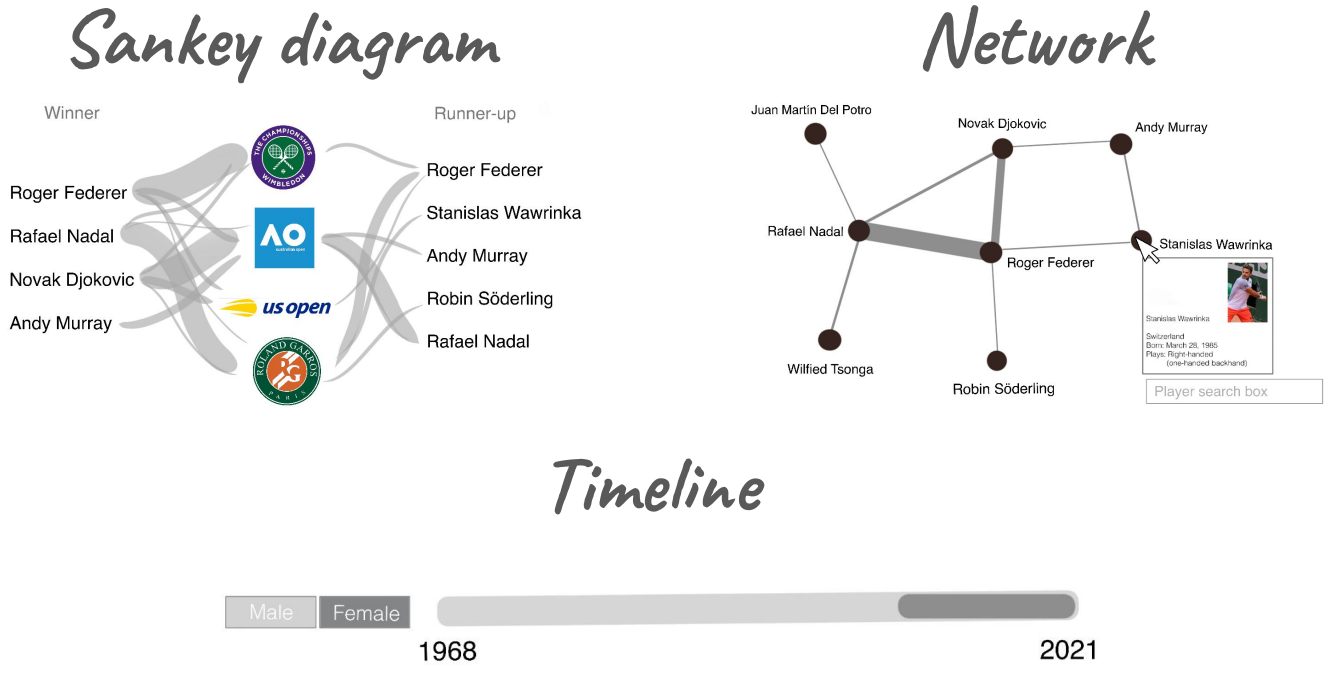


Figure 1. Sketches of main elements of our visualization.

The data input is controlled by the timeline slider as well as a gender switch button. Both the lower and upper bound on the year can be modified, allowing the display of different granularity.

Website layout

In our initial plan, the placement of the visualization elements was supposed to be one next to each other, as can be seen in Figure 1. Having the diagram and the network in the same viewpoint would have helped the user to better interact with both of the visualizations at the same. However when we created the diagram and network, their size was too big to place them next to each other. Making them smaller would have only made the visualization worse. Therefore we decided to place them one after another and fix the bar containing the slider, the female/male switch and the search bar. In this way the user is able to interact with both of the elements without needing to scroll down or up to find the filtering features.

Sankey Diagram

Inspiration

We wanted to show the dominance of the most successful players per Grand Slam. Inspired by the part about members' contributions in their biggest hits in the Beatles visualization, the Sankey diagram was the perfect choice as the width of the flow would be proportional to the number of wins/finals in the tournaments.

Challenges

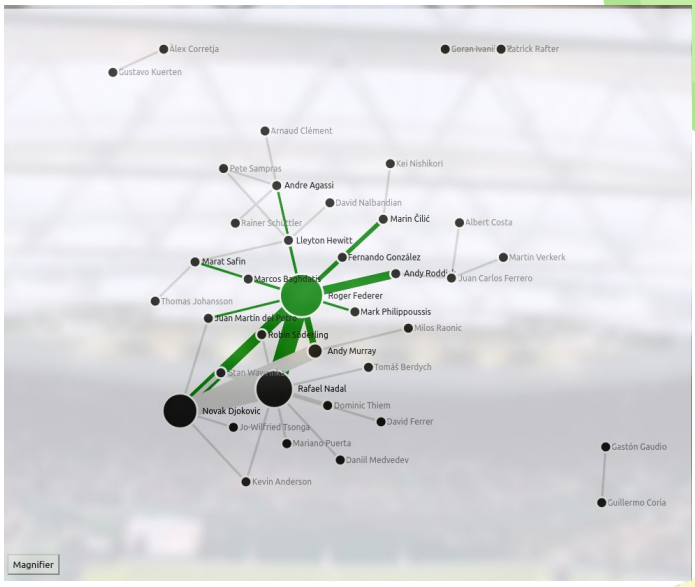
One of the main problems was dealing with how to represent the players when there were too many distinct ones in the selected range. There are more than 50 distinct winners for both, men and women, in the Open Era. Representing all of them is infeasible if we want to show each of them in a clear manner. To solve this issue, we went back and thought about the use case and the goal of our application. If the chosen range is small enough, then we do not have many players in the first place. Otherwise, the users would be interested only in the most successful players in the wide range, so we decided to group the less dominant into a single node. The players we choose to display is also motivated by our goal – we want to show the users the most successful players. So, for each Grand Slam in the selected period we compute the top 4 winners / runner-ups.

