

32146 Data Visualisation and Visual Analytics

Tutorial Monday 11:00 am – 1:00 pm

Assessment Task 2: Advanced Data Visualisation

Author:

Viet Kien Nguyen-14443939

1. Executive Summary

This report analyses 120 years of Australian Open men's and women's singles finals using advanced data visualisation techniques. Drawing on a dataset enriched with manually collected match durations, the study highlights long-term trends in player dominance, match competitiveness, and the globalisation of the tournament.

Key findings include the shifting balance of national success: Australia dominated the early decades, particularly in men's finals, before the United States, Europe, and more recently Serbia and Eastern Europe emerged as powerhouses. In women's competition, the United States and Australia contributed the largest number of champions, with modern contributions from nations such as Russia, Belarus, and China demonstrating the tournament's expanding global reach.

Performance analysis identified a small group of "top players" with five or more titles, including Novak Djokovic, Roger Federer, Margaret Court, Serena Williams, Nancye Wynne Bolton, Roy Emerson and Daphne Akhurst. Treemaps and win rate charts highlighted their dominance across eras, while parallel coordinate charts revealed both their set-level consistency and the gradual decline in win rates across later sets, indicating closer competition in extended matches. Scatter plots of match duration showed a clear increase in the length of men's finals over time, with the 2012 Djokovic–Nadal final lasting 353 minutes as the longest in history. Women's finals, while generally shorter, also contained notable competitive outliers.

The visualisations, developed in Tableau, provided both high-level overviews and detailed player-level comparisons. Treemaps, scatter plots, geographic maps, and parallel coordinates were used to enhance readability and highlight breakthrough points such as unseeded champions and extreme outliers. Tableau's interactivity and flexibility allowed for effective storytelling, though some data transformations in Excel were required to prepare fields such as sets played, win rates, and flags for retired matches.

Overall, the analysis demonstrates how advanced visualisation can uncover insights into the evolution of the Australian Open, revealing both historical patterns and the extraordinary performances of the tournament's greatest champions.

2. Dataset Overview

2.1 Data Source

The dataset for this analysis was provided in Excel format and was enriched by manually collecting match duration values from reliable online sources. It covers approximately 120 years of men's and women's singles finals at the Australian Open, offering a comprehensive historical perspective on the tournament.

2.2 Data Fields & Formats

The dataset contains a wide variety of fields in different formats. Year is treated as a continuous numeric field, suitable for time trend and scatter analyses. Gender is categorical, enabling direct comparison between men's and women's competitions. Champion and Runner-up are categorical fields used in treemaps and win rate charts, while Nationalities are also categorical but assigned a geographic role for mapping visualisations. Champion Seed is numeric but includes the category "Unseeded," making it useful in scatter plots and parallel coordinate analysis. Match Score and individual Set results are stored as text, which were parsed into separate values for total games and sets played. Match duration (Mins) is numeric and supports scatter plots and outlier analysis. A range of derived fields were also created, including Sets Played, Set Win Rate, Total Games Played, Retired/Walkover flag, Total Wins, Total Losses, and Match Win Rate. These additions enabled deeper analysis of player performance and tournament patterns.

2.3 Data Transformations

Several transformations were undertaken to ensure consistency and analytical usefulness. Match duration values were filled where possible; if unavailable, rows were marked as N/A or given estimated values to approximate performance. Match duration data was collected from reliable online sources using ChatGPT which includes the official Australian Open website, ATP/WTA statistics pages, and Wikipedia match records. Where official data was not available, estimates were generated using a set-based approximation method (~30 minutes for short sets, ~50 minutes for long/tiebreak sets) and cross-referenced against similar matches (see Appendix). For example, a straight-set 6–2, 6–4, 6–3 final was estimated at approximately 110–120 minutes, while a five-set final with multiple 7–6 sets was estimated closer to 250–300 minutes. These values were flagged as approximate and used only to preserve dataset continuity for trend analysis.

Champion Seed values were standardised, with blank entries replaced by “U” (Unseeded). Scores were reformatted for consistency, such as changing periods to commas in row 5. A per-player summary table (refer to figure 1 below) was created to identify champions with five or more wins, forming the basis for top-player analysis. Additional calculated fields were derived, such as set-based win rates and total games played, while pivot tables were used to isolate the top performers. Some calculated fields, such as win rates for 4th or 5th sets, returned #DIV/0! when those sets were not played (e.g., straight-set matches). These errors were left as-is in the Excel file since they reflect the reality of shorter matches. For analysis and visualisation, these were treated as ‘not applicable’ rather than true errors. These transformations ensured that the dataset could support advanced visualisations such as win rate performance charts and parallel coordinates.

