

## Executive Summary

This report analyses the Australian Open Tennis Champions dataset covering matches between 1905-2024. Harnessing Tableau we identify key trends and performance comparisons across genders, individual players and historical periods. Tableau is especially effective at quickly developing geographic maps, scatter plots and tree maps. These charts facilitated effective story-telling and pattern recognition as they are particularly relevant to the type of attributes in the dataset, such as location data. This allowed for in-depth analysis of top player (those with five or more championships) performance, win rates and match dynamics. Visual clarity and storytelling were further enhanced through intuitive annotations and labels. Challenges encountered include addressing dataset anomalies like missing values and ambiguous data points (e.g., “walkover” match records). Furthermore, visualizing dense data like performance scores required careful graph design, employing axis rescaling and grid adjustments to improve readability and clarity.

The primary trends reveal distinct differences between men and women, specifically men’s matches are typically significantly longer, and champion seeding correlates to greater dominance with lower seeds. Despite this, both genders have remarkably similar win rate trends, revealing top player’s all show early dominance in matches. Additionally, top players’ average seedings and performance over time indicated a slight increase in competitiveness, due to the spread of knowledge and advanced training methods. Australia dominates the championship count due to historical factors, however after the 1908s Europe and the USA have outperformed Australia.

Top-players are specifically analysed for their unique abilities. All players showed a sustained dominance, most notably Djokovic and Court outclassed all top winners with double digit championship titles. Djokovic has the potential to surpass Margaret Court’s status as the greatest of all time with her 11 championships. The top players excel with distinct strengths over other players. Djokovic thrives under pressure with unmatched mental resilience, while Federer is known for his versatility and composure. Serena Williams dominates with aggression and toughness, Court for her adaptability, and Emerson for his endurance. Bolton and Akhurst are successful from their mental resilience and strategic consistency. Together, they represent the pinnacle of tennis excellence through their unique abilities.

## 1. Introduction

This report outlines the key information and trends found in Australian Open Tennis championships matches. We analyse the relevant data in the 2024-Spring-Ass2-v15.xlsx dataset which contains matches dating back to 1905. By effectively graphing the data, we outline key findings and trends for both the game and key individual players. This includes Australia’s dominance and Margaret Court’s unmatched 11 championships.

## 2. Data Collection and Preparation

To enable data analysis, we first collect and prepare the data. We download the 2024-Spring-Ass2-v15 dataset from an online repository and save it as a xlsx file. This file is stored on a cloud drive with version control. Each row represents a specific year, and each column represents an attribute. The data possess a column/row format as we are using excel. Subsequently, the data must be formatted so that it is in the correct category (Column). This is ensured during data entry. There are several abnormalities in the dataset that must be addressed to prepare for graphing and analysis in Tableau. Tableau is selected as it is more suited to handling high dimensional data, such as sporting match histories. We make changes in a sperate sheet which references the original dataset, following best practices to ensure the raw data is not lost and updates to the original data is reflected in subsequent calculations. This ensures data consistency.

Firstly, in 1977 there were two championships for each gender, in the original data these years are labelled 1977-(1) and 1977-(2). This formatting prevents analysis tools from recognising these values as years, hence we remove the annotations. Likewise, the year column is originally formatted as text, we change this to a number format to enable graphing. Once this has been done, Tableau is successfully able to recognise the attribute formats automatically to enable relevant graphing. For example, the Year is recognised as numerical

values and the Champion Country is recognised as a Geographic Location. Additionally, the score attribute possesses abnormalities. In the 2024 Men's game the score is separated by “.” instead of “,” characters which breaks calculations, this has been fixed.

Calculated values are used to extract additional insights and information. Firstly, we calculate the win rate of the Champion and Runner-Up for the Nth set in each match using the following formulas:

$$\text{Champion } N^{\text{th}} \text{ Rate} = N^{\text{th}}\text{-Won} / (N^{\text{th}}\text{-Won} + N^{\text{th}}\text{-Loss})$$

$$\text{Runner-up } N^{\text{th}} \text{ Rate} = N^{\text{th}}\text{-Loss} / (N^{\text{th}}\text{-Won} + N^{\text{th}}\text{-Loss})$$

Subsequently, we calculate the average win rate for each championship player across each set. First, we must calculate the average across all Runner-up and Champion games for a player separately:

$$\text{Champion } N^{\text{th}} \text{ Rate Avg} = (\text{Champion } N^{\text{th}} \text{ Rate}_1 + \dots + \text{Champion } N^{\text{th}} \text{ Rate}_i) / i$$

$$\text{Runner-up } N^{\text{th}} \text{ Rate Avg} = (\text{Runner-up } N^{\text{th}} \text{ Rate}_1 + \dots + \text{Runner-up } N^{\text{th}} \text{ Rate}_i) / i$$

Next, we calculate the average between these two values for the same player, to get their overall win rate. We also calculate the years since debut according to the following formula:

$$\text{Years Since Debut} = \text{Current Year} - \text{Debut Year}$$

Additionally, we calculate the average champion seed according to the following formula:

$$\text{Champion } x\text{'s Avg Seed} = (\text{Champion } x\text{'s Seed}_1 + \dots + \text{Champion } x\text{'s Rate}_i) / i$$

### 3. Dataset Characteristics and Data Types

#### 3.1 Data Quality

First, we evaluate the relevant dimensions of data quality. The dataset is incomplete as it has several missing values in the ‘Mins’ and ‘Seed’ attributes, however the data is sufficiently complete to support analysis. Specifically, Seeds before 1924 and Mins before 2024 are missing. A note is added to the raw data sheet explaining these missing values. The timeliness of the dataset is sufficient as it has entries for the current year, 2024. The dataset is accurate as it complies with Australian Open records and real-world results. The validity cannot be evaluated as we do not know how the data was obtained; however, the provider is a reputable university. The dataset captures detailed match statistics making it precise. Additionally, the dataset has high accessibility due to its widely accessible excel format and standardised data point formatting. The volume, or quantity, of the data is relatively small, hence it does not fall into the category of big data. This enables analysis of the data to be more manageable and handled by simple analysis tools. The dataset contains numerical values and text which encodes names and geographic locations, making it sufficiently diverse. This simplifies the tools necessary to analyse the data, enabling the use of conventional techniques. Specifically, we collect the following structured internet data attributes for the years 1905-2024 (See Appendix A). We define the type with the following key: Quantitative interval-scale (I), Categorical Ordinal (O), Categorical Nominal (N), Quantitative ratio-scale (R).

#### 3.2 Abnormal Values

Figure 1 was designed such that 1-1 correlations between champion country and champion nationality are on the centre diagonal; this addressed the challenge of intuitively highlighting any uncorrelated values. Additionally, we circle and annotate these data points with red line art to highlight their abnormal nature. Although it is possible for there to be a disparity between champion nationality and country, when, for example, a champion's citizenship is different to their place of birth, in practice countries seem to appoint their tennis champions based on their nationality. The “CAE” nationality does not refer to any currently recognised country code and is likely an error in the dataset that should be “CZE”, as all champions with the “CAE” code are of Czech descent. This has been addressed during graphing.

