Australian Open Data Analysis and Visualization

XIN DENG (DORA)

Table of Contents

Introduction

This report analyses and visualizes Australian Open data since 1905 to provide insights for tennis enthusiasts. It identifies key changes in the dataset, focusing on the top 5+ competitors. Utilizing various visualization techniques via Tableau, the report presents high-level trends of the competitions and delves into specific match-related statistics for individual players and their games. Additionally, it highlights the advantages and disadvantages of using Tableau for data analysis and visualization.

Data Summary

This dataset summarizes the finals of the Australian Open throughout its history since 1905. It includes information such as the champions' names, countries, genders, scores, as well as the corresponding runner-up names and nationalities.

This dataset encompasses various data types and formats, reflecting the diverse nature of the Australian Open and its historical evolution. Additionally, it includes special characteristics and outliers inherent to the tournament's dynamic history.

Year is quantitative(interval-scale) as the order and the difference between two years is meaningful and it cannot be 0.0. (*1.2.2 The different types of data*, 2024)

Characteristics & Outliers:

- The data starts from 1905, ends 2024.
- There was only male data point between 1905-1921.
- Data points are missing between 1916-1918 due to World War I.
- Data points are missing between 1941-1945 due to World War II.
- There are dual records in 1977 due to a change in scheduling.
- There were no records in 1986 due to various factors, including a change in venue and financial difficulties.

WinRate is quantitative(ratio-scale) as there is a meaningful order to the data, and a clear definition of 0.0.(*1.2.2 The different types of data*, 2024)

	Data Type	Data format
Year	Quantitative (interval-scale)	number
Gender Champion Champion Nationality Champion Country	Categorical nominal	text
Score		
WinRate	Quantitative (ratio-scale)	percentage
1st-won 1st-loss 2nd-won 2nd-loss 3rd-won 3rd-loss 4th-won 4th-loss 5th-won 5th-loss Wins Loss	Quantitative (ratio-scale)	number
Runner-up Runner-up Nationality Runner-up Country	Categorical nominal	text

These are derived from the Score, contains number of games for each set and the totals. They are quantitative(ratio-scale).

These four dimensions are categorical nominal as they describe variables with categories that do not have a natural order or ranking.(*1.2.2 The different types of data*, 2024)

Characteristics & Outliers:

Gender: There were only men's records between 1905-1921.

Champion Nationality is the corresponding country code, same as champion country.

Score: This is a series of number regarding number of games for each set in the whole match.

- In 1966 Margaret Smith won the final as walkover and do not have Score Records.
- The score system was different in the early years.

Runner-ups are the players that finishes in second place in the tournament.

Runner-up nationality is the corresponding country code, same as runner-up country.

These three are categorical nominal as they describe variables with categories that do not have a natural order or ranking.(*1.2.2 The different types of data*, 2024)

Data Transformations & Calculations

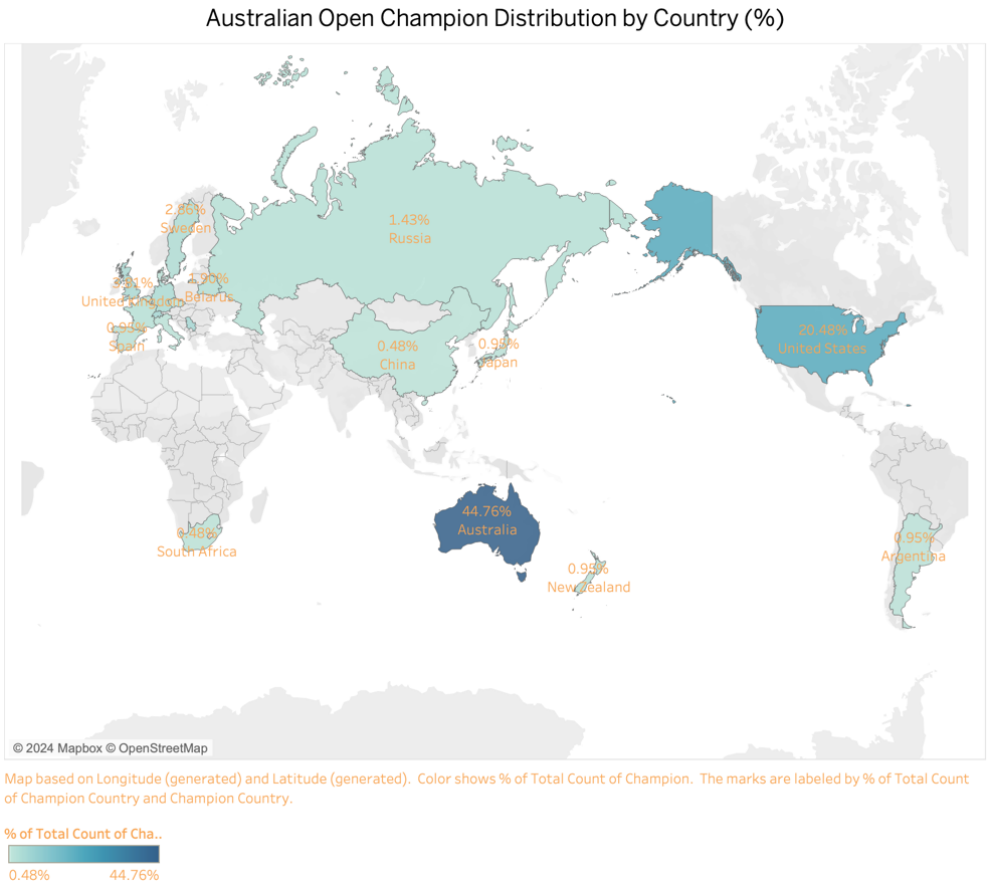
Below is a summary of the data transformations and calculations performed during the analysis, with some conducted in Excel and others in Tableau.

Excel	Tableau
• Sum up all the wins and losses for Australian Open, calculated the Win Rate. • Basic statistic about championship tenure. • Summarisation of % Single-Time Champions. • The calculation of Average Game Per Match for Top5+.	• Change Year from string to number for grouping and filtering purposes. • Calculate the championship tenure by subtract first year as champion from last year as champion. • Calculate Win Rate for each set. • Grouping year into three (before 1922, between 1922-1977, after 1977) for filtering and comparison purposes.

Analysation & Visualisation of Australian Open

First, I conducted a general summary of the entire dataset. A geographic map was generated to provide an overview of the distribution of champions across different countries throughout Australian Open history.

