

A dark blue vertical bar on the left side of the page. A blue arrow points to the right from the bar, containing the date.

4/28/2023

Assessment Task 2: Advanced Data Visualisation

Australia Open Championship

Several thin, curved lines in dark blue and light grey originate from the bottom left and curve upwards and to the right.

Vi Nguyen

13592629

CHI.V.NGUYEN@STUDENT.UTS.EDU.AU

Table of Contents

| | |
|--|----|
| EXECUTIVE SUMMARY | 2 |
| TASK 1 & 2: DATA PREPARATION AND DATA COMPREHENSION..... | 2 |
| TASK 3: WIN RATE CALCULATION..... | 3 |
| TASK 4: VISUALISATION METHOD 1. PARALLEL COORDINATE | 4 |
| TASK 5: VISUALISATION METHOD 2. TREEMAPS | 6 |
| TASK 6: VISUALISATION METHOD 3. GEOGRAPHIC MAP | 10 |
| TASK 7 & TASK 8: VISUAL ANALYTICS - CHAMPION PERFORMANCE | 12 |
| TASK 9: SUMMARY | 15 |

Executive summary

The report examines and analyses the Australia Open Championship match record for both genders over the last 118 years, from 1905 to 2023. The row header contains the 118 years arranged in reverse order from 2023 to 1905. On the other hand, there are 22 columns illustrating the characteristics of data such as gender, nationality, score, runner-up, etc.

Initially, the report will illustrate the data preparation process, including collecting, formatting and transforming data. This stage allows any audience of this report to comprehend the dataset's characteristics before proceeding to the visualisation state. Comprehending the dataset and the situation is crucial so the audience will not be confused or misinterpret the data visualisation charts.

Furthermore, a data table will be constructed to display the attribute of the datasets, the data type, the format/example, and the detailed description.

Regarding the visualisation, this report focuses on the overall pattern and trend of the entire data set, not any specific individual. Particularly, the report will analyse and compare the performance of many countries across the globe, as well as the win rate of champions across continents. Moreover, the win rate trend of all players in both genders and all countries will be plotted and compared by nearly twenty high-dimensional data visualisation graphs. The three main data visualisation techniques used in this report are Parallel Coordinate, Treemap and Geographic Mapping. In addition, the win rate of each champion was computed to analyse the winning pattern. Lastly, the top player who won from more than five champions was also analysed carefully to discover any usual pattern.

Task 1 & 2: Data Preparation and Data Comprehension

Before applying any data analysis technique, it is very crucial to prepare the data and ensure that the data is ready to use. There are few steps including collecting data, ensuring the data format meet the requirements, and data transformation.

- The Data collection step is completed and does not required any further modification.
- Regarding the format requirements, the champion's name, the country name or the nationality code are consistent in the runner-up and champion sections. The score is divided into ten separate columns for more accessible analysis. Hence, the format requirements are satisfied.

The table below identify the data type and explore each component of the data set.

| Components | Type | Format/ Example | Description |
|-------------------------|---------------------------|---|--|
| 1. Year | Quantitative: Interval | 4 digits: YYYY | Year of tournament ranging from 1905 to 2023 |
| 2. Gender | Categorical: Nominal | Women's or Men's | Distinguish between women and men tournament |
| 3. Champion | Categorical: Nominal | String | Name of the champion of that year tournament |
| 4. Champion nationality | Categorical: Nominal | String 3 capital letters E.g. USA | The country code of where the champion |
| 5. Champion country | Categorical: Nominal | String E.g. United States | Similar to component #4, but the country/nationality of the champion player are written in full name |

| | | | |
|------------------------------------|---------------------------|---------------------------------------|--|
| 6. Champion Seed | Quantitative: Ratio | N/A | The seed is determined by the ranking points of the most recent tournaments |
| 7. Mins | Quantitative: Ratio | N/A | The duration in minutes of the match |
| 8. Score | Categorical: Nominal | String E.g. 6-4 | Set of pair of integers represent the score for the players |
| 9. 1st - 5th lost 1st - 5th win | Quantitative: Interval | Numeric 0 - 12 | The score results of the players split into 10 columns |
| 10. Runner-up | Categorical: Nominal | String | Runner up of tournament that year |
| 11. Runner-up nationality | Categorical: Nominal | String - 3 capital letter E.g, USA | The nationality of runner-up player |
| 12. Runner-up Country | Categorical: Nominal | String E.g. United States | Similar to component #11, but the country/nationality of the runner-up player are written in full name |
| 13. Runner-up seed | Quantitative: Ratio | N/A | The seed is determined by the ranking points of the most recent tournaments |

Task 3: Win rate calculation

Step 1: Sum up all the score results from 1st won to 5th won into a new column called Win.

Step 2: Sum up all the score results from 1st lost to 5th lost into a new column called Lost.

Step 3: For each champion of the year, Win rate = Win / (Win + Lost)

Example

Step 1: Sum up 1st won, 2nd won, and 3rd won give the total of 16

Step 2: Sum up 1st lost, 2nd lost, and 3rd lost give the total of 13

Step 3: Win rate = $16 / (16 + 13) = 55.17\%$

| Year | Champion | Champion Nationality | Win rate | Score | Win | Lose | 1st won(%) | 2nd-won(%) | 1st-won | 1st-loss | 2nd-won | 2nd-loss | 3rd-won | 3rd-loss | 4th-won | 4th-loss |
|------|-----------------|----------------------|----------|-------------------------|-----|------|------------|------------|---------|----------|---------|----------|---------|----------|---------|----------|
| 2023 | Aryna Sabalenka | BLR | 55.17% | 4-6, 6-3, 6-4 | 16 | 13 | 36% | 50% | 4 | 6 | 6 | 3 | 6 | 4 | | |
| 2023 | Novak Djokovic | SRB | 57.14% | 6-3, 7-6(7-4), 7-6(7-5) | 20 | 15 | 55% | 58% | 6 | 3 | 7 | 6 | 7 | 6 | | |

Unusual data

1. Walkover: There is a case where all the scores of each match are, and the string Score data = Walkover in 1966, where the Champion is Margaret Smith. Walkover is a common term in tennis where one player withdraws from the championship match before the match starts or does not complete all the required sets. All the relevant data will be left blank except for the string "Score."

| | A | B | C | D | E | F | G | H | I | J |
|-----|------|----------------|----------------------|----------|----------|-----|------|------------|------------|----------|
| 1 | Year | Champion | Champion Nationality | Win rate | Score | Win | Lose | 1st won(%) | 2nd-won(%) | 1st-loss |
| 116 | 1966 | Margaret Smith | AUS | | walkover | | | | | |

2. Retire: The 1965 champion Margaret Smith had the 4th set score = retired. This mean that her opponent, or the Runner-up Mario Bueno had retired at the 4th set. The data is already being handled, where the score of the 4th set is removed.

| Year | Champion | Champion Nationality | Win rate | Score | Win | Lose | 1st won(%) | 2nd-won(%) | 1st-won | 1st-loss | 2nd-won | 2nd-loss | 3rd-won | 3rd-loss | 4th-won | 4th-loss |
|------|----------------|----------------------|----------|------------------------|-----|------|------------|------------|---------|----------|---------|----------|---------|----------|---------|----------|
| 1965 | Margaret Smith | AUS | 55.17% | 5-7, 6-4, 5-2, retired | 16 | 13 | 45% | 50% | 5 | 7 | 6 | 4 | 5 | 2 | | |