Champion Country	ARG	AUS	BEL	BLR	CZE	CHN	DEN	FRA	GER	ITA	JPN	NZL	RUS	SRB	RSA	ESP	SWE	SUI	GBR	USA	CAE	YUG
Argentina	2																					
Australia		94																				
Belgium			2																			
Belarus				4																		
Czechia					1																	
China						1																
Denmark							1															
France								3														
Germany									7													
Italy										1												
Japan											2											
New Zealand												2										
Russia													3									
Serbia														10								
South Africa															1							
Spain																2						
Sweden																	6					
Switzerland																		10				
United Kingdom																			8			
United States																					43	

Figure 1

There are several outlier and abnormal values in the dataset. This includes the 1990's Men's game which ends with Stefan Edberg retiring, and a score ending in "retired[k]". Retired refers to when a player voluntarily withdraws during a match, likely due to an injury, illness or fatigue making them unable to continue. This results in the opponent winning by default. The "k" may refer to a key or additional information to provide context on why the player retired. Likewise, in the 2006 Women's game Mario Bueno retired, there is no key in the scoreline. A Walkover is when a player is unable to start a match at all, and the opponent wins by default. 1966 was a controversial year where Nancy Richey forfeited the championship with a walkover.

Interestingly, the 1993 Women's champion country was Yugoslavia which is no longer recognised by tableau, we mapped this entry to its successor country Serbia to ensure the data point is still visualised. This explains why there are disparities between the champion country and champion nationality for "Serbia" and "YUG".

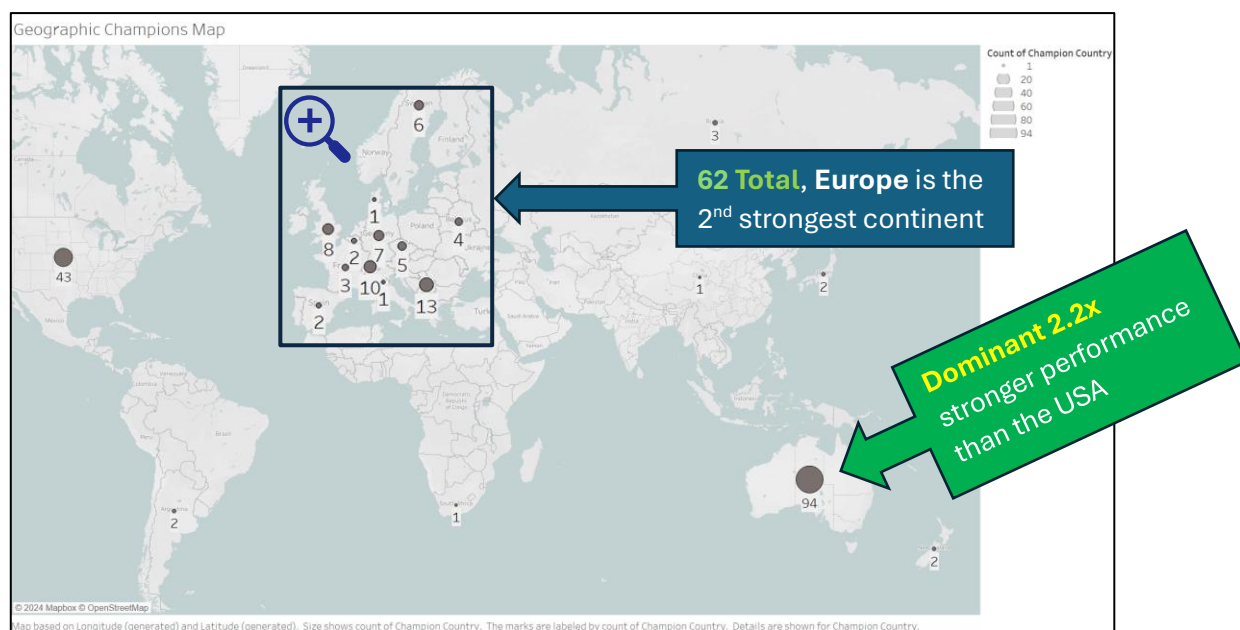
Additionally, the Champion Seed is a numerical value representing the placement of champion players in the tournament, however the attribute also contains the character value "U" for some data points. Seeding is a system used by tournament organizers to rank the top players based on their world rankings, recent performance, and sometimes past performance in the specific tournament. The purpose of seeding is to ensure that the best players are spread out in the tournament draw, reducing the likelihood that top-ranked players face each other in the early rounds. We take "U" to refer to unseeded, meaning that the player did not receive a ranking position in the draw, as a result they can be placed anywhere in the bracket. This typically occurs when a player is of lower ranking in the world or does not have a strong performance for the tournament.

## 4. Data Visualisation

### 4.1 Geographic Champions Map

Here we plot the location of the championship countries in the data, we chose a geographic map as it clearly conveys locations in an intuitive and standardised manner. Tableau recognises two outlier data points, Australia and the United States, which have an abnormally high number of championship titles. We represent the quantity of championships by the size of the circular plot. This enables quick and intuitive visual comparison while drawing the viewers' attention to the most prominent countries. However, to ensure clarity we add data labels to clarify the exact numerical differences. This was a challenge as it can be harder to differentiate smaller counts of champions. The labels have a white background to differentiate them from the country borders. The colour of the circular plot was chosen to contrast with the background, which used

intentionally muted colours. A border was added to more distinctly separate the circles from the country borders. A description is also included underneath the chart and a key is included to the right to contextualise the plot and scale. The map was intentionally zoomed in to maximise readability while hiding irrelevant landmasses like Antarctica. This simplifies interpretation of the chart.



### Figure 2

Australia has the disproportionately largest number of champions on the map. This is expected as Australia is renowned for its tennis and hosts the Australian Open. This could suggest a home advantage, as when players compete in their home country, they have the confidence gained with a supportive crowd. However, this dominance is mainly in earlier years of the tournament preceding the 1980's. This could also be due to less international participation in earlier years. Individual European countries have negligible counts compared to Australia, but when considered as a continent Europe has the next strongest performance of 62. This grouping was highlighted by zooming on Europe to group and increase legibility of smaller datapoints. Switzerland's and Serbia's strong performance can be explained by Roger Federer and Novak Djokovic's dominant performances in Australian Open titles. This is followed by the United States of America reflecting its long-lasting historical strength in tennis. Other continents have extremely low counts indicating their preference for other sports which fosters less champions. We highlight these findings with colour coded annotations and bolded text.

