

# Data Visualization and Visual Analytics

## Assessment Task 2: Advanced Data Visualization

Student Id: 24883896

<b>Summary</b>	<b>3</b>
<b>Data Exploration</b>	<b>3</b>
<b>Data Preparation</b>	<b>4</b>
<b>Tree Map</b>	<b>5</b>
<b>Parallel coordinate</b>	<b>7</b>
<b>Text Graph</b>	<b>8</b>
<b>Geographic Map</b>	<b>9</b>
<b>Scatter Plot</b>	<b>11</b>
<b>conclusion</b>	<b>12</b>

# Summary

Every year, the prestigious Australia Open tennis tournament takes place in Melbourne, Victoria, Australia, with Melbourne Park as its dedicated venue. The event kicks off in mid-January and extends for a fortnight, concluding on Australia Day. This extensive dataset spans 118 years, from 1905 to 2023, and includes the triumphant matches from both the men's and women's divisions at the Australia Open. The primary aim of this report is to thoroughly explore and visually present the dataset's findings using the Tableau software.

## Data Exploration

To effectively visualize data, it is essential to begin with a thorough exploration of the data and its characteristics. This initial step is crucial in gaining a comprehensive understanding of the data's underlying patterns. Below, you'll find a table summarizing the attributes, their descriptions, and their respective types.

Column name	Type of data	Description
Year	Int.	Year when the game held
Gender	str.	Players' gender
Champion	str.	Players' name
Champion Nationality	str.	champions' nationality
Champion Country	str.	champions' country
Score	str.	scores achieved in each of the five sets
1st-won	Int.	Player won the first set
1st-loss	Int.	Player lost the first set
2nd-won	Int.	Player won the 2nd set
2nd-loss	Int.	Player lost the 2nd set
3rd-won	Int.	Player won the 3rd set
3rd-loss	int.	Player lost 3rd set

4th-won	Int.	Player won the 4th set
4th-loss	Int.	Player lost the 4th set
5th-won	Int.	Player won the 5th set
5th-loss	Int.	Player lost the 5th set
Runner-up	str.	Runner up player name
Runner-up Nationality	str.	Runner up's nationality
Runner-up Country	str.	Runner up's country