In our initial implementation of the Sankey diagram the nodes had a rectangular shape. Changing the tournament nodes to show their logos was a challenging part of the implementation.

Design evolution

Our design did not change much. As discussed previously, we initially did not think about the case where there would be too many players, so we were planning to show all players. However, we grouped the less dominant players in one node to make the visualization clear. In addition, we added the players' information in the Sankey Diagram which is displayed when the user hovers over the player's node. In our initial plan, we wanted to place this feature in the network graph but we decided that the diagram was a better choice.

Network graph



Inspiration

We wanted to display all the frequent rivals of the Grand Slams tournaments. Therefore, using the network graph was the best way to identify all the rival players and to show how many times they played against each other using the width of the edges.

Challenges

As in the case with the Sankey Diagram, when using a wide range of years, the network graph is cluttered with the player's nodes and edges. This made it difficult to see the rivalry between players. Therefore, firstly we implemented highlighting of the nodes and edges of the hovered player's rivals, making them stand out in the network. Yet, this did not help enough to provide a clear visualization. Then, we continued by implementing the magnifier glass that the user can activate using a click. Ensuring the proper working of the magnifier was also challenging since the force simulation of the graph needed to be stopped. Using this feature when "magnifying" the surrounding of a given node made it visually easier for the user. In addition to the magnifier glass solution, we implemented the search bar in a way that will reduce the number of nodes and edges in the graph by hiding the irrelevant ones for the searched player.

For displaying the frequency of rivalry between players we used the width of the edges, however, we thought that the user would also be interested in the exact number of games played. Adding these numbers to the network graph was challenging because of the aforementioned problem with the "busyness" of the network. We addressed this problem by displaying these numbers only when the user hovers over the edge. In addition to the number of played games, we added extra information for the matches such as the scores per year and per Grand Slam tournament.

Design evolution

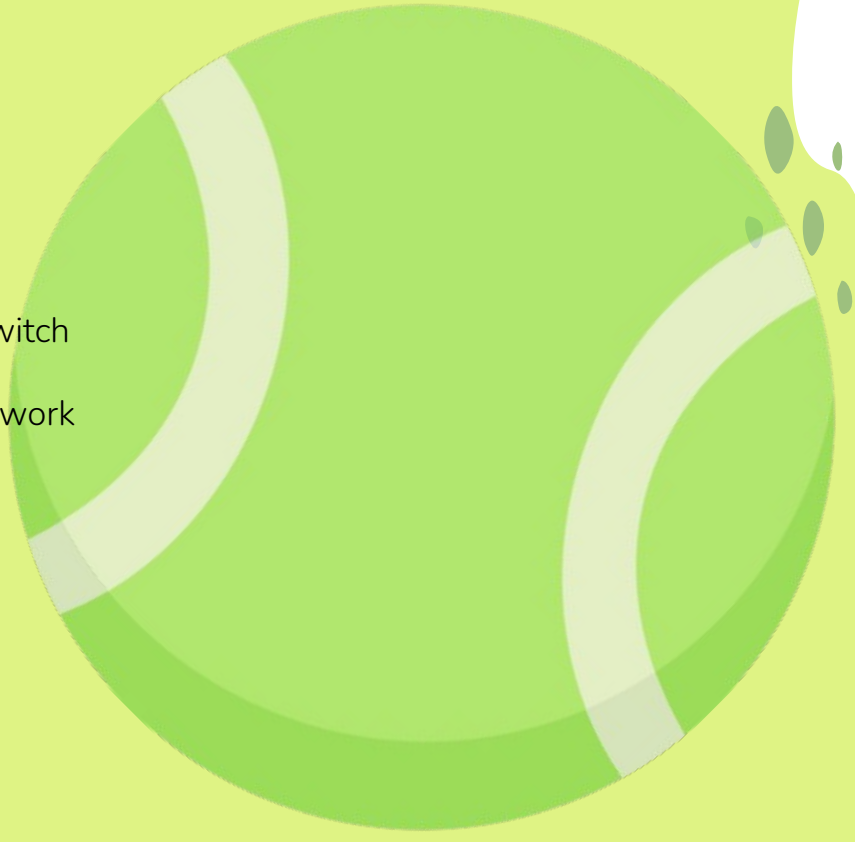
The base of our network graph design did not change much during the whole process, however we did add additional features in order to improve the visualization. We also eliminated the players' information display and added it to the Sankey Diagram since it was hard to hover over the right node when there are too many nodes.

Improvements

There are a few aspects that could be improved. First would be smoothening the transitions of the Sankey diagram. While this improves the user experience, it is not necessary to deliver the message of this visualization. Similarly, the website could be adapted to work on mobile devices, although it would probably be hard to display things clearly on a small screen, and we are not targeting mobile users. One could also link the two main components – when a user chooses a player in the Sankey diagram, the network would refocus on this player (similarly to the search box).

Finally, a possible extension to the project would be incorporating interesting highlights in the display (e.g. in the form of a pop-up) so that the users who are not so knowledgeable of tennis history know what to look for. For instance, if the time range includes the year 2007, we could display a pop-up window in the left margin saying that Roger Federer became the first man in tennis history since Bill Tilden in the 1920's to win 4 U.S. Open titles in a row that year. This could be extended to provide a complete guide for the website, where we pre-define some events that we deem interesting and the user simply follows our lead.

Peer assessment



Marie
Timeline implementation and gender switch
Layout and search box
Miscellaneous tasks for Sankey and network

Marija
Network implementation
Search box functionality
Website appearance

Ljupche
Data scraping
Sankey diagram and player profiles
Bug fixes