## 4.2 Champion Tree Map

A tree map was employed here to clearly convey the number of championships won by each champion; this is clarified by the title. We selected to group all champions with only one title as plotting them individually reduced the readability of the chart. Readability of the chart and data labels was a significant challenge that required exploration and graph resizing. This choice also reveals the impressive and relative significance of repeat champions compared to the many one-time successes. The champion's name and count of championships were annotated to ensure clarity. Additionally, tooltips were enabled to ensure that obstructed labels like Aryna Sabalenka were viewable. The colour gradient was chosen to maximise readability and highlight numerical differences intuitively. A key was added to explain this gradient.



Figure 3

From this plot we find Margaret Court has the most wins at 11. Novak Djokovic is a close second at 10 wins with the potential to overtake Court as he is still actively competing. Serena Williams, Roger Federer, Roy Emerson, Nancye Wynne Bolton and Daphne Akhurst are all notable performers too. This visualisation reveals 5 as a sufficient threshold to differentiate top-winners from standard champions, as there are many more champions with 4 or less titles. The “One Off Champions” category is a large block of 64 players who have only won a single title. This shows that while many players have had singular victories, only a small number of champions have achieved repeated success. This highlights how impressive repeat champions are and reveals their unmatched dominance and mastery of their craft. There is a notable group of mid-range champions with 2-4 wins. These findings are highlighted by colour coded and bolded annotations.

### 4.3 Win Rate

A scatter plot was selected here to convey the average win rate of players over the duration of the sport in championship matches. Scatter plots are effective for showing relationships between two variables, especially time, with many data points. The data points are depicted as open circles so that overlapping data points can be viewed more clearly.

The plot reveals a relatively stable win rate which slightly decreases over the 1905-2024 period. This implies that the game has become slightly more competitive as knowledge of optimal strategy and approaches has become more widely available and implemented. We highlight this finding with a trend line and annotation for the equation of the trend line. The annotation is coloured blue to show its correspondence to the data points. Likewise, we add a line art arrow that describes this trend. Additionally, red clustering boxes with a colour coded annotation reveals an increasingly frequent and consistent debut frequency of champions which can be explained by the increasing popularity of the tournament. A difficulty with plotting this chart included ensuring visibility of data points and trends, this was ensured by rescaling the horizontal and vertical axis such that the data points filled most of the view. Further, horizontal and vertical grid lines were made more visible to convey the location of data points and provide context to the trend line.

The chart also reveals three outlier values, Alfred Beamish, Sofia Kenin and Dominic Thiem. To highlight these values, we add annotations with a contrasting colour and line to link them to the corresponding labelled data point. To ensure readability, the font size of these annotations was adjusted. These outlier values can be explained by random chance as these players do not have enough championship matches to make these measurements a reliable indication of their unmatched performance, instead they had some notable sets each which affected the average measurement.

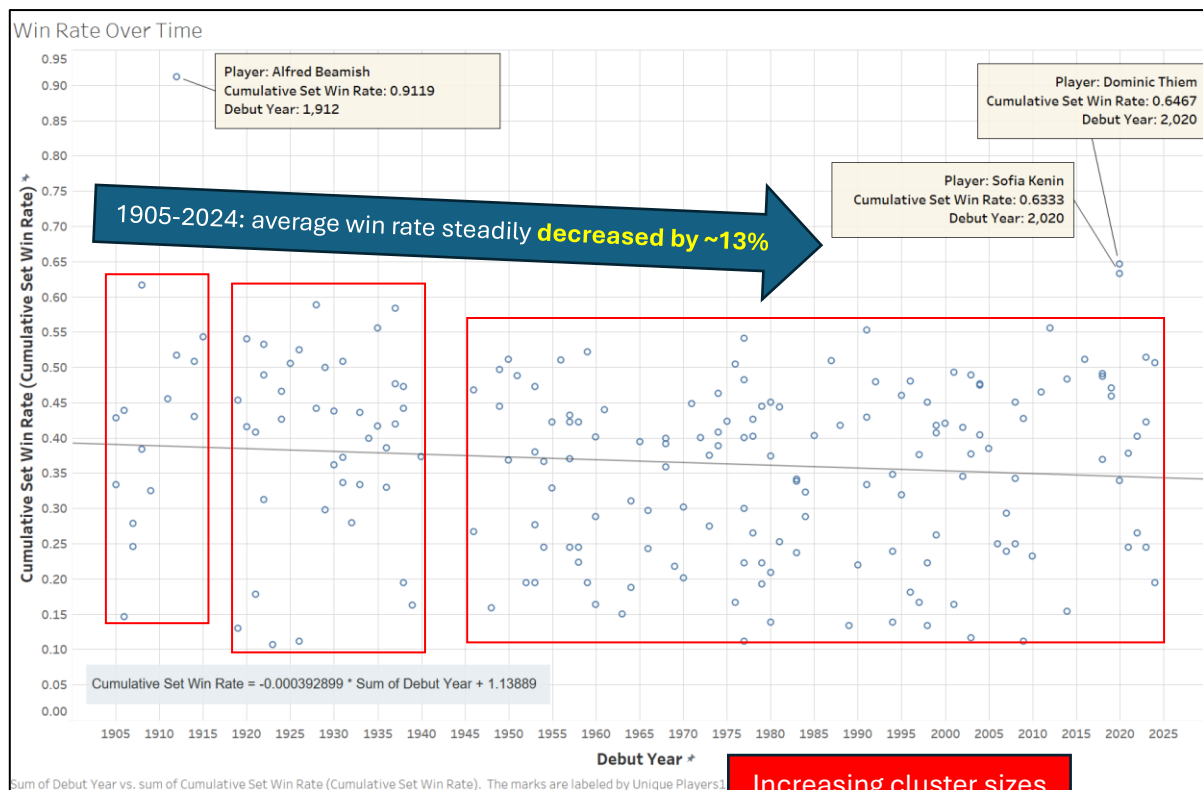


Figure 4

Increasing cluster sizes  
as champions debut  
more frequently

#### 4.4 Sets Per Match

This scatter plot intuitively visualises the number of sets per match over time from 1905-2024, comparing both men's and women's tennis matches. This provides a good proxy for match length as the "Mins" attribute is missing. Scatter plots are effective at differentiating and clustering many data points over time.