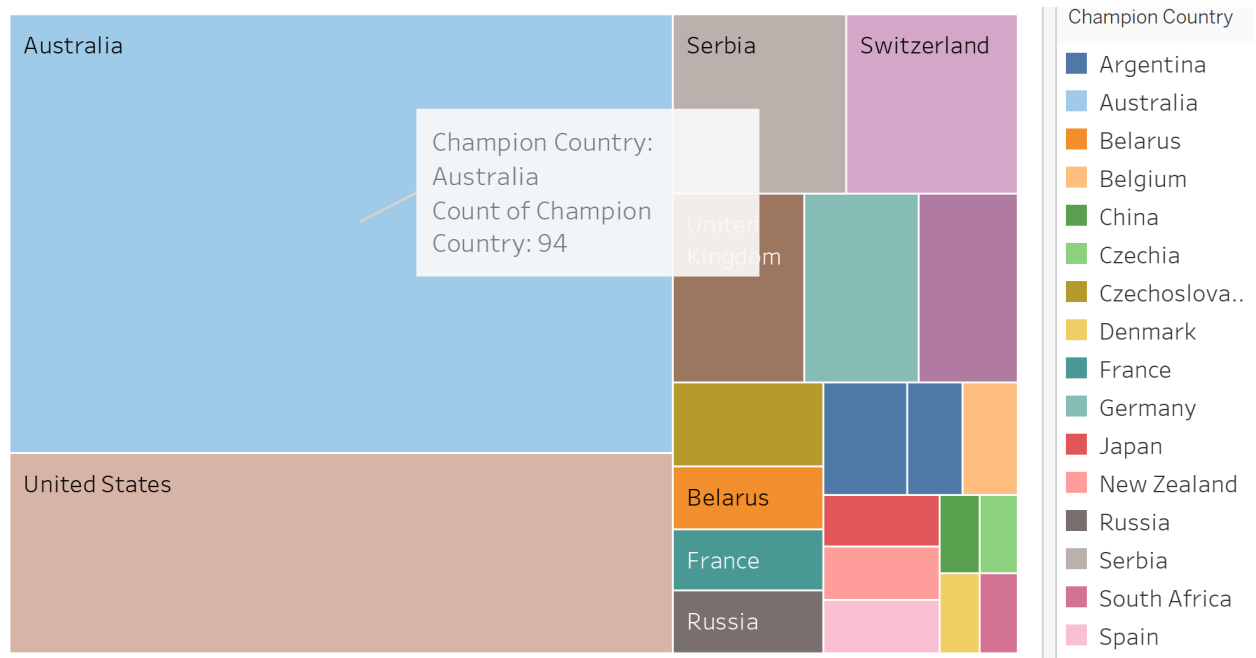
## Data Preparation

Data is a dynamic entity in a constant state of flux. As a result, it's commonplace to encounter various challenges, including dealing with data that isn't particularly useful, addressing missing values, and the necessity of converting numeric data into a string format. To prepare the data for analysis, pre-processing emerges as a pivotal step, encompassing activities like managing missing data, tidying up the dataset, and transforming it as needed. Notably, historical data is inherently subject to changes as it progresses through time. Within the dataset at hand, you might notice fluctuations in how champion and runner-up names are documented, underscoring the importance of maintaining consistency in the way names are recorded. Nonetheless, it's worth mentioning that beneath these variations, there are persistent patterns that can be discerned within the data.

I made a deliberate decision not to substitute or impute missing values with mean or derived values. Instead, I chose to maintain these gaps in the dataset as they were, without alteration.

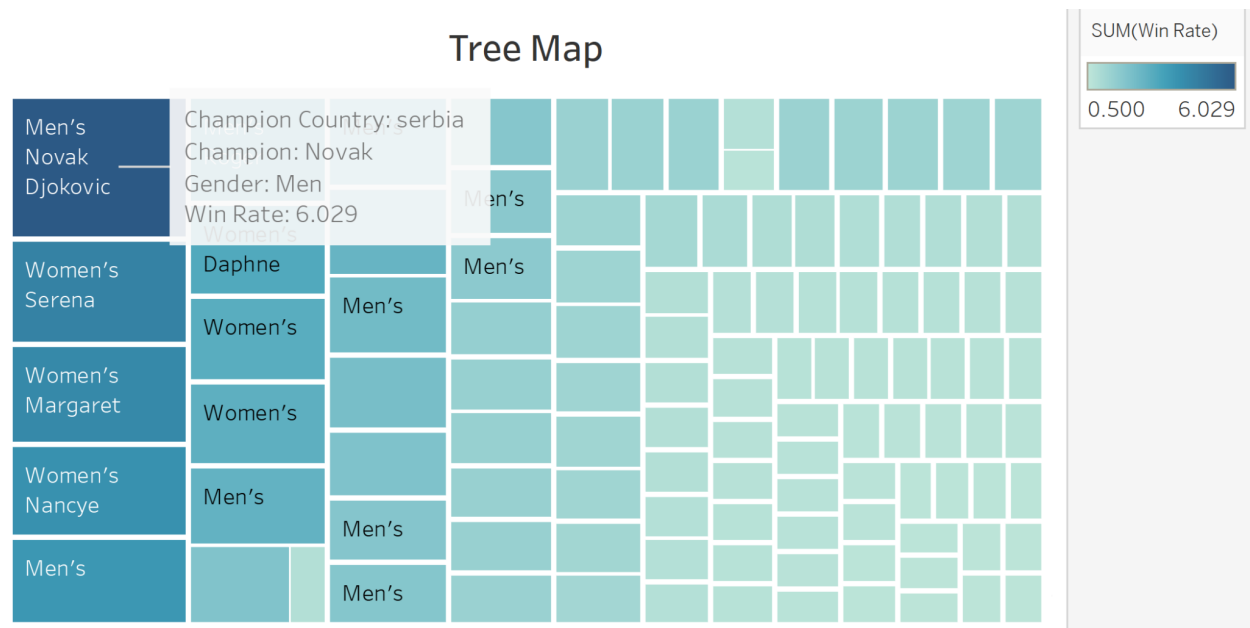
Three additional columns were seamlessly integrated into the dataset, specifically named "wins," "loss," and "win-rate." The "wins" column was meticulously computed as the cumulative total of victories across the five sets, while the "loss" column represented the summation of losses from those same five sets. Calculating the "win-rate" was an equally meticulous process, carried out using the IFERROR formula within Excel. These additions contribute to a more comprehensive and informative dataset, offering valuable insights for further analysis.