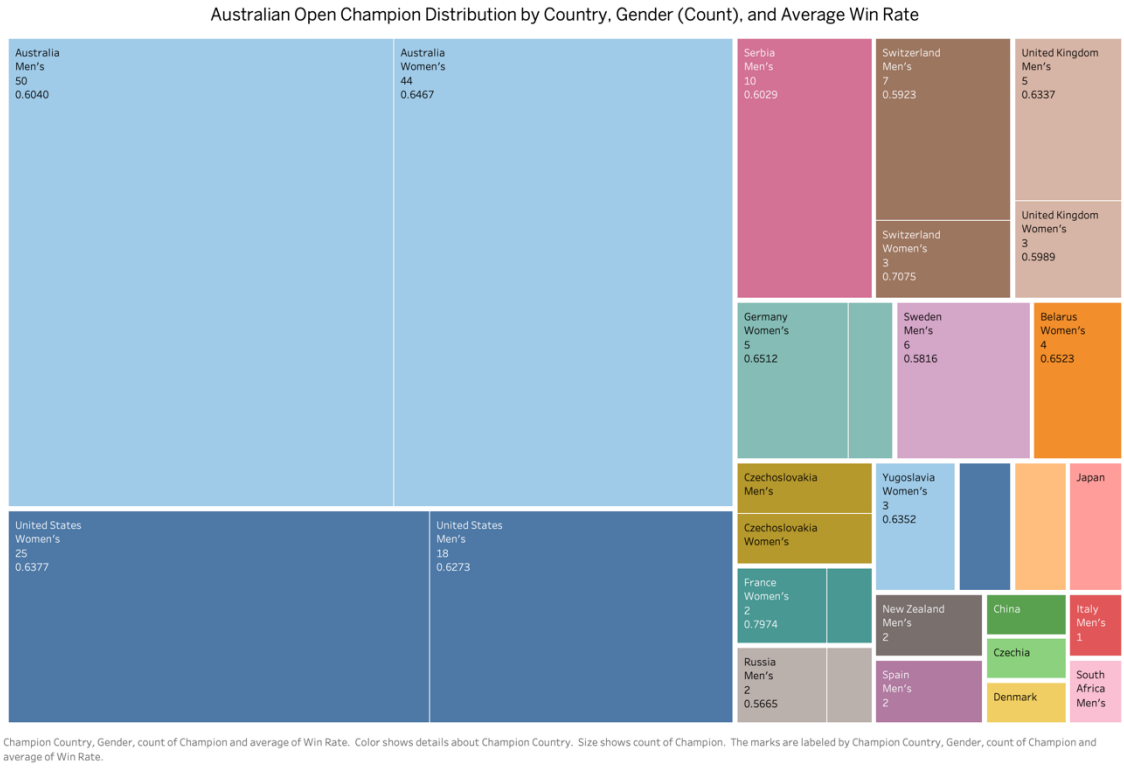
The colour represents the density of data, with darker shades indicating a higher percentage. Data labels are utilized for better clarification.



Data Insights:

- Australia has dominated the number of champions throughout the entire history of the Australian Open.
- The United States holds the second-largest percentage of champions.
- Apart from Australia and the US, Europe has also produced a significant number of champions.

This Treemap provides detailed insights into the distribution of Australian Open Champions. Colours distinguish between different countries, while the size of each block represents the number of champions. Additionally, blocks are further divided by gender where applicable. Key data, including the corresponding average win rate, are labelled to offer better insights.



Data Insights:

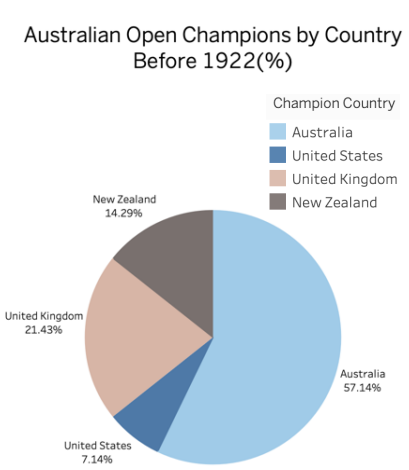
- The number of male champions in Australia slightly exceeds that of female champions, which contrasts with the situation in the US.
- championships are predominantly won by males in some countries, while in others, females dominate.

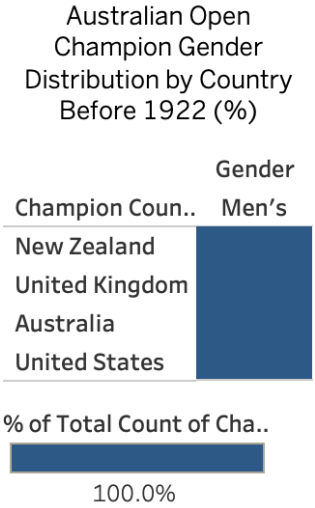
The Analysation and visualisation of nationalities, gender, and changes over time.

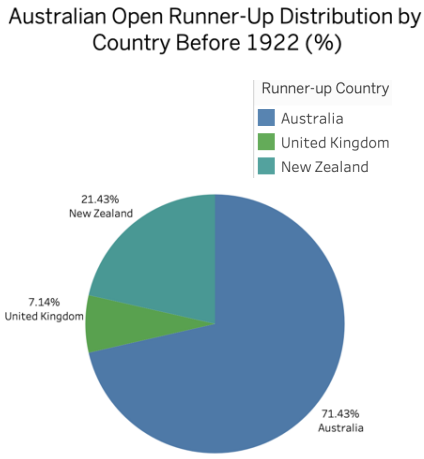
There are three significant time periods in the history of the Australian Open: before 1922, between 1922 and 1977, and after 1977. Before 1922, only males participated in the competition. During the early years, international participation in tennis tournaments was limited due to factors like travel restrictions and financial constraints. Although players from other countries may have attended the Australian Open during this time, their numbers were likely much smaller compared to Australian participants. Traveling long distances to Australia for a tennis tournament was not as feasible or common as it is today, especially for players from countries outside the region.

The year 1978 marked a pivotal moment in the tournament's history, as it was the first time a player from a non-traditional tennis powerhouse nation became champion. This signalled the Australian Open's evolution into a global tournament. Additionally, 1978 saw the transition from grass courts to hard courts as the playing surface for the tournament. This change had a significant impact on the game and introduced new challenges for the players, leading to different styles of play. Moreover, the shift to hard courts made the tournament more accessible to players from around the world who were accustomed to this surface, further contributing to the internationalization of the event.(*Australian Open*, 2024)

Before 1922





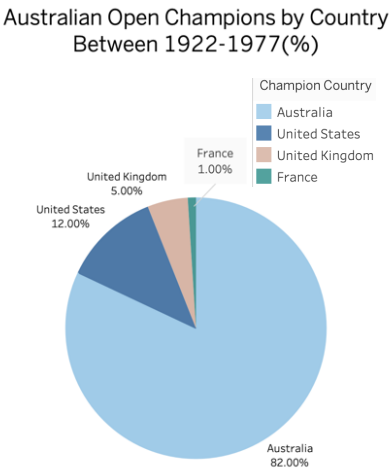


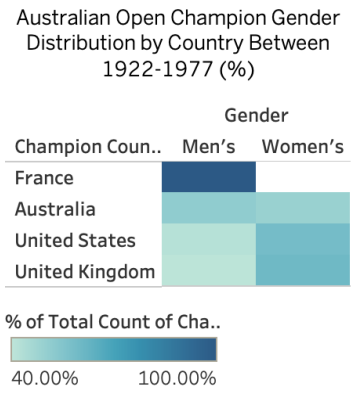
These two pie charts provide insights into the distribution of champions and runner-ups by country before 1922 in the Australian Open. The colours differentiate between different countries, while the labels provide precise percentages for each segment. Additionally, a small table displays the gender distribution by country during that period.