3. Retire [k]: The score of 1990 has the word 5-2 retired[k]. This is because the tennis player retired during the match, while the champion have not scored 6 yet. This could possibly be due to an injury during the match. The score for win and loss are still recorded.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|------|-------|------------|-----|----------------|-------------------------------|---|---|---|---|---|---|---|---|
| 1990 | Men's | Ivan Lendl | TCH | Czechoslovakia | 4-6, 7-6(7-3), 5-2 retired[k] | 4 | 6 | 7 | 6 | 5 | 2 | | |

4. Score after 6-6. In order to win, the player has to score 6 and be 2 scores in advance. If both score 6, the player must play one extra set called 'Tie breaker', and whoever scores seven first wins. The tie breaker winner will have a set score of 7 while the opponent's score is 6. The score of the tie breaker will also be recorded in the string "Score" in the parentheses, e.g. (7-2) in 2022. However, the score in each column (1st won, 1st-loss...) will only record the result of the first set -> set score.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|------|---------|--------------------|-------------|-------------|---------------|------|-------------------------|---------|----------|---------|----------|---------|----------|
| Year | Gender | Champion | Champion Na | Champion Co | Champion Seed | Mins | Score | 1st-won | 1st-loss | 2nd-won | 2nd-loss | 3rd-won | 3rd-loss |
| 2023 | Women's | Aryna Sabalen BLR | Belarus | | | | 4-6, 6-3, 6-4 | 4 | 6 | 6 | 3 | 6 | 4 |
| 2023 | Men's | Novak Djokovic SRB | Serbia | | | | 6-3, 7-6(7-4), 7-6(7-5) | 6 | 3 | 7 | 6 | 7 | 6 |
| 2022 | Women's | Ashleigh Barty AUS | Australia | | | | 6-3, 7-6(7-2) | 6 | 3 | 7 | 6 | | |

Interesting Findings

Women champion only have to win 2 sets (2-0 or 2-1) in order to win the match while men's champion require to win 3 set to win the game.

Task 4: Visualisation Method 1. Parallel Coordinate

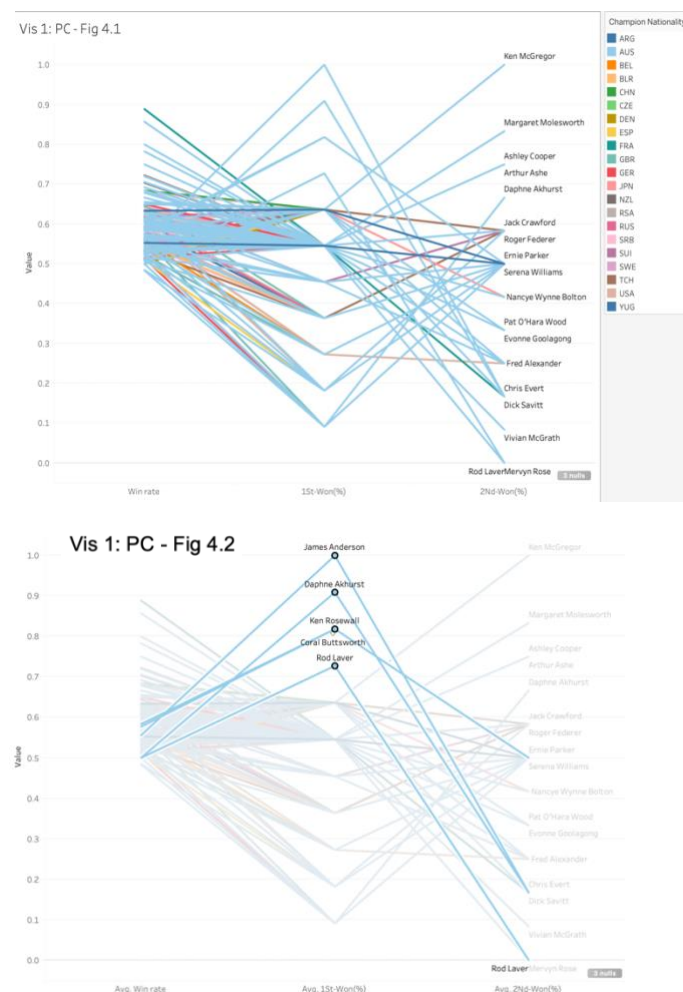
The measure value included 1st-won, 2nd-won and Win rate because these measures are in number. The marks include champion, score and nationality since these variable types are qualitative. However, since the win ratio ranges from 0 - 1 (0% - 100%), but the scores are from 0 - 11 (for 1st-won) and 0-12 (for 2nd-won), it would be more meaningful to standardise the parallel coordinate and convert the 1st-won and 2nd-won scores to percentages based on the highest value.

NOTE: 0.55 in 1st-won mean the champion won 6 game in the 2nd match while 0.5 mean in 2nd-won mean the champion won 6 games.

Note: PC stand for Parallel Coordinate

4.1 High score (Fig 4.2)

As shown in figure 4.2, there are five Australian Champions with a medium Win rate (0.5 -0.6) and have a very high score for their first win. However, this is not necessarily a good thing to have a high score in tennis. This means that the champion takes more effort to win a set. He/she can win 6 games but was not able to score 2 games higher than his /her component and beat them quickly. This denoted that the runner up's ability is not essentially lower than the champion and they can immediately gain back the score. In figure 4.1, there are 5 outliers. The champion with a lower win rate (50 - 55%) that takes more effort to win the first set have a higher chance to loses the second set. With a slightly higher rate (around 58%), Ken Rosewall and Coral Buttsworth were able to score 6 points and win in the second set.

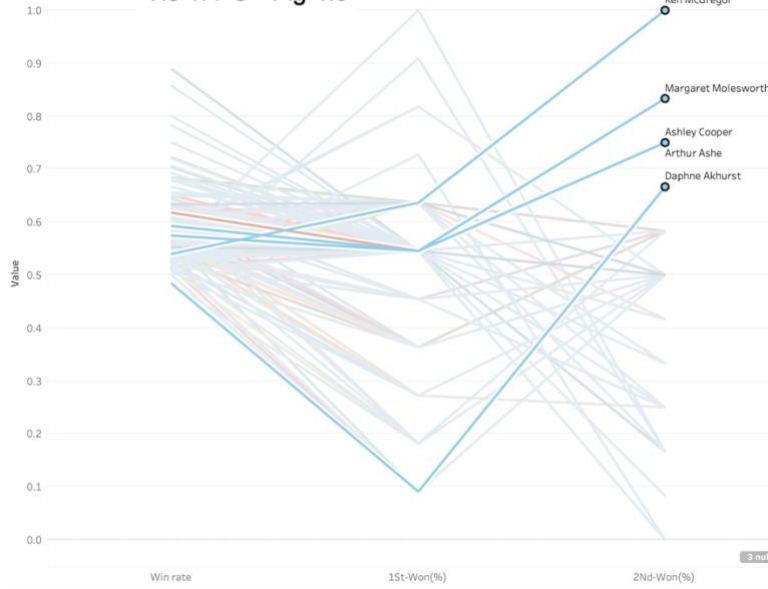


Similarly, the champion who won the 2nd game with a much more effort (higher score) tends to have an average to low Win rate as well. For example, the lowest win ratio champion we can observe from the figure 4.3 is Daphne who lost the first set, winning only one game.