# Tree Map



**Figure 1: Tree map**

Figure 1 provides a straightforward application of the tree map, presenting the combined count of both male and female champions. It uses various color shades to indicate the magnitude of this total. A darker shade corresponds to a higher numerical value, resulting in a larger box. Notably, Australia emerges as the country with the highest number of champions, closely followed by the United States. In the period spanning from 1905 to 2023, Australia boasted 94 victorious champions in total.

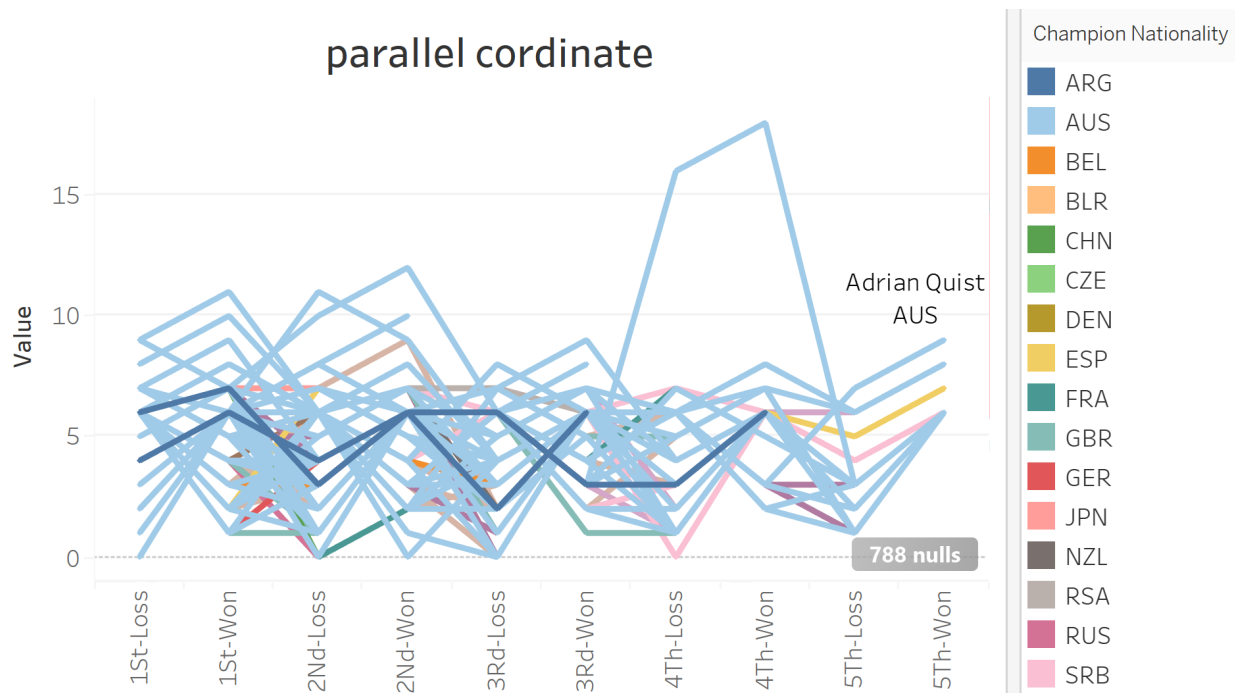


**Figure 2: Tree Map 2 by win rate**

In Figure 2's tree map visualization, the presentation of champion data is ingeniously structured, efficiently conveying crucial information about champions, their countries, gender, and their win rates. This visualization is thoughtfully designed to enhance data comprehension and engagement. The hierarchical layout of the tree map is a standout feature, with distinct branches dedicated to different genders, clearly illustrating the gender-based distribution of champions. Within these branches, the inclusion of champion names and their corresponding countries allows viewers to swiftly identify individual champions and their home countries. Furthermore, the use of color-coded shading to represent win rates adds a layer of depth to the visualization, making it effortless to pinpoint champions with exceptional records. The addition of a well-placed title at the top of the visualization provides vital context, acting as a guiding headline for the viewer, setting the stage for the data's exploration. Additionally, the strategically located win rate bar in the corner serves as a valuable reference point, offering a clear scale for comprehending win rates and facilitating comparisons among champions. Altogether, Figure 2 is a testament to effective data visualization, combining structure, color, and context to facilitate a deeper understanding of the dataset's champion-related information in a visually engaging manner.