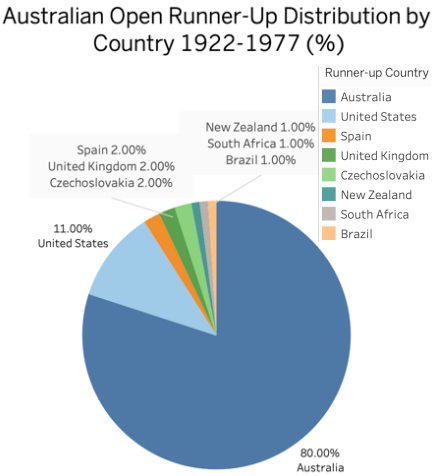
Data Insights:

- Champions before 1922 only come from four countries, Australia, UK, New Zealand, and the US.
- Up to 57.14% champions are from Australia, followed by UK 21.43%.
- Runner-up only from three countries including Australia, NZ, and UK, with Australian takes up to 71.43% of the total.
- Only males were in these data during this time.

Between 1922 and 1977







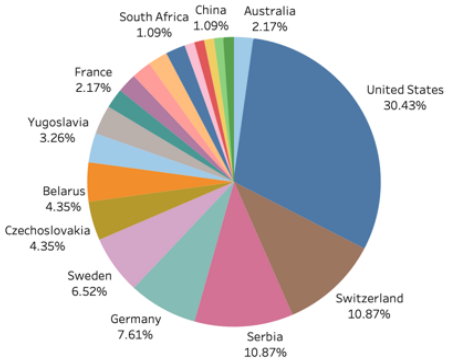
These two pie charts provide insights into the distribution of champions and runner-ups by countries between 1922 and 1977 in the Australian Open. The colours differentiate between different countries, and the labels provide precise percentages for each segment. Additionally, a small table displays the gender distribution by countries during that time, with colour indicating the density of the data — darker shades representing higher percentages of the total.

Data Insights:

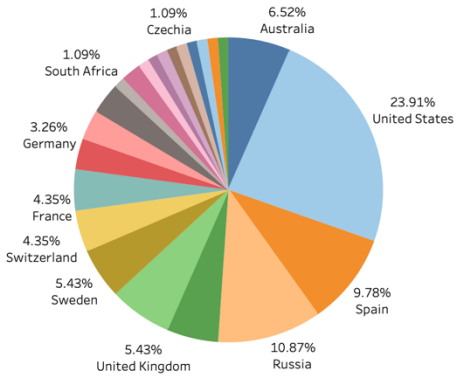
- New Zealand is no longer listed among the champion countries and only represents 1% of the total runner-ups. This could be attributed to the increased number of players participating in the tournament, coupled with New Zealand's relatively small population.
- France has begun to secure a position on the list of champion countries.
- Australia maintained its dominant position in both the champion and runner-up lists, accounting for more than half of the percentages.
- Females began to establish a presence in the championship. Except for France, where men dominate, the rest of the countries have both genders evenly represented in the champion list.

After 1977

Australian Open Champions by Country After 1977(%)



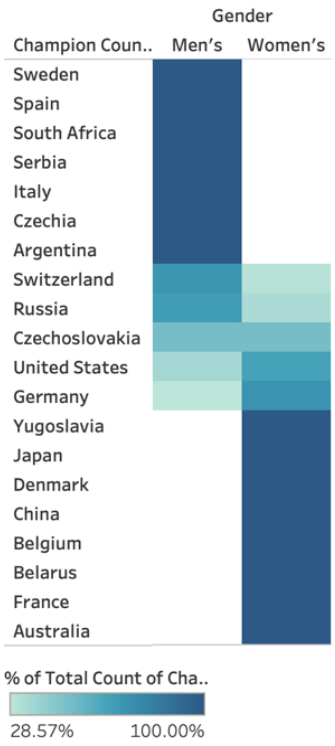
Australian Open Runner-Up Distribution By Country After 1977 (%)



Australian Open Champions by Country After 1977(%)

Champion Coun..	
United States	30.43%
Serbia	10.87%
Switzerland	10.87%
Germany	7.61%
Sweden	6.52%
Belarus	4.35%
Czechoslovakia	4.35%
Russia	3.26%
Yugoslavia	3.26%
Argentina	2.17%
Belgium	2.17%
Japan	2.17%
Spain	2.17%
France	2.17%
Australia	2.17%
China	1.09%
Czechia	1.09%
Denmark	1.09%
Italy	1.09%
South Africa	1.09%

Australian Open Champion Gender Distribution by Country After 1977 (%)



Australian Open Runner-Up Distribution by Country After 1977(%)

Runner-up Cou..	
United States	23.91%
Russia	10.87%
Spain	9.78%
Czechoslovakia	6.52%
Australia	6.52%
Sweden	5.43%
United Kingdom	5.43%
France	4.35%
Switzerland	4.35%
Belgium	3.26%
China	3.26%
Germany	3.26%
Chile	2.17%
Austria	1.09%
Croatia	1.09%
Cyprus	1.09%
Czechia	1.09%
Greece	1.09%
Kazakhstan	1.09%
Romania	1.09%
Serbia	1.09%
Slovakia	1.09%
South Africa	1.09%

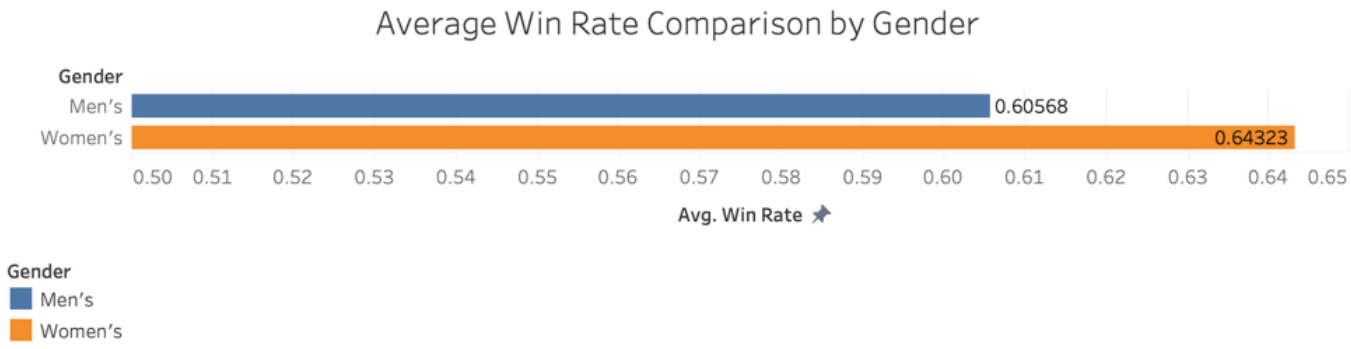
These two pie charts help identify the percentage of champions and runner-ups by countries after 1977. The colour distinguishes different countries, and labels provide exact size information. Additionally, the middle table show gender distribution by country, with darker shades indicating higher percentages. To accommodate the increasing data, two tables list all percentage distributions for champions and runner-ups, providing clearer insights.

Data Insights:

- Australia's dominance in both the champion and runner-up lists has diminished, representing only 2.17% of champions and 6.52% of runner-ups.
- The US has emerged as the leading country in both the champion list (30.43%) and the runner-up list (23.91%).
- The US is the only country that maintains a consistent presence in both lists. Other countries either dominate the champion list or the runner-up list, but not both.
- It is intriguing to note that some countries exhibit a male-dominated champion list, while others have a female-dominated one, with the gender distribution being relatively even across nations.

More analysis regarding the players' performance over the Australian Open History.