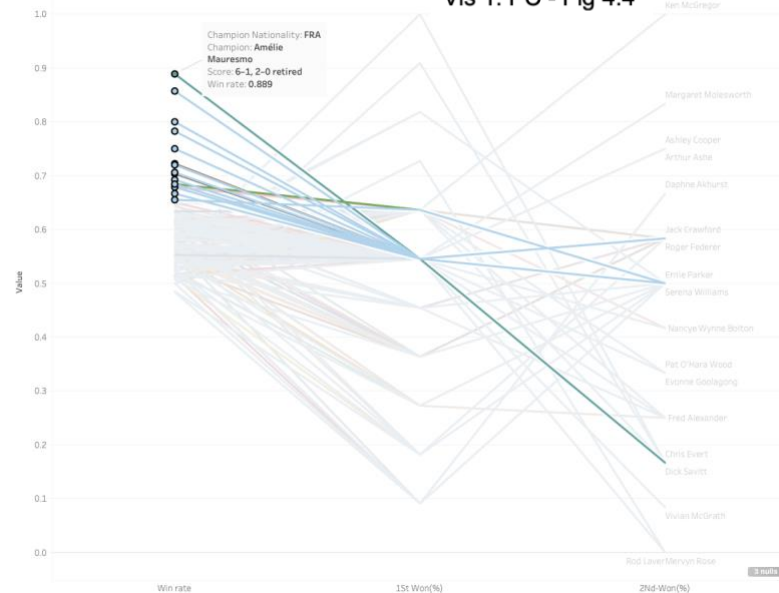
4.2 High Win rate (Fig 4.3 and 4.4)

Win rate denoted the percentage of won games compared to the total game a player plays. Higher wins signal that a player tends to win more games in an entire match, and potentially have a chance to win more sets to become a champion.

Vis 1: PC - Fig 4.3



Vis 1: PC - Fig 4.4



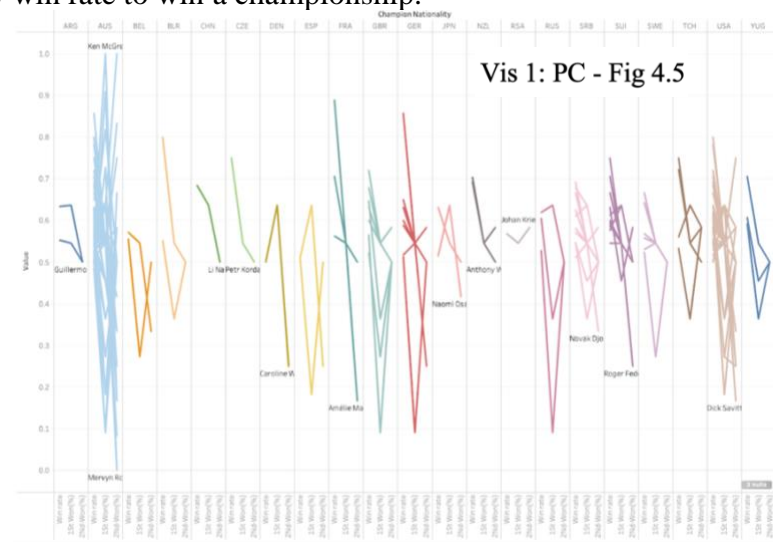
The highest ratio goes to Amélia from France, as shown in Figure 4.4. However, this is because she only played two sets. In the first set, she won 6 games; in the second set, she won 2. Her opponent won one game in the first set, then retired after Amélia won 2 games in the second set. Hence, this is the reason for her high win rate.

Apart from Amélia, all the champions with a win rate above 65% tend to win the first two consecutive games, as shown in Figure 4.4. The Win rate below 65% varies significantly and has no stories to tell. Furthermore, a tennis player must have at least a 50% win rate to win a championship.

4.3 Champion Nationality (Fig 4.5)

According to figure 4.5, the majority of the champions come from Australia and USA. England, Germany, and Serbia are the next group of countries that have a lot of champions. There are five countries that only have one champion including Argentina (ARG), China (CHN), Czechia (CZE), New Zealand (NZL) and Russia (RUS).

Vis 1: PC - Fig 4.5



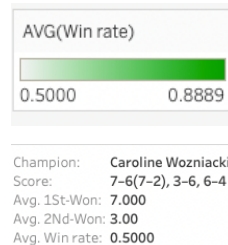
Task 5: Visualisation Method 2. Treemaps

Treemap is a very helpful visualisation technique that utilise the set of rectangles and the colour difference to visualise data from different categories. In this case, we want to examine the relationship between different champions, their nationality, gender, score and win rate.

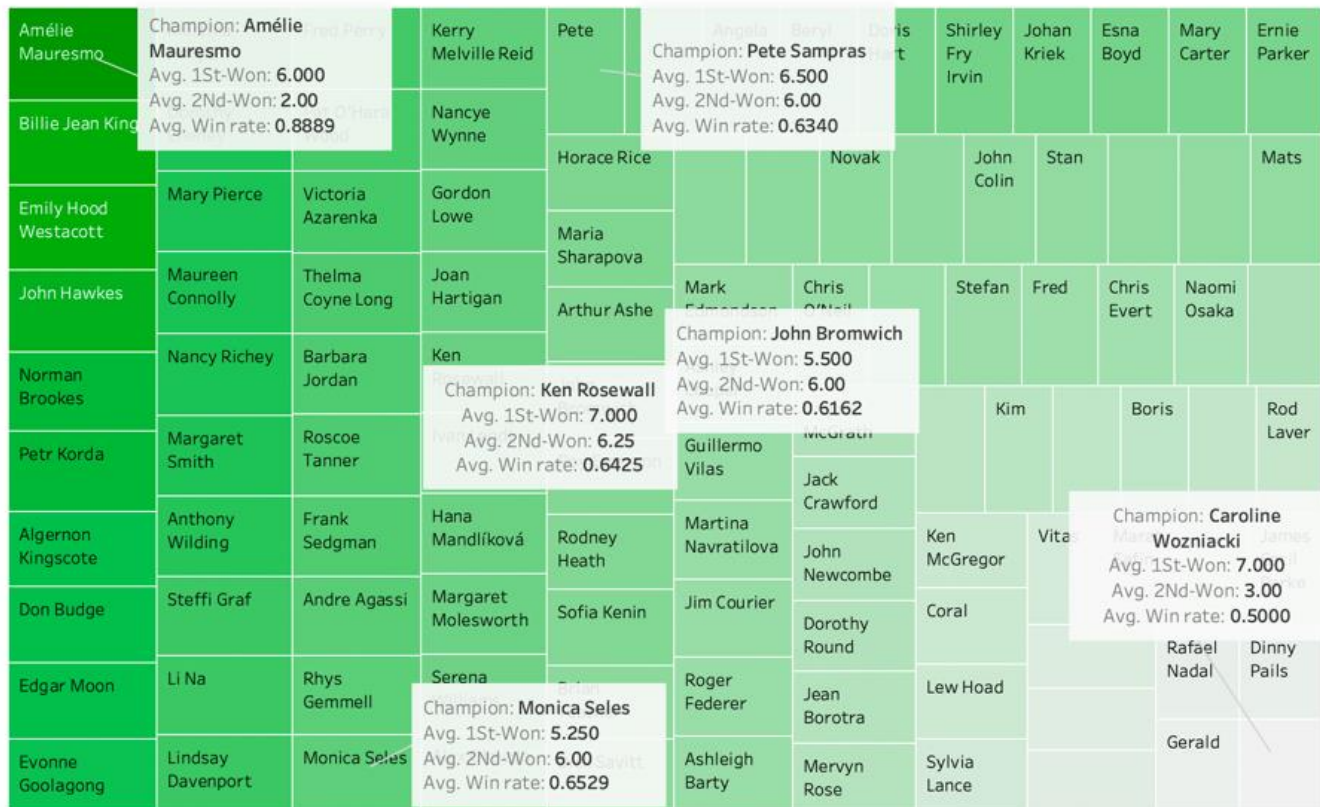
5.1 Champion win rate (Fig 5.1)