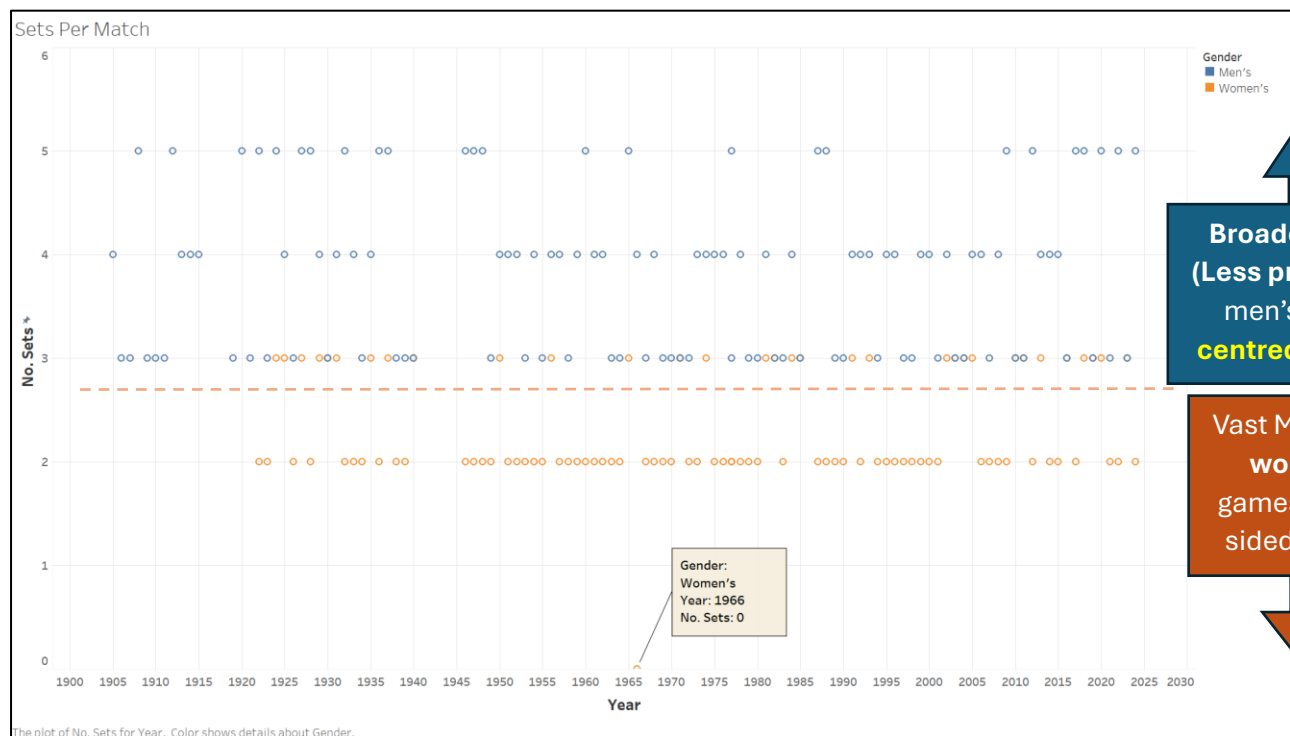


Figure 5

We do not find matches to have gotten longer, indicating the competitiveness of the championship matchups across the sports history has remained relatively stable. The data is colour coded by gender and a key is provided to visually differentiate the data points for easier visualisation of trends. From this we find that Men's games typically have 4 sets. Women's matches almost always feature 3 sets, interestingly, Men's matches have a greater spread of sets compared to women's, ranging from 3-5 sets. We can explain these findings by the rule structure of tennis, where men's matches are played in a best-of-5 format, while women's are played in a best-of-3 format. We highlight this separation of set totals by adding a dashed line to separate the men's and women's scorelines, using a dashed format to not confuse the line for a trend line. Colour coded text boxes also emphasise and explain this finding on the chart.

A challenge was visibility of the location of data points. This was addressed by making the grid lines more visible with a darker colour. By doing so we note an outlier in the 1966 Women's game which had zero sets. This is annotated with a colour coded tooltip to highlight this finding. This outlier game is explained by a walkover which prevented the match from being played. The inclusion of this data point automatically made tableau include negative set values on the vertical axis. The axis was rescaled to avoid displaying impossible values to reduce confusion and increase the size of the used chart area, ultimately increasing readability.

#### 4.5 Champion Seeds

We employ a box plot as it clearly conveys the distribution of seeds for male and female champions. We differentiate men and women with distinct colours and a key to improve readability. To further improve readability axis rescaling was included to ensure the chart maximises view space. Additionally, data points were not plotted within the box plot, so that only outliers had circular plots on the chart to differentiate them from the standard range, which was a challenge with the standard formatting. Further, grid lines and attribute labels were included to address the difficulty of reading values on the extremes of the chart. Likewise, axis labels and a title are included to provide context.