# Parallel coordinate

In Tableau, the parallel coordinate plot, also referred to as the parallel coordinates chart, stands as a dynamic visualization tool used to analyze intricate multidimensional data. This visualization consists of parallel axes, each representing distinct data attributes, with lines connecting data points along these axes. This graphical representation offers an efficient way to uncover patterns, relationships, and exceptional data points within datasets that encompass numerous dimensions. In contrast to conventional charts, parallel coordinates provide a comprehensive perspective on data interactions, simplifying the identification of trends and anomalies. Tableau's interactive features enable users to construct and explore these plots, making it a valuable resource for data analysts seeking to reveal valuable insights within a wide range of datasets, from examining customer behavior to delving into financial data.



**Figure 3: Parallel coordinate by top player**

In Figure 3, we have a visual representation that presents the Australian Open's championship history. It displays five sets of wins and losses, offering a straightforward view of each champion's performance. The graph further breaks down the data by champion nationality, gender, match scores, and years. It includes champion names and their countries, adding a personal touch to the data. The use of different colors for each line helps distinguish data points visually. Clear labels and organized axes make it accessible for various viewers. Notably, one top player, Gerald from Australia, achieved the highest score with his 4th win. This graph provides a way to track champions over

time, compare countries, and understand the performance of male and female champions. Figure 3 is a functional tool for exploring the Australian Open's history, offering a data-driven perspective.

## Text Graph

The use of text graphs in data visualization is relatively uncommon due to the challenge of maintaining clarity when dealing with numerous attributes. In this specific scenario, we've associated the "champion's name" attribute with the shapes in the graph, and we've transformed its essence into a count. As a result, it yields an integer value that serves as a representation of how many times a player has achieved victory and earned the champion title. The intriguing part is that this integer value dictates the size of the shapes, causing them to either expand or contract accordingly. Notably, when this integer value is higher, the player's name takes center stage with a larger and more prominent font size, effectively shining a spotlight on their accomplishments.