The win rate is critical, so placing it in both the size and colour marks would be appropriate. The darker the green colour and the bigger the size of the rectangles, the higher the win rate is. For instance, the highest win rate goes to Amelia Mauresmo at 0.8889, with the largest rectangle and darkest green on the screen. On the other hand, the lowest win ratio goes to Caroline Wozniacki at 0.5. The lowest win ratio had to be 0.5 as the champion must balance the amount of game loss at least and the game won in order to win the final match. With the equal game won and lost, the champion had to win at least one tier breaker in order to win the final match(as explained above).

The largest and the darkest green rectangle are placed on the left and the top, while the lowest win rate is located at the far right bottom of the tree map. Among the middle is the normal green colour, with the win rate ranging from 0.6 - 0.7.



Vis 2: Tree map - Fig 5.1



5.2 Champion nationality and win rate (Fig 5.2)

After that, we should also look at the win rate of different nationalities win rate. Let's change the colour mark from average win rate to Champion nationality and remove the champion as detail to see only the country's average win rate.

The top win rate goes to Czechia with a 75% win rate, and both the score in 1st won and 2nd won are 6. The other big top win rate goes to France, New Zealand and China. This means the champions must

have scored very high at the later set (3rd, 4th and 5th). The country with the lowest win rate is Denmark and Spain.

A positive correlation exists between the win rate and the average 1st won and 2nd won score. Moreover, as the gap between the average 1st-won and 2nd-won enlarges, the win rate is often quite low. E.g. DEN average 1st-won: 7; average 2nd-won: 3 -> win rate = 0.5, similarly for UK and Russia. The gap in the win score between 2 sets raises a concern about the ability to control the game of the champions. France, in this case, is quite unusual where their average 2nd-won is nearly 25% lower than their average first-won.

Vis 2: Treemap - Fig 5.2



Regardless, they might have some strategy to maintain a stable win toward the later set.

5.3 Win rate, champion nationality and gender (Fig 5.3)

After considering gender as a variable, we can see that Czechia is now ranked 2nd in the hall of fame for the win rate, and all of its champions are male. On the other hand, the female tennis players from France take the lead with a 79.74% winning rate, denoting that the French male tennis players dragged this rate down by nearly 8%. Clearly, French women can handle everything by themselves without help from men.

Overall, the differences between male and female champions vary quite significantly except for Australia, America and TCH. These three countries have quite balanced average win rates between men and women sitting at the middle of the tree map ranging from 0.62 - 0.65.

A few countries do not even have a men's championship such as China or YUG. On the other hand, some nations do not have any women's champion, like New Zealand, Spain or Serbia.

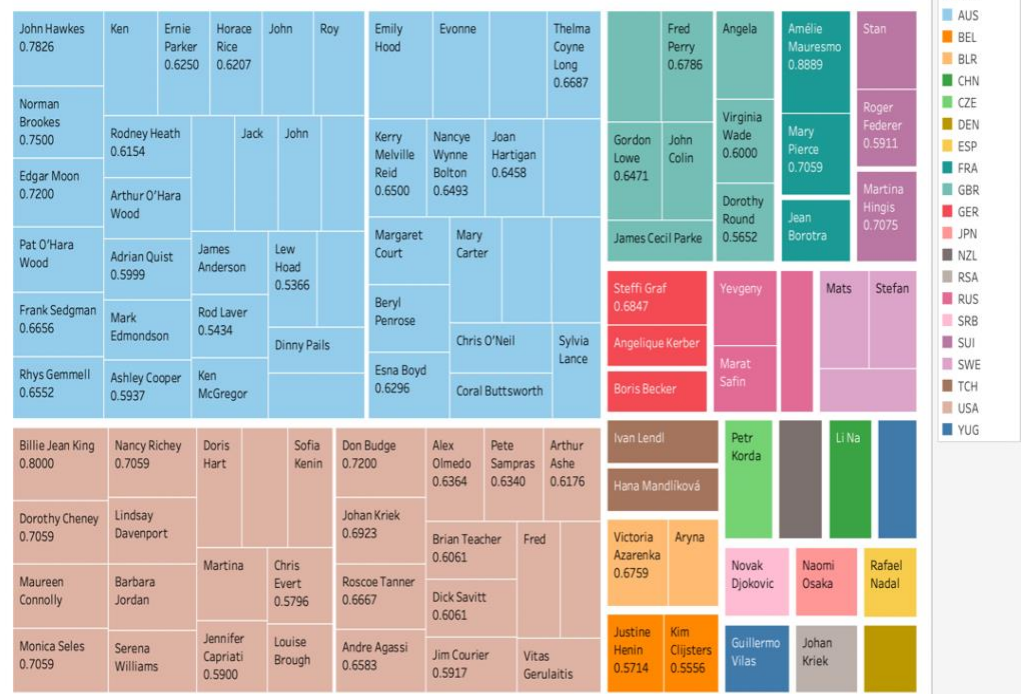
Vis 2: Treemap - Fig 5.3



5.4 Champion nationality, gender, win rate and champion

Finally, we put everything into place for the final step in the hierarchy. We can see that most champions come from America and Australia. The numbers of champions from these two countries are pretty balanced, which we could not observe from the Parallel Coordinate in the previous visualisation. The next large group of champions came from England (GBR) with eight champions. Most commonly, one country only has one champion with 11 countries.

Vis 2: Tree map - Fig 5.4

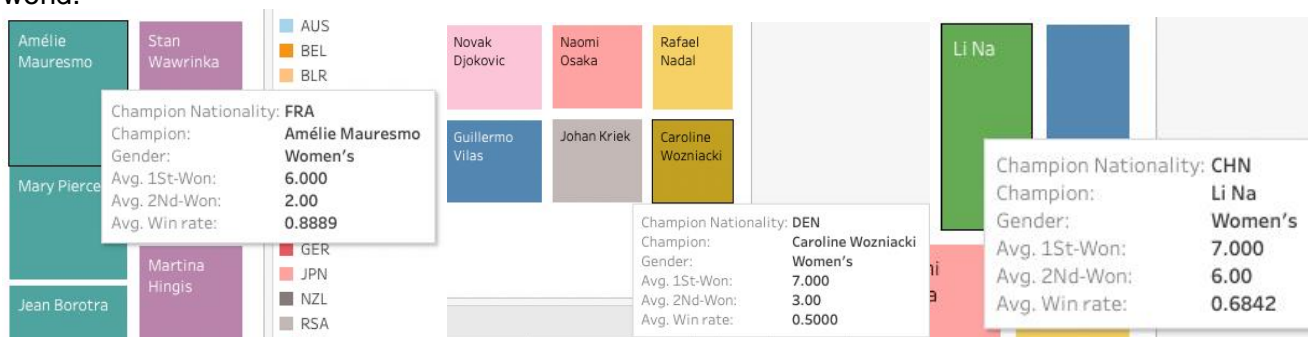


Unlike other sports like soccer or football, where male is the dominant sex, we can observe that the ratio between male and female is quite balanced in the top-winning countries like Australia & America.