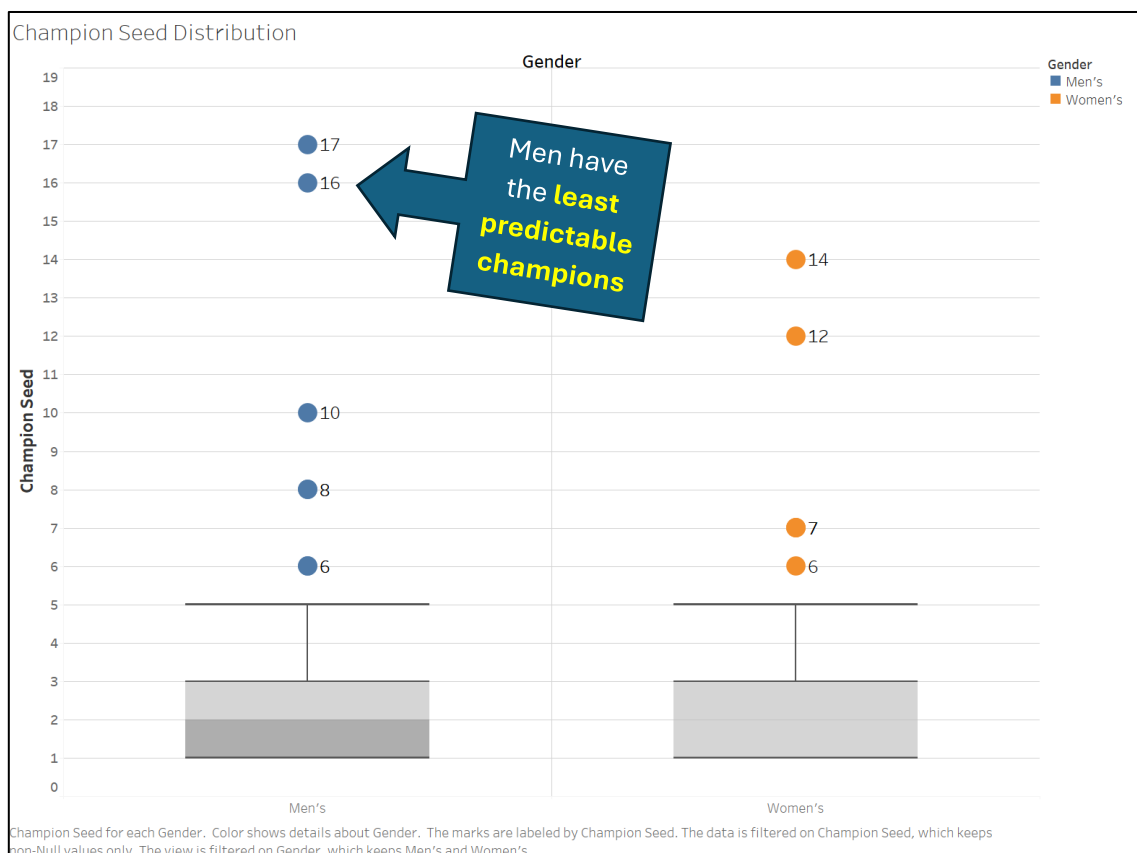


Figure 6



Interestingly both men and women have the same IQR of 2, and the same minimum value of 1 meaning that 75% of champions across all genders are either seeded as 1, 2 or 3. However, women have 50% of champions seeded as 1, implying there are more dominant top players in the cohort which aligns with Serena Williams' outsized skill for example. There are several outliers above the IQR in both men's and women's data, indicating champions with lower seed rankings can still win. In alignment with the previous conclusion about the dominance of lower seeded players in women's tennis, the outliers are seeded higher for men than women. This implies a slightly more even competitive scene in men's tennis compared to women's tennis. We highlight this with a colour coded annotation on the chart. Contrastingly, there were 3 unseeded winners who were not plotted: 1976 Men's, 1978 Women's and 2007 Women's. Additionally, there were several null values not plotted on the chart due to seeding not being available in the dataset preceding 1924.

## 5. Top-Winner Analysis

We define a top winner as a player who has at least 5 championships. Consequently, over the 1905-2024 period the following 7 top winners emerge. There are 109 Champions meaning that less than 7% of champions are classed as top winners. Hence, the following players are in an elite tier of their own who have shown a sustained excellence over time:

Top Winner	Debut Year
Novak Djokovic	2008
Roger Federer	2004
Serena Williams	2003
Margaret Court	1960
Roy Emerson	1961
Nancye Wynne Bolton	1936
Daphne Akhurst	1925

### 5.1 Top-Winner Championship Counts

We visualise the Number of Championships for top-winners below. We colour code each data point to visually differentiate each player, the size of the circle reflects the number of championships for quick interpretation and comparison of players. To improve clarity data labels outline the differences between visually similar plot sizes, which posed a challenge for precise interpretation.

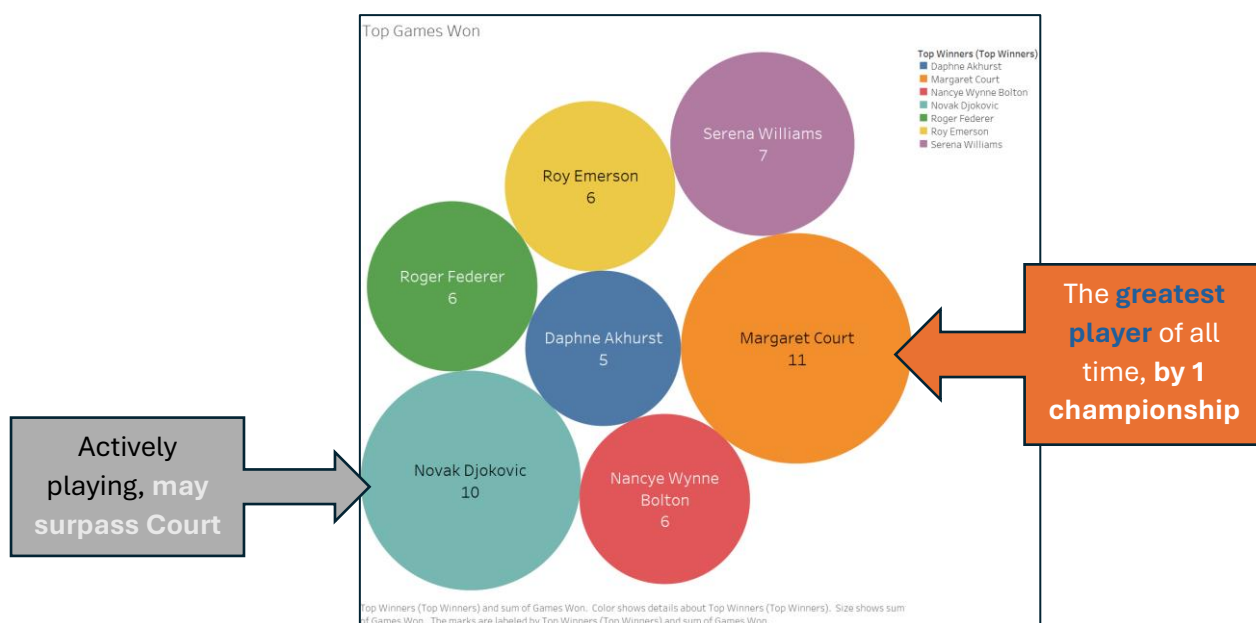


Figure 7



From this we see Margaret Court holds the record with an impressive 11 titles, making her the most successful player on the list and greatest of all time in the sport. We highlight this finding with a relevant colour coded annotation on the chart. Women were only allowed to compete in the Australian Open almost two decades after men. This makes Margaret's performance even more impressive as she showed dominance as a less supported class. Close behind is Novak Djokovic in the Men's division with 10 titles since his debut in 2008. Given that Djokovic is still an active player he has the potential to surpass Court's record. Additionally, Djokovic has showed dominance despite being the latest champion to debut. This makes him uniquely impressive as differentiating yourself is harder once a sport and its techniques has been developed and standardised. The modern era is defined by Novak Djokovic, Roger Federer and Serena Williams who has specifically separated herself greatly from other female players.

During the 1960's and 1970's Margaret Court and Roy Emerson dominated the Australian Open and were both legends of their era. Some of the earlier top champions of the sport are Nancye Wynne Bolton and Daphne Akhurst who must have had an intuitive understanding of the sport to outperform others despite optimal technique not being fully developed and available to learn.

## 5.2 Average Set Performance

Parallel coordinates work by plotting multidimensional data as lines across multiple parallel axes. Adjacent axes reveal relationships. The following figure outlines the average set win rate for each top-winner. A parallel coordinate chart highlights how these tennis champions perform across various set stages by visualising the link between adjacent data points.