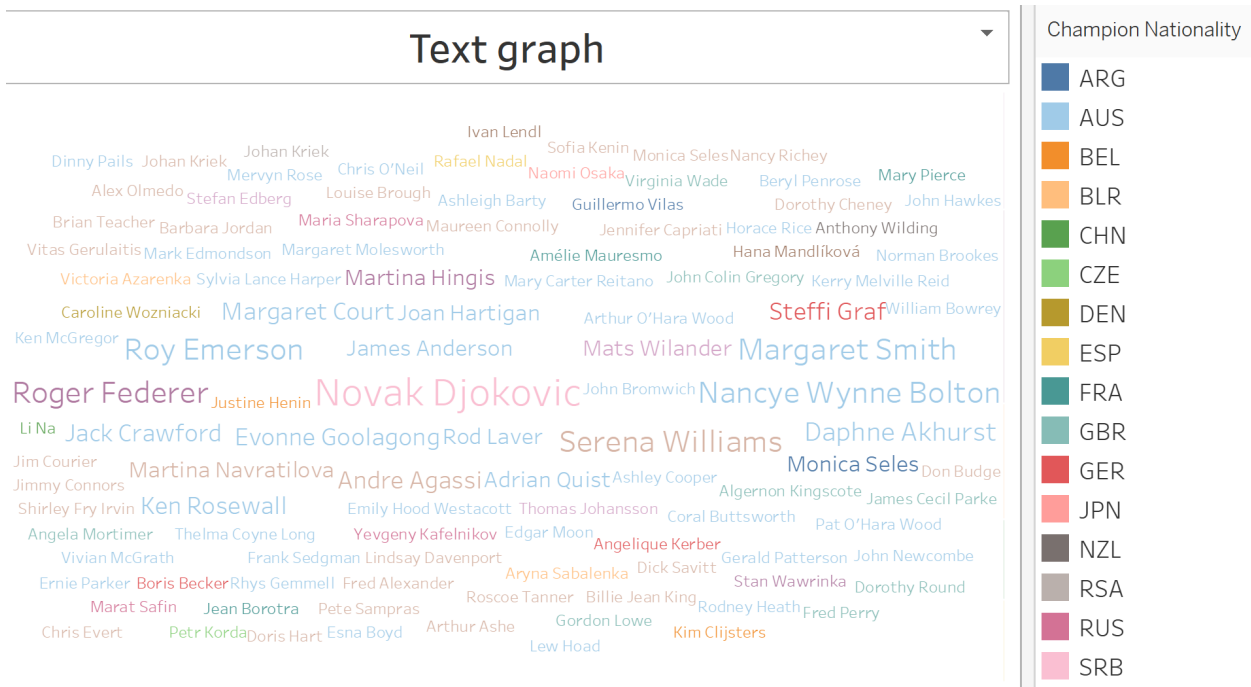


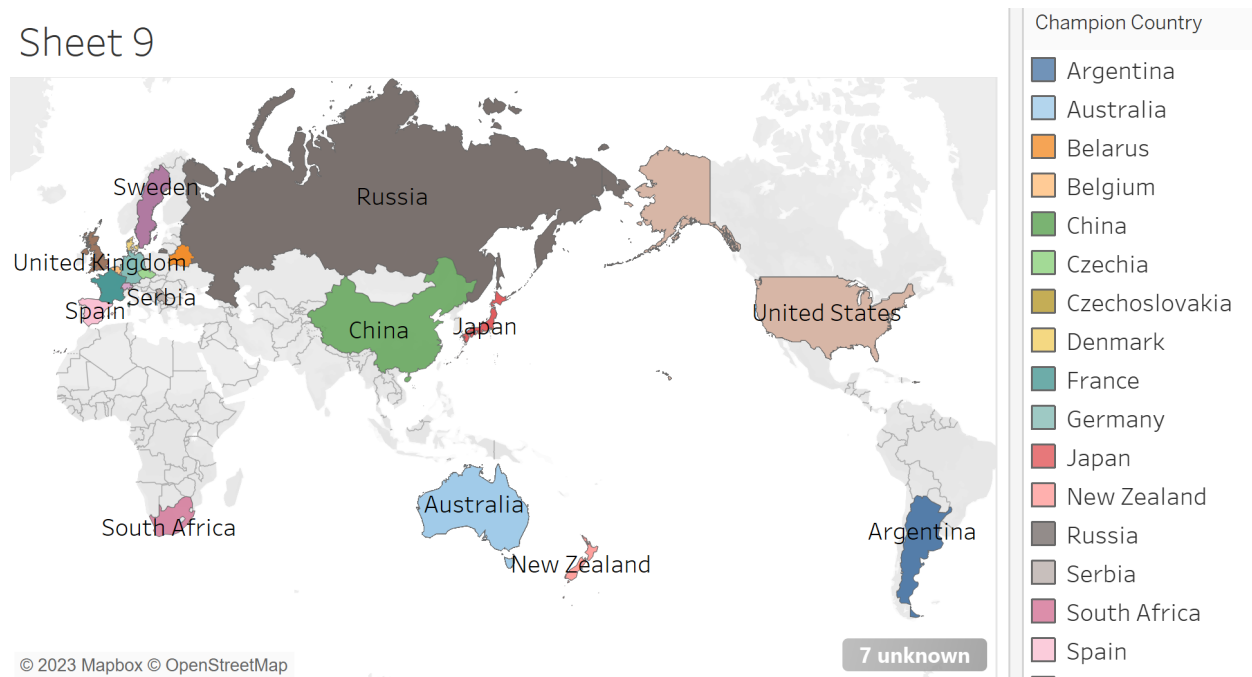
Figure 4: Text Graph by champion

In figure 4, the champion's home country is represented through a diverse array of colors. It's worth highlighting that Novak Djokovic's name takes the spotlight with the largest font size who is a top player, denoting his exceptional achievement of clinching the Australia Open title a remarkable ten times within the given timeframe. In contrast, as the number of victories decreases, the font size undergoes a natural reduction.