Average Win Rate VS Gender



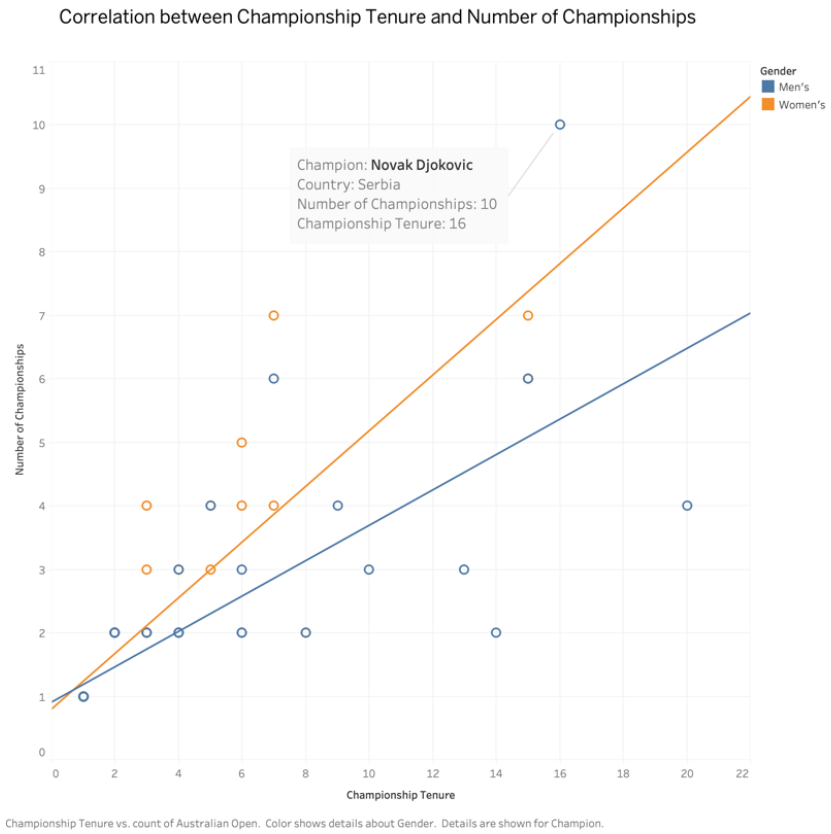
This bar chart compares the average win rate between genders since 1905. The axis has been rescaled between 0.5 and 0.65 to facilitate clear comparison between the two datasets. The use of distinct colours clearly distinguishes between genders, while labelled markers provide additional clarification.

Data Insights:

Over the history of the Australian Open, women have exhibited stronger performance compared to men, with an average win rate higher by 0.03755.

Championship Tenure VS Number of Championships

This scatter plot evaluates the correlation between championship tenure and the number of championships for all male and female players in Australian Open history. Gender is distinguished by different colors for better visualization, and annotations have been added to clarify outliers. Trend lines are included for both genders to provide a more accurate analysis.

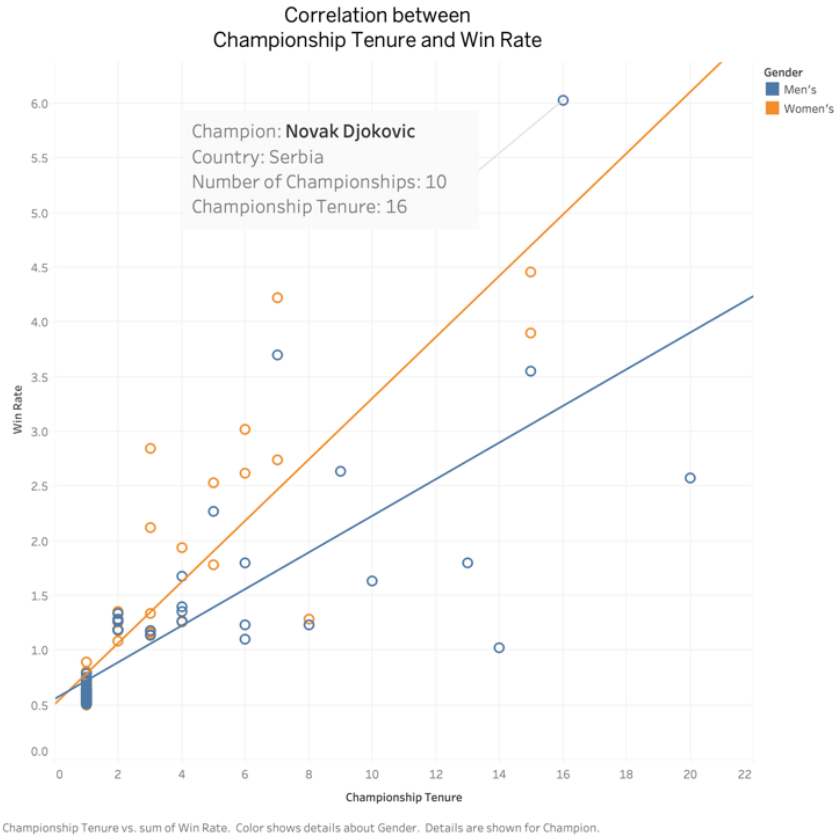


Data Insights:

- The longer the championship tenure, the more wins for the players, as indicated by the positive correlation observed in the scatter plot.
- The trend line for female players appears steeper, indicating a more pronounced increase in wins with longer championship tenure compared to male players.

Championship Tenure VS Win Rate

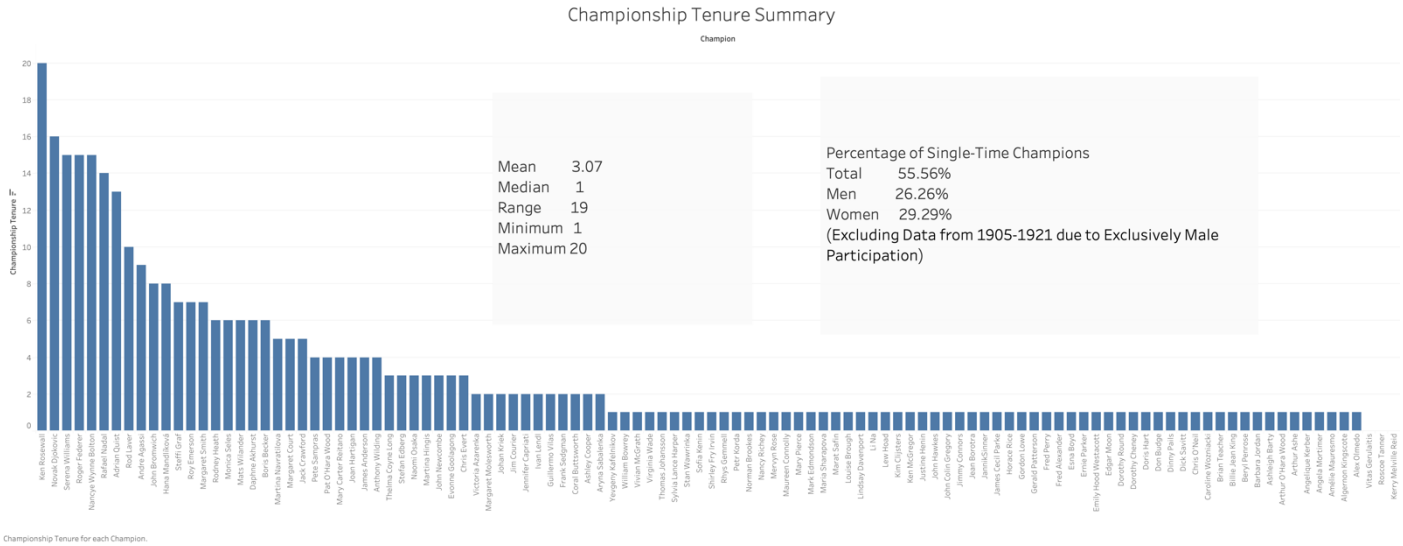
The scatter plot evaluates the correlation between championship tenure and win rate for male and female players in Australian Open history. Gender is differentiated by colour for clarity, and annotations highlight outliers. Trend lines for both genders provide further insights into the data.