Interesting finding

Now that we have a complete view of the data set, there are a few things that are quite interesting. The win rate follows the normal distribution, which means that as the sample size gets larger, the sample mean (or average) tends to move closer to the population mean. Countries like Australia and America, regardless of male or female, all have the average win rate in the middle of this dataset. Sitting on the two tails are countries with just a few champions.

For instance: one of the lowest win rate (0.5) goes to Caroline Wozniacki from Denmark, and this win rate determines the entire nation's win rate since there is only one champion from this nation. On the other hand, on the right tail of the distribution, we have Amelie Mauresmo coming from France with the win rate of 0.8889. France has the 2nd highest average win rate for the nation. If Amelie Mauresmo's nationality changed to Australia or America, her winning rate would not greatly impact this nation's average win rate. Or, Lina from China determines the win rate of the largest population country in the world.

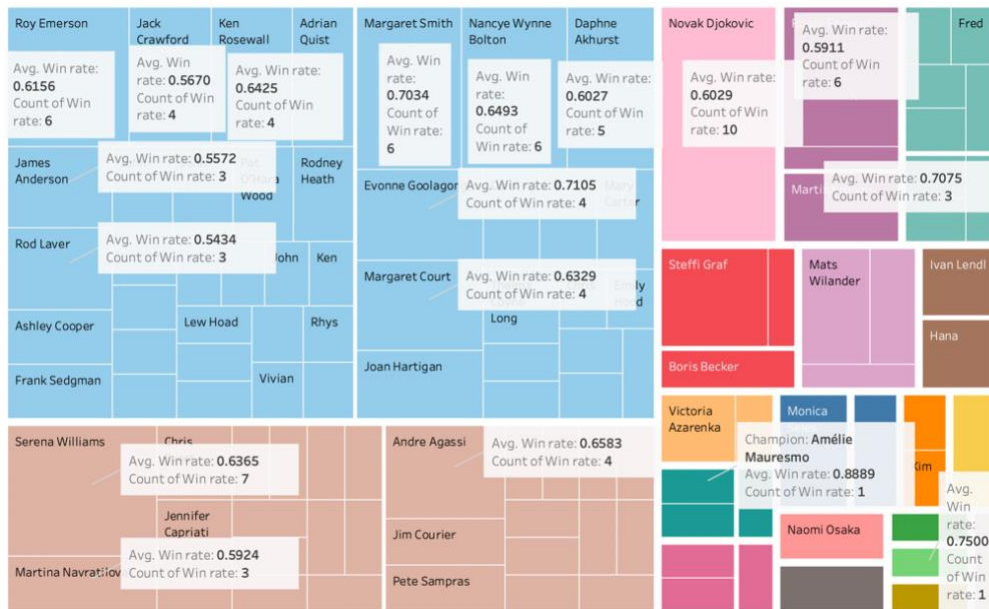


5.5 The contribution of the champion toward the win rate

Acknowledging that the champion can compete many times, we must also consider this factor. Let's change the size to the count of win rate while keeping champion nationality as colour and gender as detail. A few champions' continuous participation contributes quite significantly toward the win rate of their nation, while some drag it down dramatically.

For instance, Australia's average win rate (AUS) equals 0.604 -> Jack Crawford, Rod Laver, and James Anderson drag down the Win rate of Australia with an average win rate lower than 0.6. On the other hand, Ken Rosewall with a win rate of 0.6425, or Evonne Goolagong with an average win rate of 0.7105 lift the average win rate of Australia quite significantly.

Vis 2: Tree map - Fig 5.5



5.6 Win rate vs win count

However, there is a slight negative correlation between the champions' win rate and the count of times that they win. Naturally, keeping up with the younger people is more challenging as these champions age. These champions can beat the younger players with experience, but their strength and flexibility decrease over time. Hence, they will lose more games and decrease their win rate over time.

In Figure 5.6, we can see that the top win rate champion only participated once. Others won around 2-4 times. There are a few exceptions, such as Margaret Smith, who won 6 times, but it is quite rare.

Vis 2: Tree map - Fig 5.6



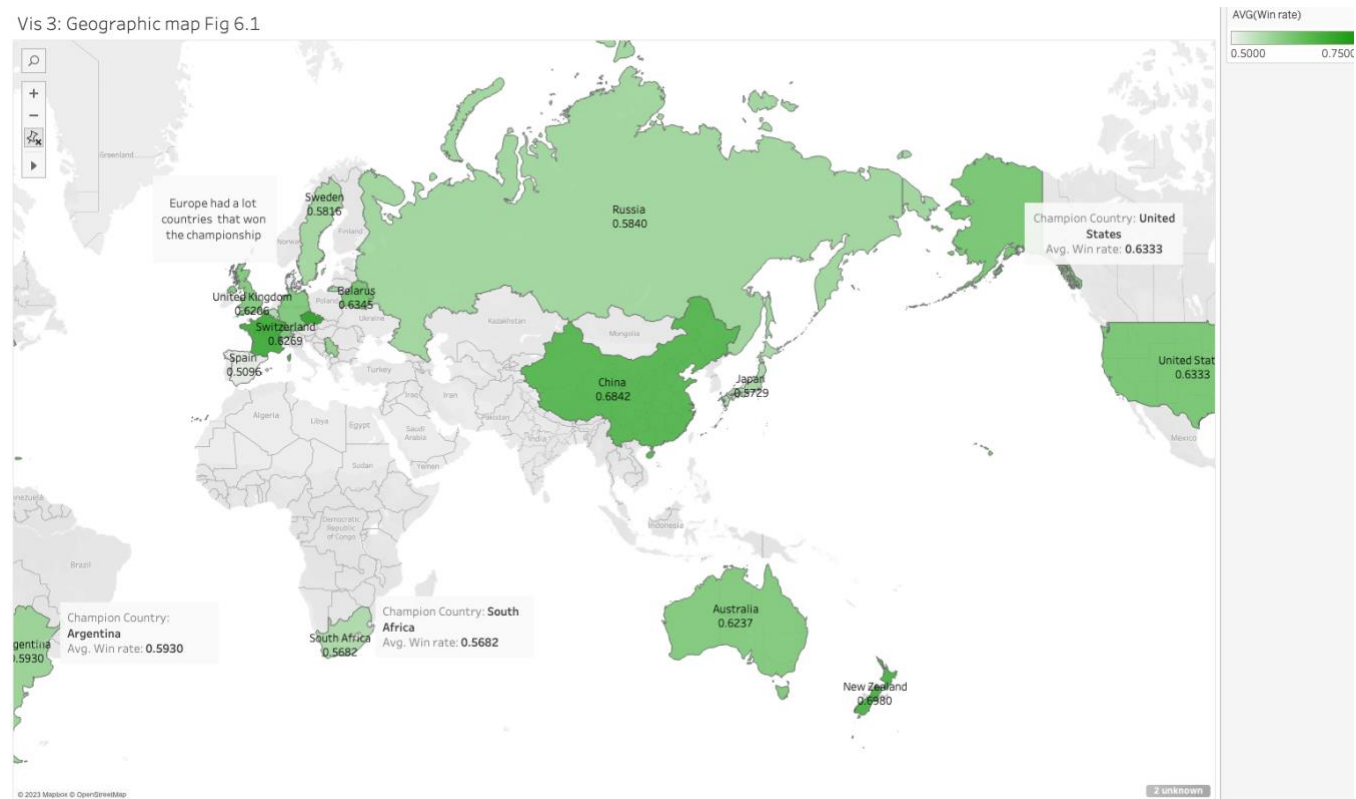
Visualisation Method 3. Geographic Map: Task 6.