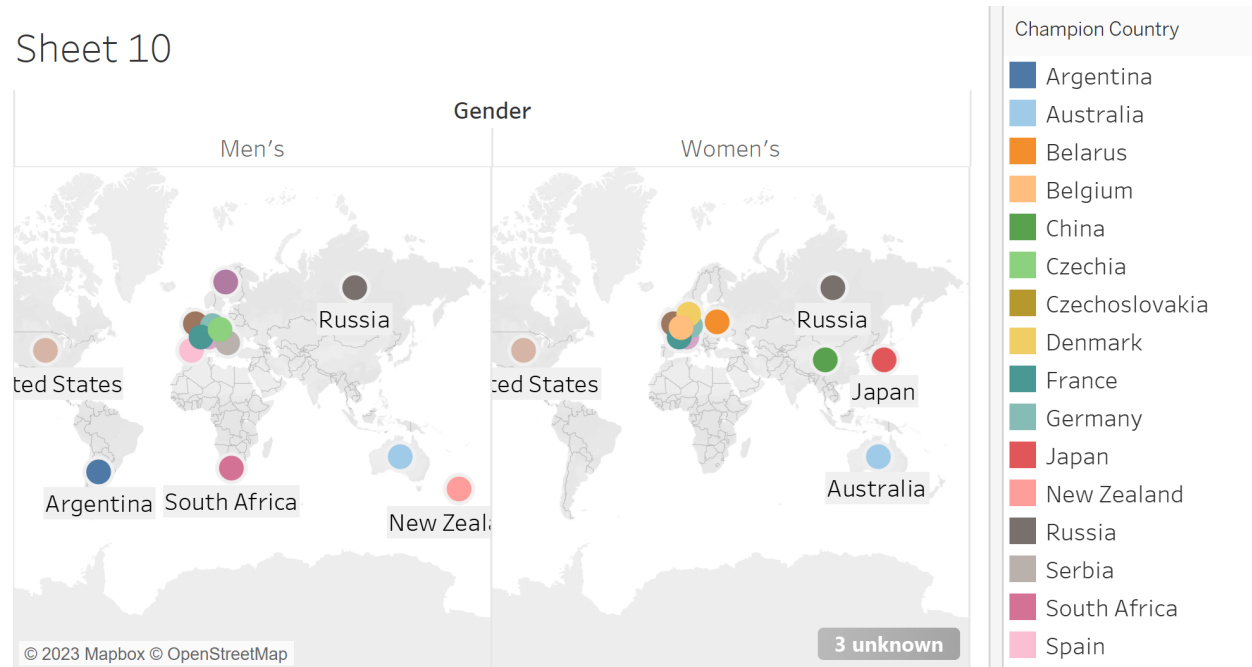
# Geographic Map

Sheet 9



**Figure 5: Geographical Map by country**

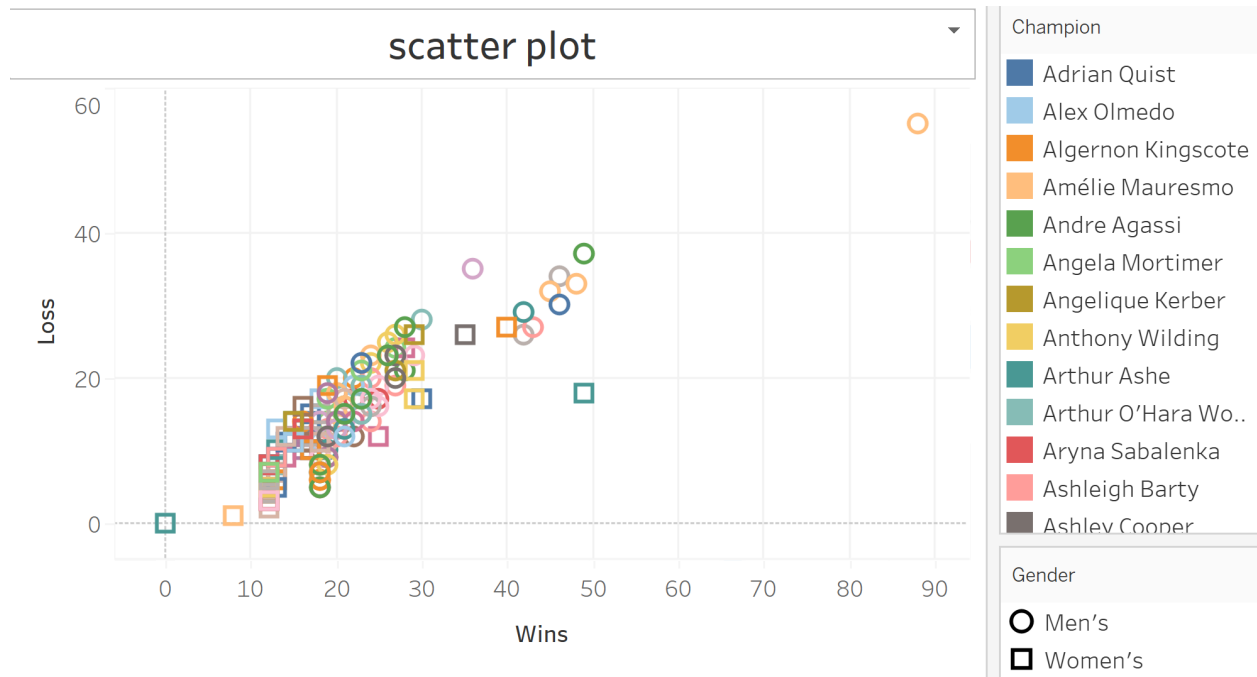
In the process of creating a geographic map which shows in figure 5, the key components were thoughtfully aligned to generate a rich and informative visualization. Longitude values were assigned to the columns, while latitude values were designated for the rows, forming the foundational coordinates for plotting geographical data. The "champion country" attribute was thoughtfully integrated into the filter, enabling users to dynamically select and view specific countries of interest. What truly enhances this map's interactivity is the way champion country names, along with additional details such as gender and score, are seamlessly displayed when hovering the cursor over each data point. This feature transforms the map into an