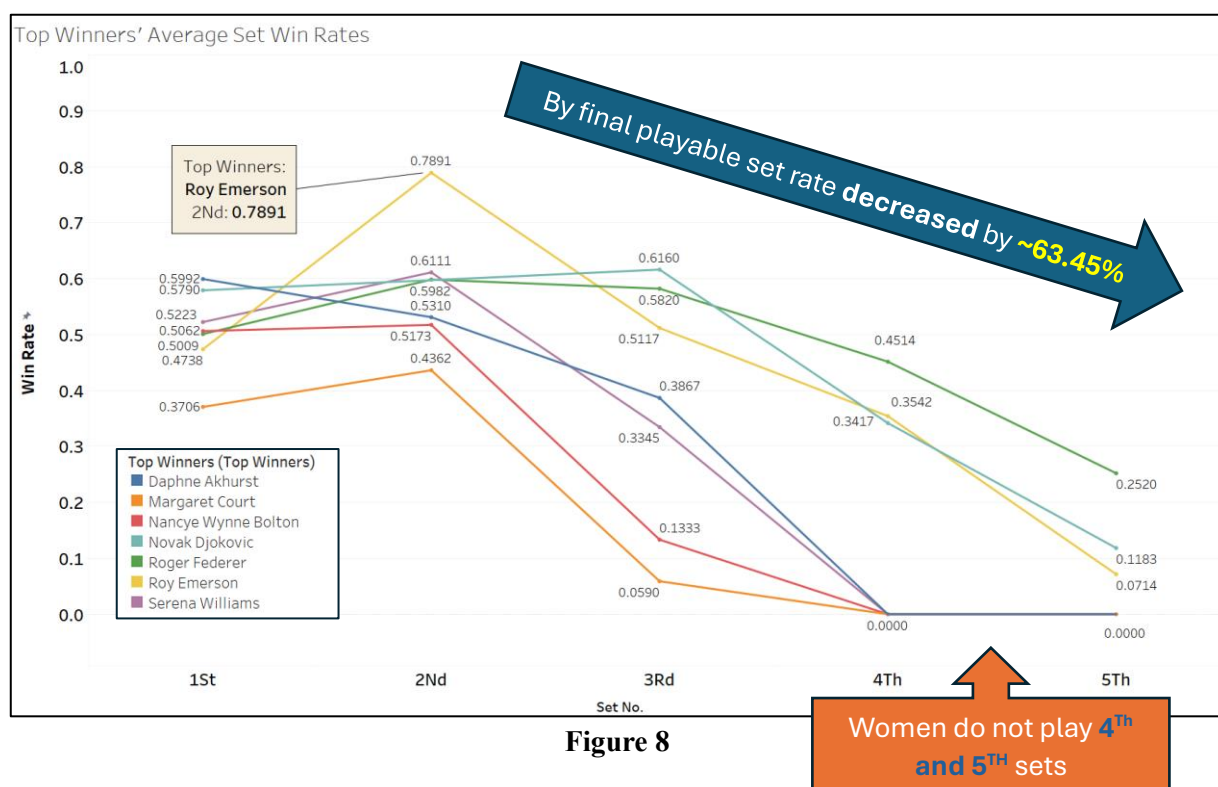


Figure 8

We see that the win rate in the first 3 sets was remarkably higher than the 4<sup>th</sup> and 5<sup>th</sup> set for men. This reveals that these players often closed out the championship game within the first 3 sets to dominate their opponent. The need to show early dominant performance is evident in all genders. Additionally, this chart reveals that Roy Emerson has a uniquely high peak win rate of 0.7891 in the 2<sup>nd</sup> set and consistently performs well across the first three sets. This peak was highlighted by a colour coded annotation. Most top winners peak at the second set (possibly due to them needing time to understand their opponents play style and weaknesses in the

first set), and all players decline in win rates after the 3<sup>rd</sup> set. This is highlighted by a colourful arrow annotation which describes the trend. This can be explained by physical and mental fatigue setting in. Notably, female players show this decline earlier to men due to physical limitations. Further, the 4<sup>th</sup> and 5<sup>th</sup> set win rates for female players is zero as they play in a best of 3 game format, compared to the men's best of 5. This is explained on the chart by a salient annotation pointing to these sets for women.

Novak Djokovic and Roger Federer perform the most consistently in the first 3 sets and overall, as evidenced by their straighter lines across the course of a game. Federer specifically maintains the highest win rate in the 5<sup>th</sup> set revealing his unique ability to endure longer matches and remain calm in decisive movements. Serena Williams shows the strongest performance for women across the first 3 sets, especially in the first two suggesting her dominance tends to be earlier in matches. Roy Emerson shows the worst performance for men in the 5<sup>th</sup> set possibly due to a lack of effective endurance training in earlier years of the sport.

Most players show strong early set dominance but experience significant declines in win rates in later sets. Federer is the exception, maintaining better performance in the later stages, while other players struggle in prolonged matches. These findings are highlighted by stylistic choices in graphing. Specifically, the top winner is each colour coded with distinct colours and a well-positioned key to enable visual differentiation. Further, the axis has been rescaled to maximise readability by increasing the area taken up by data points. More intuitive axis labels and titles were included to add context. A particular challenge with this chart were the overlapping labels and data points. This was alleviated by font changes, deletion of duplicate labels and the repositioning of overlapping labels to intuitive but readable locations.

### 5.3 Top-Winner Seeds

Here we use a Packed Bubbles chart to convey the average seeds of top-winners. This provides a quick visual comparison of the values based off of size. We support this by differentiating players by a unique keyed colour. Additionally, for clear comparison for similar values we add large annotations to address this difficulty.



Figure 9

The data shows that top players such as Djokovic, Federer, Serena Williams, and Margaret Court typically win tournaments when they are among the top seeds. The consistency of seeding in the women's category is notable, with fewer outliers in the champion seeds, whereas the men's category shows a bit more unpredictability with champions occasionally coming from lower-seeded positions. Nancye Wynne Bolton and Roy Emerson stand out as particularly dominant players, with very low average seed numbers (1.6 and 1.2, respectively), reflecting their overwhelming status as favourites in the tournaments they won. Despite this, Margaret Court and Djokovic still outperformed them greatly with notably similar higher seeds. This implies that seeding at the highest levels is not as important, beyond a certain threshold.

The champions fall below the third quartile for champion seeds supporting the notion that seeding is an accurate measure for dominance of champions. This excludes Roger Federer who has an average seed of 4 which is notably higher than other top-winners. This shows he has initially more variable performance than other top-winners which affects his placement. This results in less favourable positioning as higher seeded players are more likely to come up against lower seeded players earlier on in the tournament. Despite this he is still able to achieve many wins, indicating his unique endurance in tournaments. Likewise, had he been placed more competitively, he may have shown a similar level of dominance to Djokovic.

#### 5.4 Top-Winner Seeds Over Time

Here we visualise the average seed of top player's based off their debut year. This chart was chosen as it is the standardised method to convey time series data. Hence, we intuitively show how the average seed of a champion has progressed throughout the sport's development.