Problem: There are two countries that the Geographical map needs help to recognise. After researching, it turns out that Czechoslovakia only exists before 1992. After that, this country split into the Czech Republic (Czechia) and Slovakia. Yugoslavia has the same situation where this country is also being broken up into several countries: Bosnia and Herzegovina, Croatia, Macedonia, Montenegro, Serbia, and Slovenia. **Solution:** Ignore it

6.1 Champion nationality and win rate (Fig 6.1)

The higher the win rate, the darker the blue and vice versa. Looking at Fig 6.1, we can see an unequal distribution of the country that won the championship worldwide. The continents that contain the most champion countries are Europe, containing the UK, Belarus, Sweden, Spain, Serbia, Switzerland (label did not display on screen), France (label did not display on screen) as well as Russia, even though this country is also a part of Asia. Despite having a lot of winning countries, the win rate in these European countries is sitting at an average and low level. (Spain had the lowest winning ratio of 0.5096).

Vis 3: Geographic map Fig 6.1

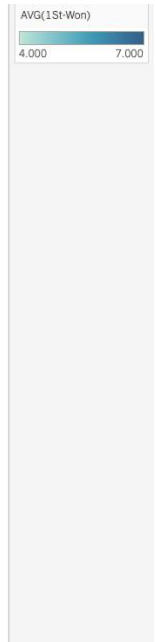


Asia is the 'runner-up' continent with China, Japan, and Russia. Oceania has two representatives, Australia and New Zealand. There are three continents, with only one country that has a champion: the US from North America, Argentina from South America, and South Africa from Africa.

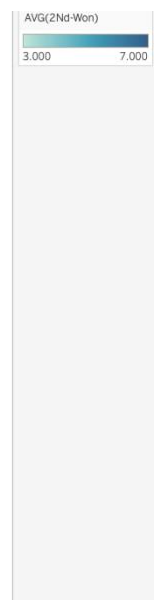
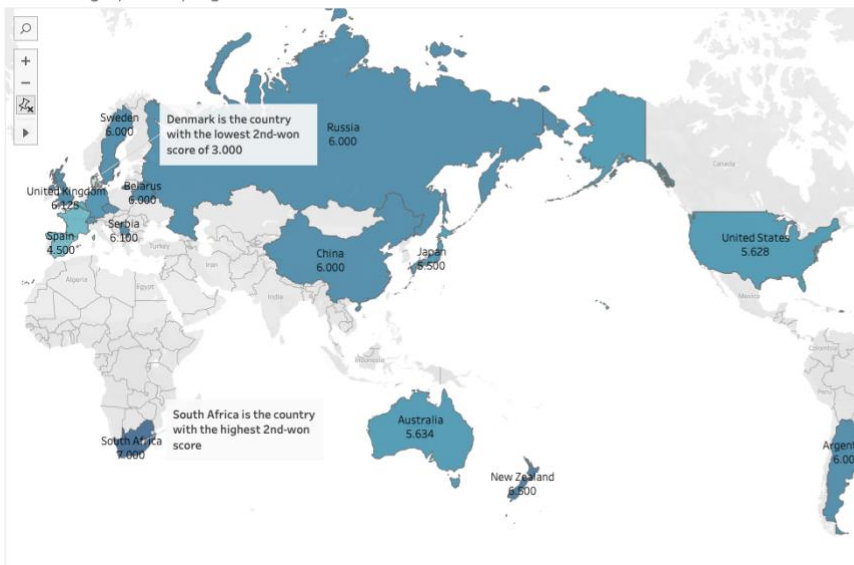
6.2 Champion nationality and the score (Fig 6.2 and 6.3)

According to fig 6.2 Russia have one of the lowest average 1st-win of 4.00 and China has the highest with 7.00. On the other hand, the country with the lowest 1st-win goes to Spain of 4.5 and highest goes to South Africa at 7.0.

Vis 3: Geographic map Fig 6.2



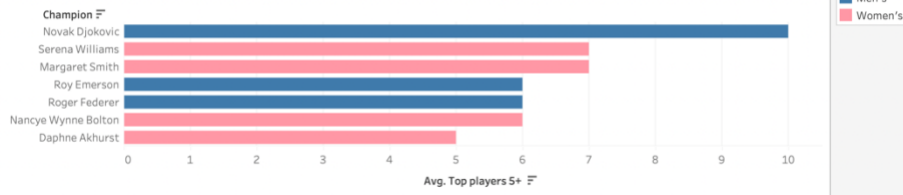
Vis 3: Geographic map Fig 6.3



Task 7 & Task 8: Visual Analytics - Champion performance

Seven champions have won more than five times, whereas Novak Djokovic won the most with ten championships. There are four women and three male champions, according to Fig 6.1.

Vis Analytics: Top player Fig 7.1



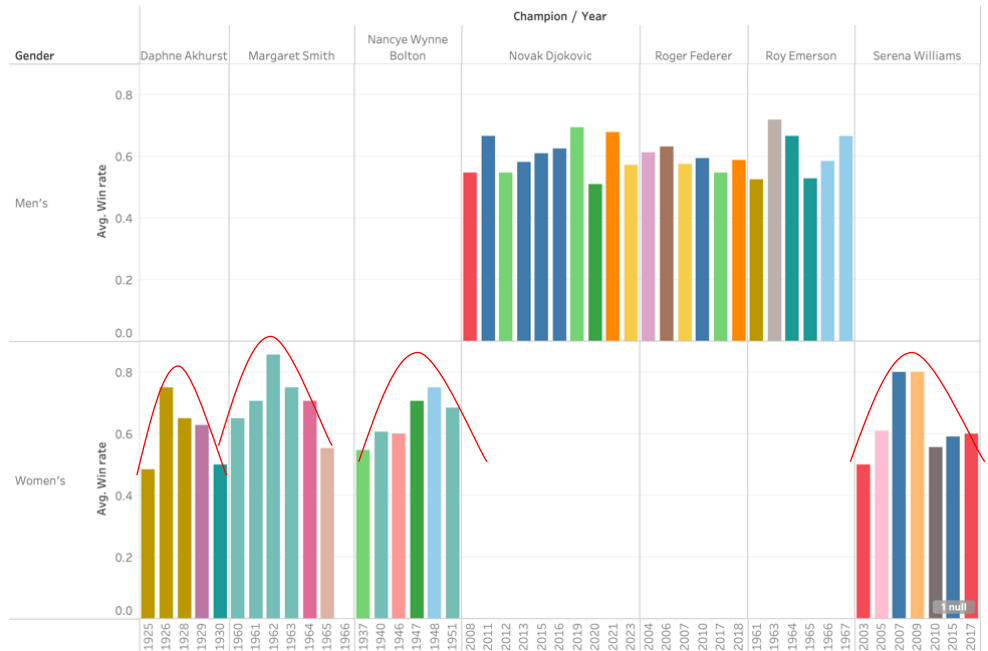
7.1 Top player, runner up, year and win rate: Bell shape win rate

For women's top players, we can observe that their win rate has a bell shape pattern. Specifically, their win rate initiates quite low, peaks in the middle, and gradually decreases as the year progresses. On the other hand, men's top players had quite a stable win rate over many years.