engaging and data-rich experience, allowing users to explore not only the geographical distribution of champions but also the specific attributes and achievements associated with each champion. The result is a dynamic visualization that not only pinpoints the geographical origin of champions but also offers a deeper layer of insight, making it an invaluable tool for understanding the diverse dimensions of champion data.



**Figure 6: Geographical Map by gender**

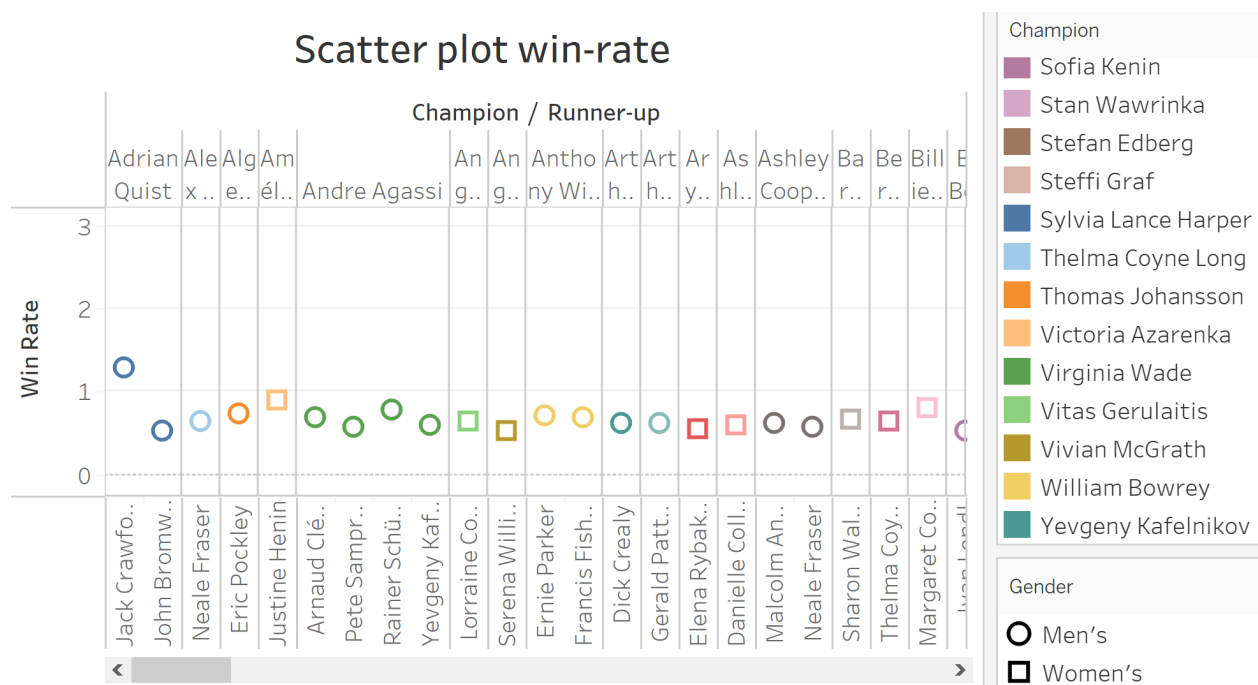
In this geographical map (figure 6), top players' countries are distinctly highlighted based on gender. The visualization is separated into two sections, one for each gender. It offers a clear and concise overview of the standout champions from each gender category within their respective countries, enabling users to easily identify and compare the top players on a global scale.