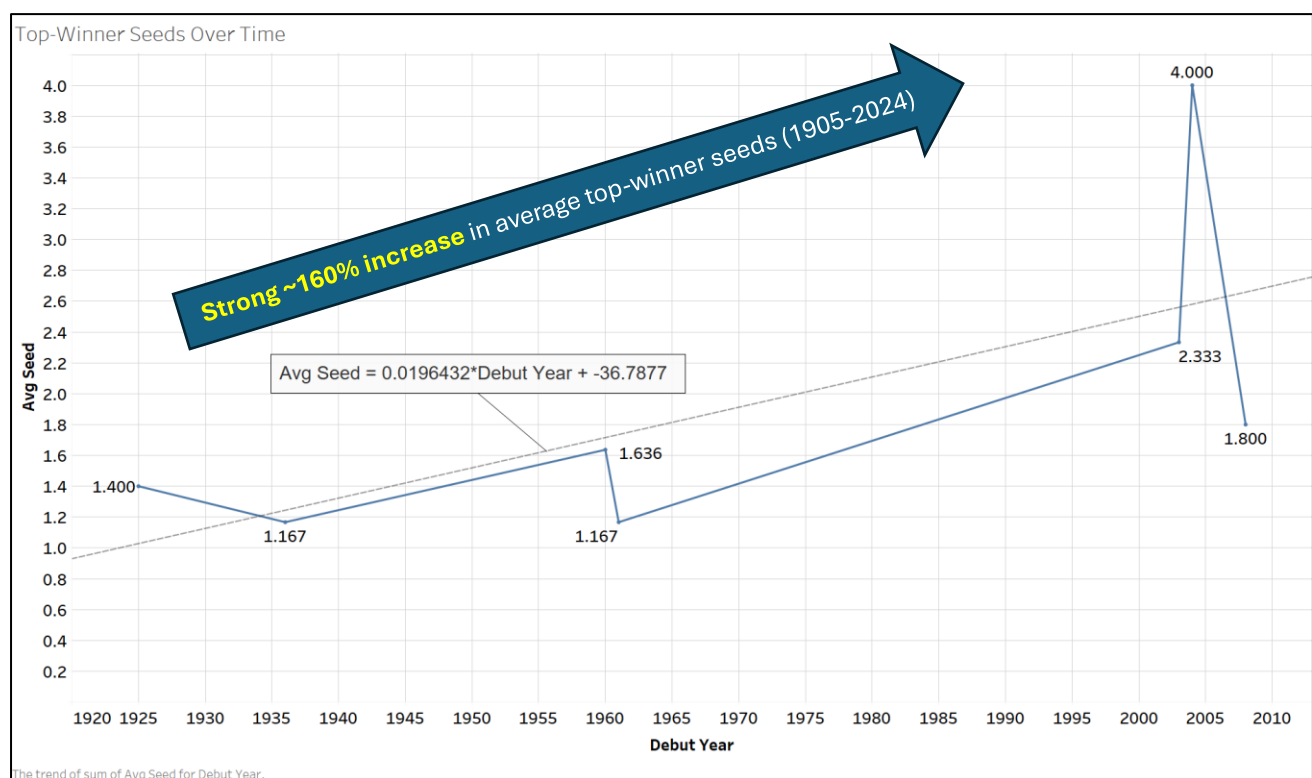


Figure 10

This reveals an upwards trend which is highlighted by an annotated trendline and a salient colourful arrow which describes the trend. This finding can be explained by the increasing competitiveness of the sport. As a sport develops the playing field gets evened by a fair distribution of knowledge and technique. Hence, there is less dominance of one single player which is reflected in higher seeded champions. This makes Novak Djokovic's achievement in the modern era exceptionally notable and impressive. We can use this trend to

extrapolate into the future. A particular challenge with plotting was clearly identifying where the data points were between the connecting lines, data labels addressed this.

### 5.5 Gender of Top-Winners

Below we visualise the distribution of top tennis winners by gender, women represent a slight majority. Pie charts are particularly effective at conveying proportions but are poor at conveying precise data values. We clarify the distribution with data labels and annotations placed intuitively to reveal that there are 4 female top winners and 3 male top winners. This makes up ~57% and ~43% respectively. We highlight this finding with a colour coded annotation. While women have a slight numerical advantage, the chart demonstrates the strong competitive excellence across both male and female tennis players in this dataset. This aligns with previous findings that suggest that the men's cohort is a more competitive environment. This makes Novak Djokovic's performance even more notable and poses an argument that his performance is more impressive than Margaret Court's.

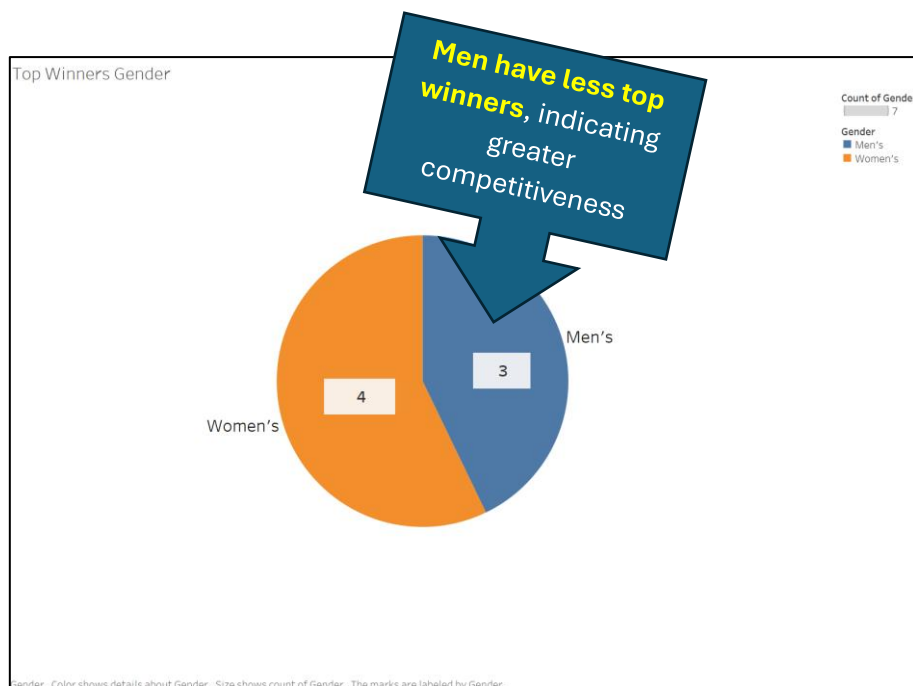


Figure 11:

### 5.6 Notable Top-Winner Championship Games

#### 5.6.1 Novak Djokovic

Year	Score
2023	6-3, 7-6(7-4), 7-6(7-5)
2020	6-4, 4-6, 2-6, 6-3, 6-4
2015	7-6(7-5), 6-7(4-7), 6-3, 6-0

Djokovic's championship victories reveal his adaptability, resilience, and mental strength. His 2023 win has two tiebreaks, demonstrating his capability in high-pressure moments. Notably, his 2020 five-set comeback, after losing two consecutive sets, highlights his flexibility and unmatched focus. Similarly, Djokovic's 2015 win reveals his fitness and ability to dominate late into matches, as evidenced by his dominant fourth set. Across these matches, Djokovic consistently displays his incredible defence, strategic returns, and stamina, making him one of the sport's strongest competitors.