There are many hypothesis that can explain this situation:

1. Naturally, men have more advantage in physical strength compared to women.
2. In the long run, men tend to have better endurance. Women may had to give birth or suffered monthly from the menstrual period can significantly decrease their body's strength

Vis Analytics: Top player Fig 6.2

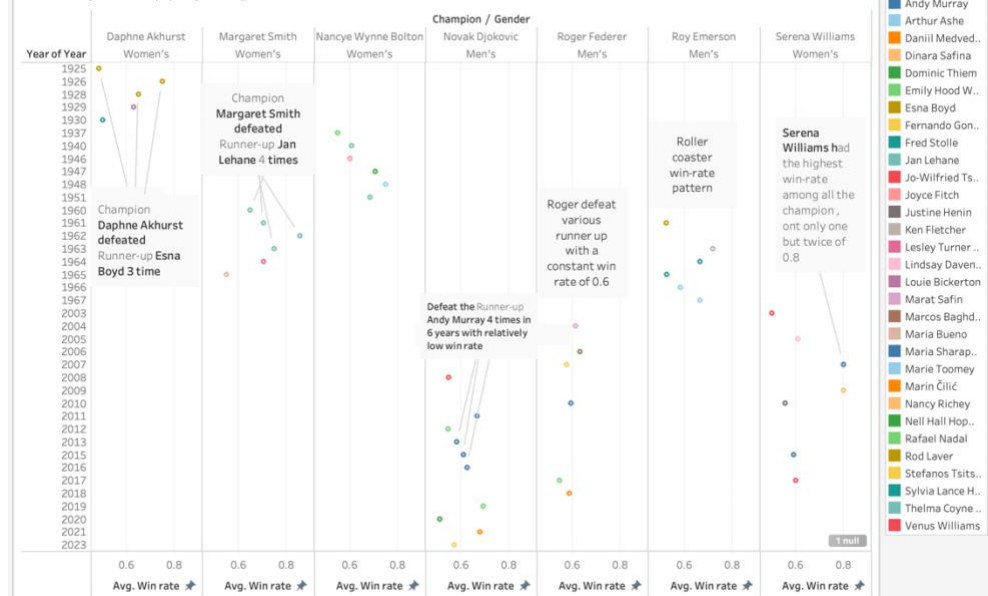


7.2 Period of playing and winning of champion

However, we can observe something different in Figure 7.3. Even though the top women players in the past tended to hold their championship for a short period, men and women after 2000 tend to play much longer. In the past, there were four top-players champions, but the period that they held their championship is quite short :

- Women: Daphne Akhurst won 5 championships in 6 years (1925 - 1930)
- Women: Margaret Smith also won continuously 6 times but also last for 6 years (1960 - 1965).
- Men: Roy Emerson won 6 times in 7 years (1961 - 1967)
- Nancy Wynne Bolton last for 14 years (1937 - 1951) - an outlier to this theory

Vis Analytics: Top player Fig 7.3



However, after the year 2000, both men and women show a much longer win period. For instance, Novak Djokovic won 10 times in 12 year (2012 -2023), or Roger win 6 times in 15 years (2004 - 2018), and Serina William win 7 times in 15 years (2003 -1017).

This incident can be because nowadays, the sports industry has a lot of medical support, advanced technology as well as financial support for the athlete that allows them to play much longer.

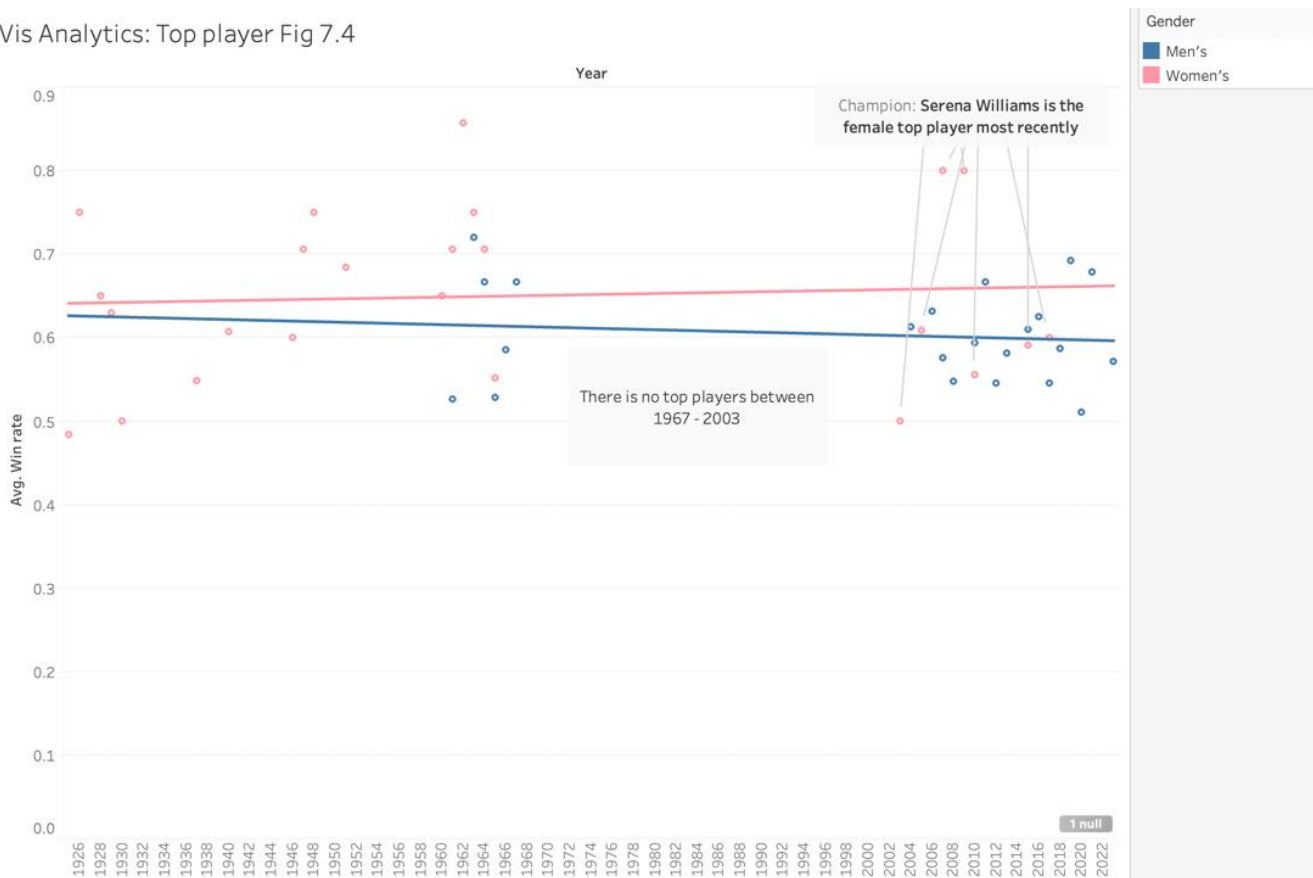
7.3 Top players vs runner up:

Runner-up are the rank second player in that tournament, who played against the champion in the final match. We can observe a repetition of runner-up in Fig 7.3. where Esna Boyd was defeated three times by Daphne Akhurst in the final match. Or Andy Murray was defeated four times by Novak Djokovic.