Data Insights:

- Higher win rates are observed with longer championship tenures.
- The trend line appears steeper for females, indicating a more pronounced relationship between championship tenure and win rate.

Championship Tenure Analysis

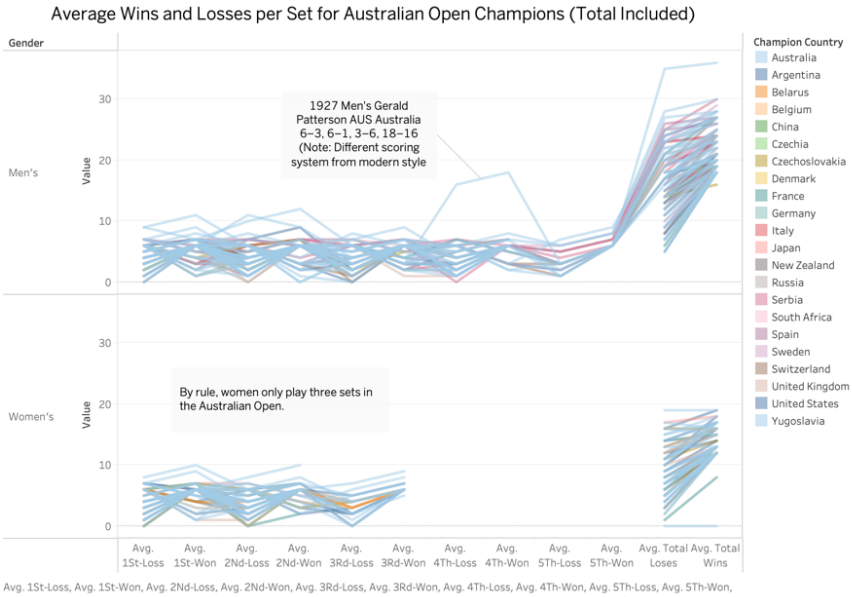


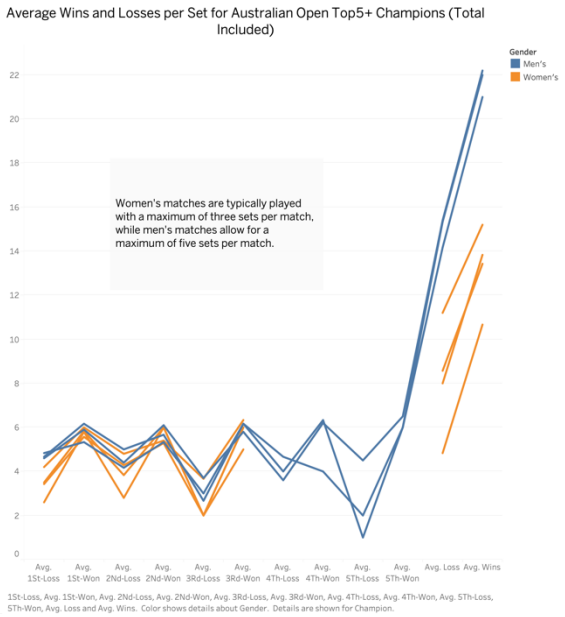
The bar chart illustrates the distribution of championship tenure for all Australian Open champions throughout history. The data has been arranged in descending order to highlight patterns, with annotations provided for further context.

Data Insights:

- Over half of the champions are single-time winners, comprising 29.29% female champions and 26.26% male champions.
- One player had a tenure of 20 years, which was four years longer than the second-ranked player.

Average Wins and Losses Parallel Coordinates





The above visualizations consist of two parallel coordinate plots. The left plot represents average wins and losses for all Australian Open champions throughout history, while the right plot focuses specifically on top5+ champions.

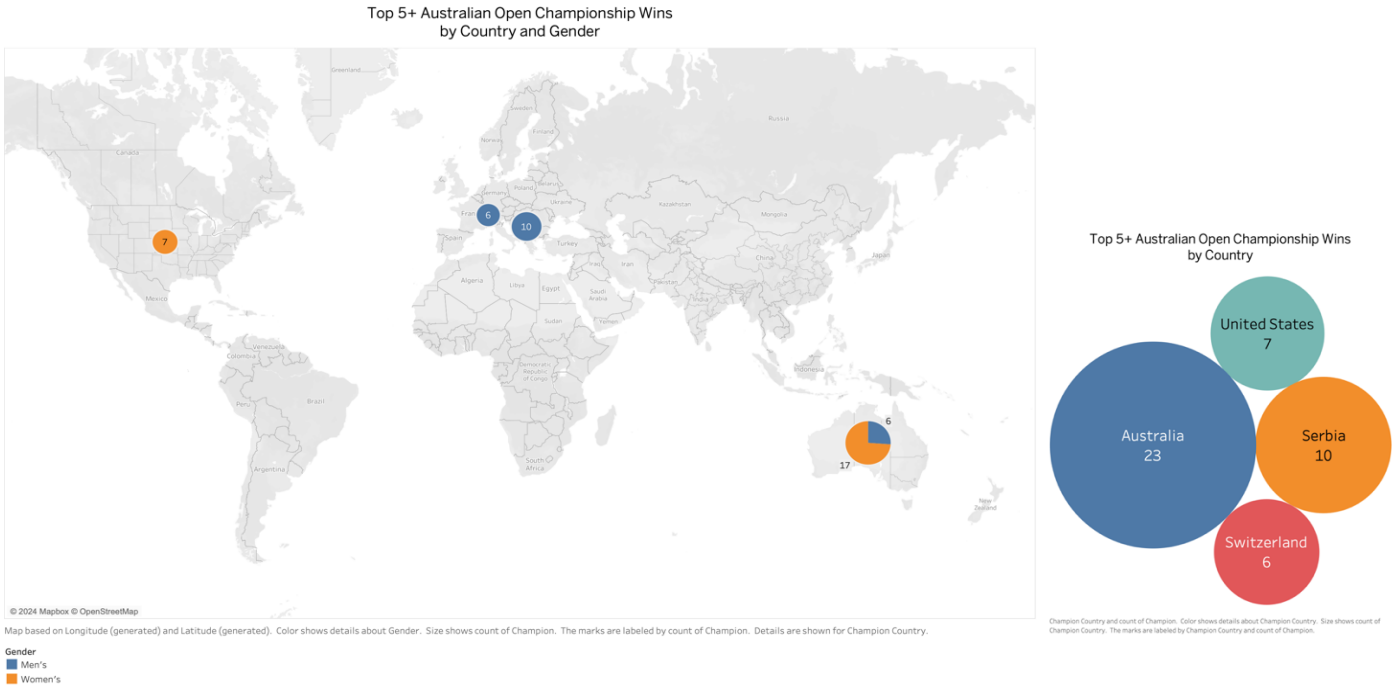
In the left plot, lines are coloured by country and separated into upper and lower sections by gender. To improve clarity, the opacity of the colours has been reduced to 50%. Annotations are included to highlight outliers and enhance storytelling.