# Scatter Plot



**Figure 7: Scatter plot top player**

This scatter plot neatly organizes and displays the data based on wins and losses. Wins are placed along the horizontal axis (columns), while losses are represented on the vertical axis (rows). A filter for gender allows users to explore the data from the perspective of male and female champions. What's particularly insightful is that it employs different shapes to distinguish between the two genders, enhancing visual clarity. In addition to this, champion names are thoughtfully displayed on the plot when hovering the cursor, providing a direct link between data points and the top players they represent. This helps users to immediately identify and associate the performance of individual champions with their specific wins and losses.



**Figure 7: Scatter plot win-rate**

The champions and runners-up data are thoughtfully organized within the columns, providing a clear and accessible layout. At the core of this visualization is the win rate, which serves as a pivotal metric for evaluating the top players. Additionally, the use of filters for "champion" and "gender" adds a layer of interactivity and flexibility, allowing users to focus on specific subsets of the data. What makes this technique particularly informative is its capacity to display not just the win rates but also the names of the top players and runners-up, providing a comprehensive view of the champions' performance based on their gender. It becomes a visual journey through the achievements of the top players, offering a dynamic snapshot of their success and highlighting the key players in the competition.

## conclusion

In the realm of data visualization, a realm where countless hidden trends and patterns lie in wait, there emerges a potential for shaping a brighter future. Following a meticulous analysis and visualization of the dataset at hand, several key revelations

have come to the fore, providing invaluable insights for both the inquisitive observer and the data aficionado.

As we delve into the visual aspects of our exploration, it's evident that the Tree Map, with its adept use of distinct colors and varying shades to represent different categories and highlight essential Key Performance Indicators (KPIs), stands out. This visualization method has a unique ability to transform complexity into clarity, making it easy for viewers to intuitively grasp information with a single glance. Furthermore, the Tree Map's feature of dynamically sizing rectangles based on specific attributes offers a straightforward means of identifying the highest and lowest values, rendering it a powerful tool for data comparison.

The Parallel Coordinates Graph, on the other hand, shines as an advanced technique ideally suited for dissecting high-dimensional data. Its unique approach involves parallel examination of multiple dimensions with equal distances between axes, thereby allowing a comprehensive view of the data. While the visualization of correlations between adjacent axes is a challenge for non-technical users, it opens the door to profound insights. Notably, Geographical Maps offer a user-friendly experience, presenting data in an accessible and appealing manner. Shifting our focus to the Australia Open Championship dataset, a tapestry of captivating narratives unfolds: Australia emerges as the unrivaled leader with an impressive 94 titles, closely followed by the USA. Women's remarkable performance is marked by both quantity and higher winning rates in Australia and the USA. The legendary Gerald Patterson's record-setting 18-point single set serves as a testament to the sport's history. A mere 11% of matches venture into a fifth set, with the majority concluding within three. Margaret Smith's razor-thin victory margin of zero points adds intrigue to her achievements. Novak Djokovic, a tennis titan, boasts an impressive ten titles. A remarkable win rate of nearly 86% against Jan Lehand positions Margaret Smith as a dominant figure of her era. Additionally, a stark contrast between Asia's female champions and Europe, Africa, and Latin America's male champions is evident, contributing to the diversity of the tournament's history.

These insights not only illuminate the rich history of the Australia Open but also underscore the profound impact of data visualization in uncovering stories hidden within the numbers. In this ever-evolving realm, data's power to reveal untold narratives and guide future decisions remains truly remarkable.