7.4 Win rate of men and women overtime

Women's top player win rate starts at a higher position than men's (0.64) and increases over the year to roughly 0.67. Men's top player win rate started at around 0.63 and declined under 0.6 recently. Looking at Fig .74, we can see that no top players were born between 1967 and 2003.

Vis Analytics: Top player Fig 7.4



Task 9: Summary

9.1 Dealing with high-dimensional data visualisation

My approach to dealing with high dimensional data is approaching it very slowly and unstacking the variables individually. Furthermore, setting a specific goal is also very important. Having a specific goal allows the data analyst to select the different attributes correctly to put on the data visualisation. Regardless, setting a good and precise goal requires the data analyst to understand the dataset comprehensively. Last but not least is the choice of a suitable data visualisation technique.

Unstacking the data set gradually allows the data analyst to discover many meaningful stories. For instance, in the Treemap in task 5, I wanted to find out the winner with the highest win rate, the nation with the highest win rate, and then the nation with the highest win rate sorted by gender. Finally, I want to find the connection between all the listed variables, including nationality, gender as well as champion in one graph. Quickly putting all the required variables into the Treemap will not allow the visualisation to recognise the player with the highest win ratio or the country with the highest win ratio.

Figuring out a goal is very important for a data analyst as it would determine the path and destination of the data visualisation. In task 7, for example, I have a specific idea that there must be some difference between the win rate of men's top players and women. Moreover, there must be some change in the win rate over the years, as well as the period that the champion wins the tournament. Hence, this led me to input suitable variables and measurements to figure out the corresponding visual pattern. Goal setting requires a data analyst to comprehend and learn the data set very carefully. For instance, the tennis scoring system was like a new language to anyone who had not played tennis before. Learning from the online tutorial and understanding the mechanic behind it is time-consuming. Furthermore, the difference between the country's name in the Geography map section requires the data analyst to have some fundamental knowledge about the country name to prepare and alter the data correctly.

After all, choosing the correct data visualisation is the final essential step. Each data visualisation technique will have its advantages and disadvantages. For instance, the parallel coordinate is the perfect solution to find the correlation between the numerical values and assist data analysts in identifying trends and patterns. However, it may not be a good idea to visualise categorical variables like gender or nationality. On the other hand, Treemap is the perfect answer for categorical variables as this technique allows data analysts to divide the categorical value into the hierarchical level of a rectangle box (leaf). The limitation of this method is that it can only work with a limited amount of numerical data (the size of the box and the colour of the box can represent the value of the numerical value). A Geographic map is an excellent method to visualise geographical measures like cities, states, countries, and continents, but only a limited amount of numerical data can be presented.

9.2 Describe graphic attribute designs and labelling techniques

Parallel Coordinate: Parallel Coordinate requires only a few labelling techniques because this data visualisation mainly aims to draw the relationship between numerical values. The score from 1st Won and 2nd Won was rescaled to show the correlation between the three variables more transparently. The outliers were labelled in figure 3.2 and figure 3.3. Regarding the graphic attribute, each nationality has different colours making much overlap. It is hard to see the dominant champion country. Hence, an additional Parallel coordinate was created to visualise the win rate of each country easier. Regarding graphic attributes, each country was displayed using a different colour.

Treemap: Treemap required much labelling. Regardless, only the essential label was displayed. The focus of the visualisation is the win rate. Hence, among the five treemaps, the label of the win rate is always displayed. The 1st-won and 2nd-won scores are displayed when comparing and analysing different nations. From Figure 4.2 - 4.5, the champion nationality was selected as the colour to display

different countries. In Figure 4.6, the darker the colour represents the more time the champion had won the championship.

Geographic map: The name of the country, as well as the win rate, was displayed. The intensity of the green colour represents the country with a higher win rate. The label and score of the country with the lowest average and highest average score were displayed.

Visual pattern: Figure 6.3 displays the time the winner defeated the same runner-up. Different coloured are used to distinguish between different runner-up players. The year was also displayed as the timeline to show the year and period the champion won. The label in Fig 6.4 also showed the gap between 1967 and 2003.

In conclusion, only the essential and relevant labels were presented to tell the story. The display of data labels is crucial, but too much can lead to unnecessary confusion for the audience. Irrelevant data labels can distract the audience and convey the wrong message.

9.3 Trend Highlights

Figure 3.4 highlights the high win rate, demonstrating that the champion with a win rate above 65% tends to win two consecutive first sets. Figure 6.4 displays the trendline to show the trend of men's and women's win rates from 1926 -2022.

9.4 Conclusion on top player's performance

After analysing and visualising the performance of the top players, a few interesting insights was finalised. First, there is bell shape pattern in the win rate of the women's top player while the win rate of the male champion fluctuates stably at the low rate from 0.5 - 0.6. Moreover, the period that the top player played in win had increased significantly by 100% - 200% after one century. The runner-up of the tournament has a high chance of winning the tournament where the champion's win rate is low. There is a repetition in the players of the final match where the champion defeats the runner-up several time. The win rate of women's top players increased quite remarkably while the men's win rate decreased.

9.5 Summarise the advantages of Tableau or other visualisation apps you have used.

Tableau is an effective and powerful tool for visualising multi-dimensional data. Overall, the application is very user-friendly and convenient in many different ways. More specifically:

- **Free-code application:** Allow data analytics from different backgrounds lacking coding skills to start on data visualisation.
- **Navigation:** Easy to navigate between different menus, options and different types of graphs using the built-in Dashboard.
- **Data manipulation:** The ease of changing data attribute in columns and rows make this tool extremely helpful when the data analyst need a specific goal to analyse the trend.
- **User-friendly interface:** The interface is similar to Microsoft Excel, allowing beginners to navigate easily. The navigation bar at the bottom allows the user to switch between worksheets effortlessly.
- **Quick graph Alteration:** "Show Me" option allows users to change the type of visualisation in a matter of seconds. This option is beneficial when the data analyst is miserable about the appropriate graph to use (in Task 7, for instance).
- **Mark option:** The mark option allows the user to define the characteristics of the graph with the choice of colour, size, label, detail, tooltip and path. Just one alteration can change the visualisation completely.
- **Data source variety:** Tableau allows users to input the data from various data sources such as Excel, web-based, SQL, etc. This characteristic is valuable in this time of Big Data when the data analyst does not have a choice of the form and format of data.
- **Quick and easy data label:** Most of the labels in Tableau are coded, and the user just needs to enable it to show on screen. On the other hand, Excel requires users to input the graph's label.
- **Data alteration within the application:** Tableau allows users to change, add, and extract information from the data source without affecting the original file.