In the right plot, each line represents a top5+ champion, with gender identified by color. Annotations provide additional context for storytelling.

Data Insights:

- In the left chart, it's evident that Australia has been the dominant country throughout Australian Open history.
- Furthermore, in the left chart, there's a discernible pattern where some champions initially experienced high losses, while others began with low losses. Interestingly, the direction of their respective lines on the chart exhibits opposite trends.
- Conversely, upon examining the top5+ champions on the right side, it's noticeable that all lines originate from a low loss point, indicating consistent performance. This pattern remains consistent across all top5+ champions.

More Analysation & Visualisation of Australian Open Top5+



The geographic map and bubble charts above illustrate the distribution of top5+ champions by championship counts and country. The bubble chart utilizes color labels to differentiate countries, while the geographic map employs color labels to differentiate gender. Labels enhance clarity in both visualizations.

Data Insights:

- Australia has the same number of championships as the other three countries combined, with a total of 23 titles.
- All top5+ champions from the US are female, while those from Serbia and Switzerland are all male.
- Female players have won 17 championships for Australia, whereas male players have won only 6 championships.

Australian Open Top 5+ Champions: Names, Number of Wins, and Nationality

Champion	Champion Country: Australia	Serbia	Switzer..	United States
Daphne Akhurst	5			
Margaret Smith	6			
Nancye Wynne Bolton	6			
Novak Djokovic		10		
Roger Federer			6	
Roy Emerson	6			
Serena Williams				7

Count of Champion broken down by Champion Country vs. Champion. Color shows details about Gender.

Gender
Men's
Women's

The table on the left summarizes the top5+ champions, including their names, number of wins, and nationality. The champions' gender is indicated by colour. On the right, word clouds display the frequency of champion names, with larger words representing more frequent names. Each name is distinguished by colour for clarity.

Data Insides:

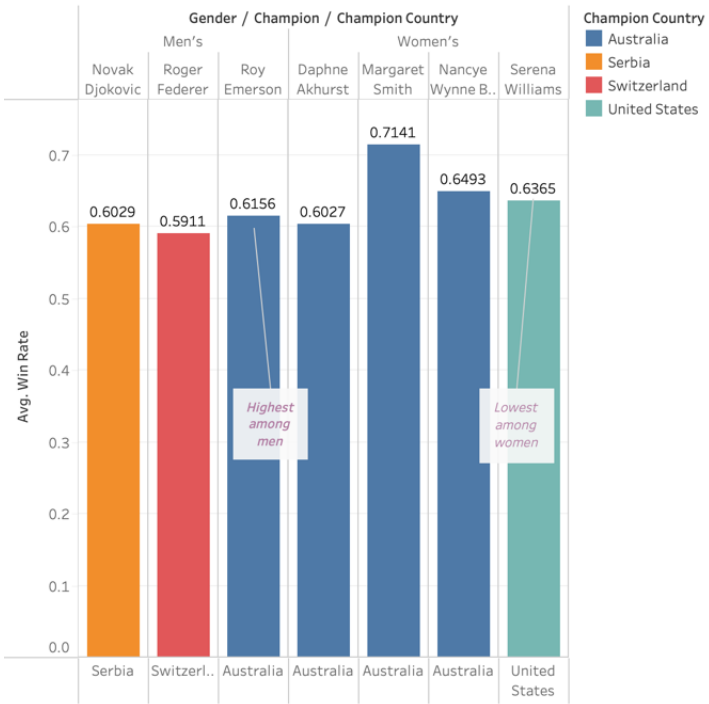
- Australia boasts the highest number of top5+ champions in Australian Open history, with four champions hailing from the country. This count surpasses the combined total of champions from all other countries.
- Novak Djokovic from Serbia notably stands out in the table, having secured the championship a record-breaking 10 times in Australian Open history. This achievement surpasses that of any other champion in the tournament's history.
- The word clouds confirm that "Novak" is the most frequently occurring name among all champions in the Australian Open history.

This bar chart compares the average win rate among top5+ champions in the Australian Open history. It displays the champions' genders, names, and countries. Each country is distinguished by a different colour, and each champion is labelled to indicate their average win rate. Annotations are included to provide additional context.

Data Insights:

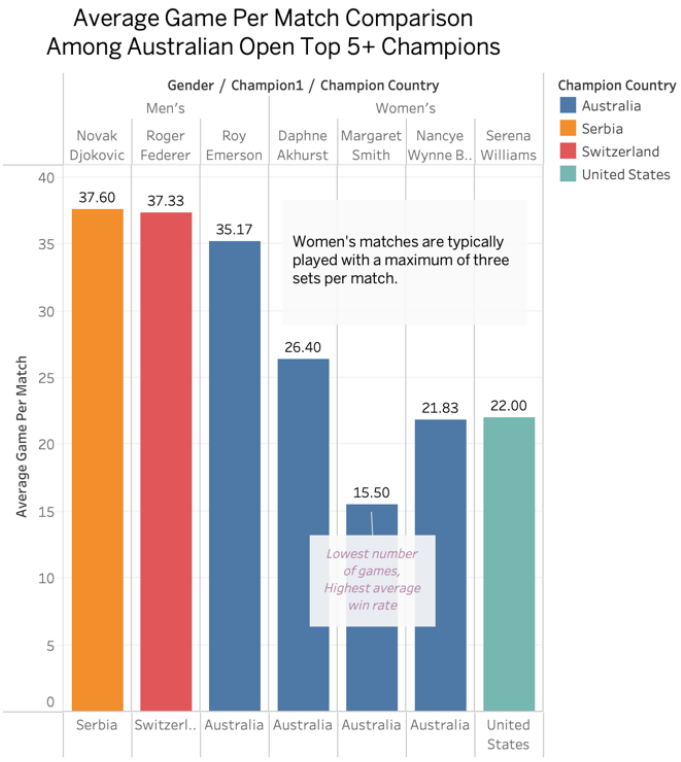
- Women have a higher win rate than men, which aligns with previous findings for all champions in Australian Open history. The lowest win rate among women exceeds the highest rate among men by 0.0209.
- Margaret from Australia holds the highest average win rate in Australian Open history at 0.7141.

Average Win Rate Comparison Among Australian Open Top 5+ Champions



Average of Win Rate for each Champion Country broken down by Gender and Champion. Color shows details about Champion Country. The marks are labeled by average of Win Rate.

The bar chart illustrates the comparison of average games per match among top5+ champions, displaying their gender, names, and countries. Each champion's average number of games per match is represented by a label, with countries distinguished by colour. Annotations provide additional insights for better storytelling.