#### 5.6.2 Roger Federer

Year	Score
2018	6–2, 6–7(5–7), 6–3, 3–6, 6–1
2010	6–3, 6–4, 7–6(13–11)
2007	7–6(7–2), 6–4, 6–4

Federer's performance throughout his career reflects his unmatched versatility. His 2018 victory, where he came back after a four-set slog to dominate the final set, reveals his ability to elevate his game when needed. Similarly, in 2010, Federer's impressive tiebreaker victory (13-11) underlined his composure and excellence under pressure. His 2007 victory, where he controlled the match from the first-set tiebreak onward, is another testament to his court mastery.

### 5.6.3 Serena Williams

Year	Score
2010	6–4, 3–6, 6–2
2007	6–1, 6–2
2015	6–3, 7–6(7–5)

Serena Williams' dominance in championship matches reflects her power, mental toughness, and tactical prowess. In 2010, after losing the second set, she responded with a dominant third set, demonstrating her ability to reset and overwhelm opponents. Her dominant victory in 2007 (6-1, 6-2) was a clear display of her aggressive baseline play. Additionally, her 2015 victory through a tight tiebreak further solidified her reputation as a clutch performer. Serena's relentless baseline play has made her a strong force in women's tennis, consistently outperforming her competition.

### 5.6.4 Margaret Court

Year	Score
1971	2–6, 7–6(7–0), 7–5
1962	6–0, 6–2

Margaret Court's championship victories highlight her adaptability and strategic prowess on the field. The 1971 flawless second-set tiebreak (7-0) showcases her ability to regain control and shift momentum. Her dominant 1962 win, where she lost only two games, underlines her power and dominance. Court's resilience is particularly evident in her many comeback victories, where she turns matches around through tactical changes and stamina. This made her a formidable champion throughout her career.

### 5.6.5 Roy Emerson

Year	Score
1965	7–9, 2–6, 6–4, 7–5, 6–1
1961	1–6, 6–3, 7–5, 6–4

Roy Emerson's ability to endure and outlast his opponents is best captured in his 1965 victory, where he rallied after losing the first two sets for five sets. This match exemplifies his resilience and adaptability, traits that defined his and other top-winner's success. Similarly, his 1961 comeback highlights his mental strength and strategic depth to follow through a poor start to the game. Emerson's dominated the sport and his opponents, particularly in longer matches where his endurance and mental fortitude were essential.

### 5.6.6 Nancye Wynne Bolton

Year	Score
1940	5–7, 6–4, 6–0
1947	6–3, 6–2

Bolton's victories reveal her mental resilience and strategic consistency. In 1940, after losing the first set, Bolton turned the match around, finishing with a perfect third set. Her performance in 1947 showcases her ability to consistently outplay her opponents in straight sets, this reveals her aggressive baseline play and powerful serving. Bolton's ability to adjust her strategy and maintain her focus has placed her as one of the most consistent champions of her time.

### 5.6.7 Daphne Akhurst

Year	Score
1930	10–8, 2–6, 7–5
1929	6–1, 5–7, 6–2

Daphne Akhurst's championship matches exemplify her mental toughness and ability to perform in long and competitive matches. Her 1930 win involved a drawn out 10-8 first set and a third-set comeback, illustrating her resilience and tactical skill. Likewise, her 1929 win, required her to regain control after losing the second set. This exemplifies her adaptability. Akhurst's strength is in her consistent baseline play and her ability to stay composed under pressure, making her one of the few standout players of her era.

## 6.0 Conclusion

The exploration of the key trends in the Australian Open championship history dataset revealed valuable insights. This includes interesting revelations between men's and women's tennis. Specifically, men had longer matches even when accounting for their 5-set games. Additionally, Women's champions tend to emerge from lower seeds more frequently than men, indicating a more predictable and dominant dynamic compared to men's games. Tableau's optimisation for visualisations such as scatter plots, box plots and tree maps were vital in highlighting these trends. This provided clarity in interpreting the highly dimensional data. Europe and the USA have dominated recent games, however, before the 1980's Australia outclassed all other competitors.

We also found interesting revelations amongst top winners, including key strengths and commonalities which were responsible for their dominance. Djokovic, with his 10 championships will likely surpass Margaret Court's record 11 titles, showcasing their remarkable consistency and ability to handle high-pressure situations. Serena Williams and Roger Federer exhibit early match dominance, like all top winners, where Federer showcases a particularly strong ability to endure long matches. Roy Emerson shows resilience by recovering from early setbacks in matches, while Nancy Wynne Bolton and Daphne Akhurst dominated during earlier eras when the sport was less developed, showcasing their natural affinity for the sport. Additionally, they possessed resilience and consistency.

Despite several data anomalies which needed to be addressed, such as incomplete seed rankings or the abnormal match endings, the overall quality of the data produced meaningful analysis. Additional challenges included effectively visualising the dense data with careful graph design and axis rescaling to improve readability and clarity. Clarity was further improved by showcasing key findings and with colour coded storytelling annotations to the graph.

**7.0 Appendix A**

<b>Attribute</b>	<b>Type</b>	<b>Description</b>
Year	I	The year of the match
Gender	N	The gender of the match (Men's or Women's)
Champion	N	The name of the winning player
Champion Nationality	N	The nationality code (ISO) of the champion player
Champion Country	N	The country which the champion player represents
Score	N	The final score of the match, comma separated sets.
Champion Seed	O	The seed ranking of the champion player for the tournament
Min	R	The total number of minutes the match lasted
1 <sup>st</sup> -won	R	The number of first sets won by the champion player
1 <sup>st</sup> -loss	R	The number of first sets lost by the champion player
2 <sup>nd</sup> -won	R	The number of second sets won by the champion player
2 <sup>nd</sup> -loss	R	The number of second sets lost by the champion player
3 <sup>rd</sup> -won	R	The number of third sets won by the champion player
3 <sup>rd</sup> -loss	R	The number of third sets lost by the champion player
4 <sup>th</sup> -won	R	The number of fourth sets won by the champion player
4 <sup>th</sup> -loss	R	The number of fourth sets lost by the champion player
5 <sup>th</sup> -won	R	The number of fifth sets won by the champion player
5 <sup>th</sup> -loss	R	The number of fifth sets lost by the champion player
Runner-up	N	The name of the player who was the loser in the final match
Runner-up Nationality	N	The nationality code (ISO) of the runner-up player
Runner-up Country	N	The country name which the runner-up player represents
Runner-up Seed	O	The see ranking of the runner-up player in the tournament.