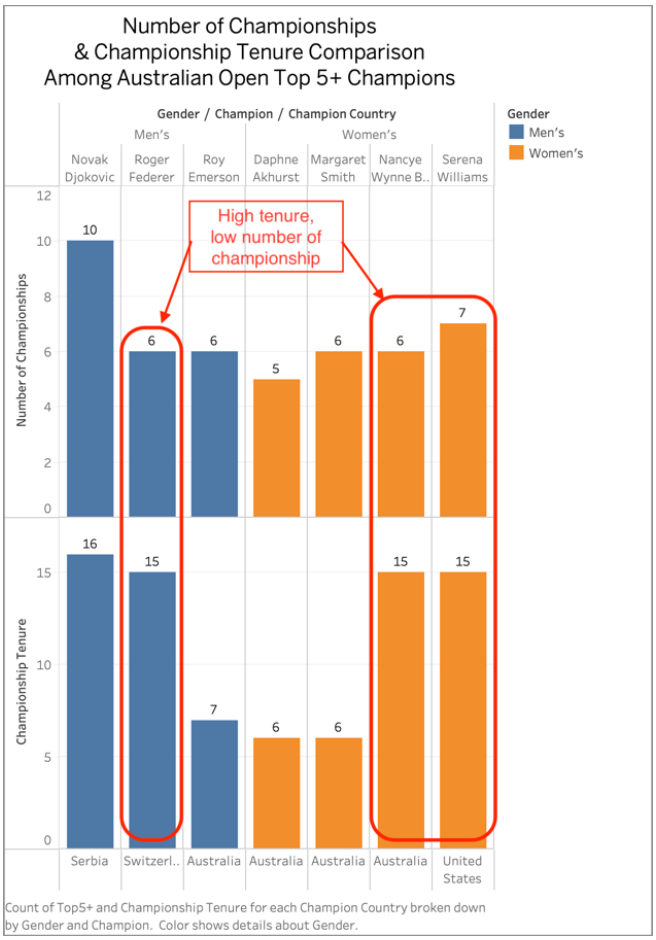


Sum of Average Game Per Match for each Champion Country broken down by Gender and Champion1. Color shows details about Champion Country. The marks are labeled by sum of Average Game Per Match.

Data Insights:

- The disparity in the number of games per match between men and women is attributed to the different set structures in tennis matches. Women typically play a maximum of three sets per match, while men compete in matches with up to five sets, resulting in a higher average number of games for men.
- The variation in the average number of games per match among men is relatively small, with only a difference of 2.43 between the highest and lowest averages.
- The average number of games per match shows significant fluctuation among women, with a difference of 10.9 between the highest and lowest averages.
- Margaret's average number of games per match is the lowest among all champions, which is consistent with her high win rate from previous analysis.

This bar chart shows the comparison of number of championship and the championship tenure among top5+ champions. It shows the champions gender, names, and country. Genders are labelled by colour and label mark has been used for better indication. Annotations are used for better story telling.

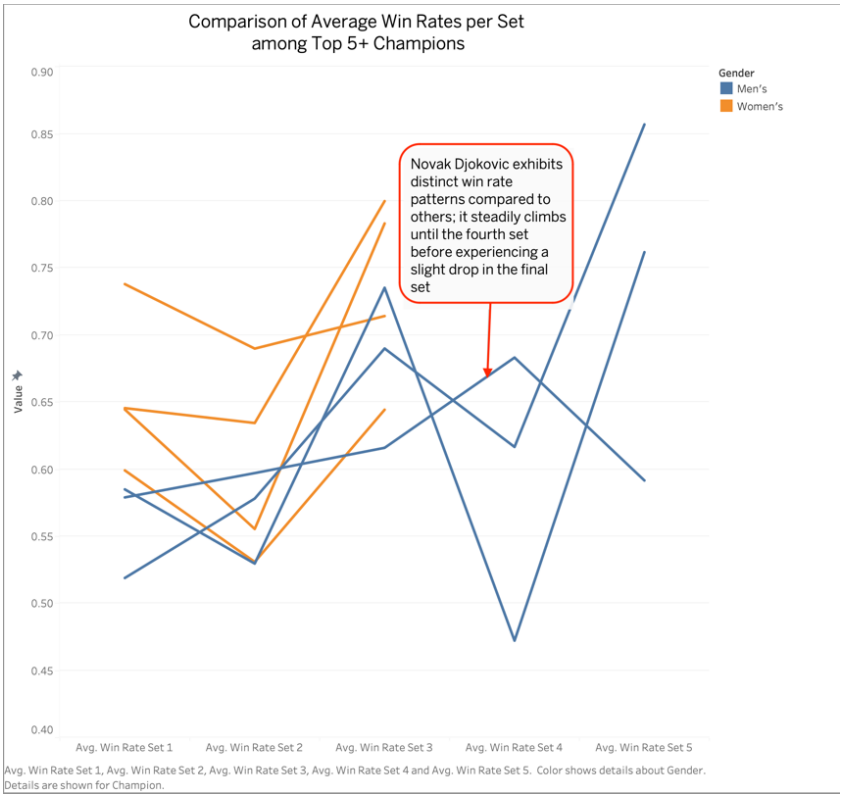


Count of Top5+ and Championship Tenure for each Champion Country broken down by Gender and Champion. Color shows details about Gender.

Data Insights:

Upon closer examination of the data, it becomes apparent that the correlation between championship tenure and the number of championships won among top5+ champions in the Australian Open history is not as straightforward as initially thought. While one might expect that longer tenure would correlate with a greater number of championships, this is not always the case. The chart highlights instances where top5+ champions had lengthy tenures of around 15 years but achieved relatively modest numbers of championships, ranging from 6 to 7 titles. This suggests that factors beyond tenure alone may influence the number of championships won by players in the Australian Open.

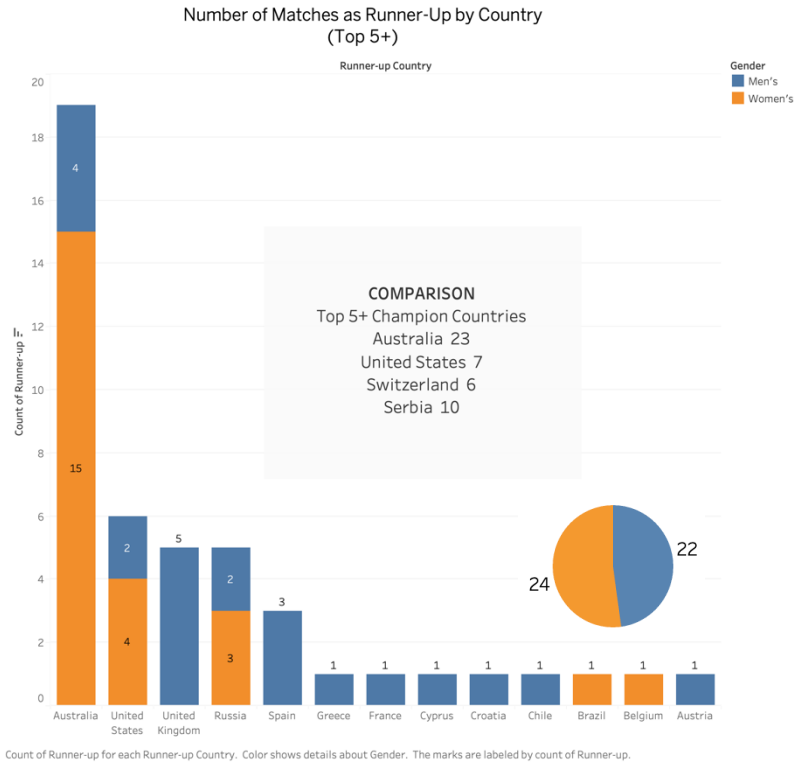
This parallel coordinate chart illustrates the patterns of average win rates per set among top5+ champions in the Australian Open history. Gender distinctions are represented by different colors for clarity. The axis has been carefully rescaled to range between 0.4 and 0.9 to facilitate easier comparison. Annotations have been added to provide additional insights into the data trends and patterns.



Data Insights:

The most intriguing observation from this chart is the unique win rate per set pattern displayed by the legendary champion Novak Djokovic. While most champions exhibit fluctuations in win rates from the first set onwards, followed by a peak in their final competitions, Novak Djokovic demonstrates a remarkable consistency. His win rate steadily climbs until the fourth set, after which it experiences a slight drop in the final competition. This suggests that Djokovic maintains exceptional stability in his performance and employs a distinct strategic approach compared to other champions.

This stacked bar chart illustrates the number of runner-up matches by country for top5+ champions. Genders are distinguished by colours, and labels provide clear indication of the data. Annotations are included to enhance storytelling. Additionally, a small pie chart in the corner summarizes the total number of matches for each gender.



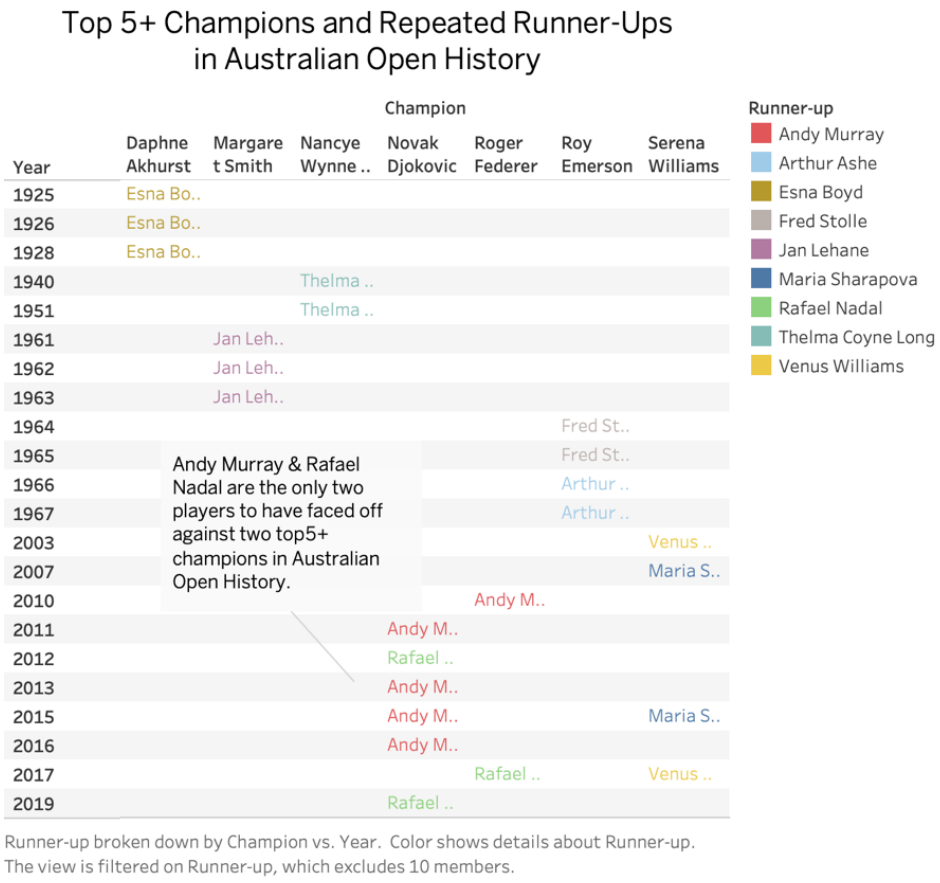
Data Insights.

- Australia maintains its dominant position in the list, with a significant number of runner-up matches.
- While the UK and Russia had 5 runner-up finishes each, neither of them is in the list of top5+ champions.
- Females maintain their leading position, with their total number of runner-up finishes surpassing males by 2.

This table displays top5+ champions alongside corresponding runner-up instances that have occurred more than once. Runner-up names are distinguished by colour, and annotations provide additional context for storytelling.

Data Insights:

- Andy Murray and Rafael Nadal are the only two players to have competed against two top5+ champions in Australian Open history.
- Andy Murray is the most frequent player in the list, having appeared five times.



Conclusion

Australian Open has experienced three distinct stages throughout its history since 1905. Before 1922, only male players participated in the tournament. Between 1922 and 1977, Australia dominated the competition. However, after 1977, Australia lost its leading position to the US, and more countries began participating in the event each year. Despite the US's dominance in the Australian Open, with the highest percentage in both the champion and runner-up lists, it only has one player listed in the top5+ category.

Overall, females have outperformed males, boasting a higher average win rate and a greater presence in the top5+. Among the top5+, four champions hail from Australia. Interestingly, the top5+ exhibit similar patterns in the average wins and losses parallel coordinate chart, contrasting with the patterns seen when analysing all champions in history. Additionally, the top5+ show different results in the analysis of championship tenure and number of championships.

Within the top5+, Novak Djokovic stands out with unique characteristics. He boasts an exceptionally high number of championships and long championship tenure. Moreover, his pattern in the average win rate per set parallel coordinate analysis differs from that of the other six champions.

Various visualization techniques were employed in this report, including treemaps, parallel coordinates, geographic maps, scatter charts, word clouds, bar charts, stacked bar charts, pie charts, and tables. Data labeling and color coding were utilized in most graphs for better clarity, and annotations were added where necessary to enhance storytelling. Axis rescaling was also implemented as needed to facilitate clearer comparisons, and trend lines were included in scatter plots for data analysis and better visualization.

While Tableau proved to be an effective tool for data analysis and visualization, offering a user-friendly interface with drag-and-drop functionality for creating interactive visualizations, it had limitations in terms of customization compared to Excel. Excel allows for greater flexibility in customizing visualizations, such as changing text colour and placing arrows where needed. Additionally, while Tableau facilitates calculations, it can be challenging to observe the entire calculation process, whereas Excel provides a more transparent view of data changes, aiding analysis in certain situations.

The utilization of various visualization techniques has greatly enhanced the readability and accessibility of the report. By presenting complex data in visually engaging formats such as treemaps, scatter charts, and parallel coordinates, readers can quickly grasp key insights and trends. The use of colour coding, data labelling, and annotations further aids comprehension, guiding readers through the analysis and highlighting important findings. These graphic techniques not only make the information more digestible but also facilitate a deeper understanding of the nuances within the data, ultimately enriching the overall narrative of the report.

References:

1.2.2 The different types of data. (2024). The University of Technology, Sydney. Retrieved 20/03, from https://canvas.uts.edu.au/courses/31371/pages/1-dot-2-2-the-different-types-of-data?module_item_id=1682737

Australian Open. (2024, 11/02/2024). Wikipedia. Retrieved 8/04, from https://en.wikipedia.org/wiki/Australian_Open#:~:text=The%20Australian%20Open%20is%20managed,and%20was%20a%20grass%20court.