| A | B | C | D | E | F | G |
|------------------------|--------------|---|---|---|---|---|
| 1 Champion | ▼ Total wins | | | | | |
| 2 Margaret Court | 11 | | | | | |
| 3 Novak Djokovic | 10 | | | | | |
| 4 Serena Williams | 7 | | | | | |
| 5 Roger Federer | 6 | | | | | |
| 6 Nancye Wynne Bolton | 6 | | | | | |
| 7 Roy Emerson | 6 | | | | | |
| 8 Daphne Akhurst | 5 | | | | | |
| 9 Jack Crawford | 4 | | | | | |
| 10 Ken Rosewall | 4 | | | | | |
| 11 Evonne Goolagong | 4 | | | | | |
| 12 Monica Seles | 4 | | | | | |
| 13 Andre Agassi | 4 | | | | | |
| 14 Steffi Graf | 4 | | | | | |
| 15 Mats Wilander | 3 | | | | | |
| 16 James Anderson | 3 | | | | | |
| 17 Joan Hartigan | 3 | | | | | |
| 18 Martina Navratilova | 3 | | | | | |
| 19 Rod Laver | 3 | | | | | |
| 20 Martina Hingis | 3 | | | | | |
| 21 Adrian Quist | 3 | | | | | |
| 22 Pete Sampras | 2 | | | | | |
| 23 Chris Evert | 2 | | | | | |

Figure 1: Top players with 5 or more wins

2.4 Data Characteristics & Outliers

The dataset reflects several interesting characteristics and notable outliers. Extremely long matches appear in the men’s finals, including the historic 2012 clash between Djokovic and Nadal, which lasted nearly six hours. Other lengthy encounters typically involved five sets and multiple tiebreaks. Retired and walkover matches also occur, specifically in three rows (40, 73, 122), which were flagged and excluded from duration-based averages to avoid distortion. Another distinctive characteristic is the presence of unseeded champions, with Serena Williams’ 2007 victory standing out as a rare breakthrough. These

features highlight both the consistency of data across most years and the exceptions that shape the narrative of the tournament.

3. Visualisation & Analysis

3.1 Treemap - Champions by Gender

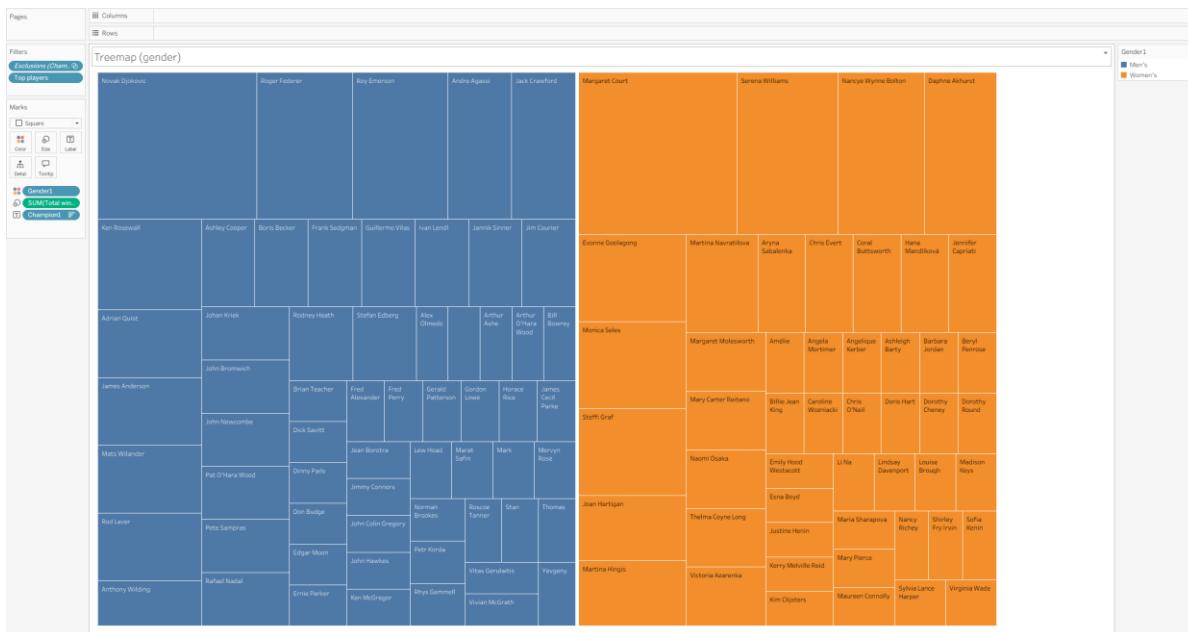


Figure 2: Treemap - Champions by Gender

The treemap (refer to figure 2 above) provides a comparative view of Australian Open champions split by gender. Each rectangle represents an individual champion, with size proportional to their total number of titles. The blue blocks correspond to men's champions, while orange blocks represent women's champions.

From the treemap, Novak Djokovic and Roger Federer dominate the men's side, with Djokovic leading the Open Era in total titles. For women, Margaret Court and Serena Williams stand out with large, prominent blocks that indicate their exceptional success. The treemap also highlights the broader distribution of champions: men's titles are concentrated among a handful of dominant players, while women's titles, although also

featuring a few standouts, are spread more widely across multiple champions. This treemap functions as a champions comparison chart, allowing a clear gender-based analysis of title distribution across eras.

3.2 Treemap – Top Players (5 or more wins)

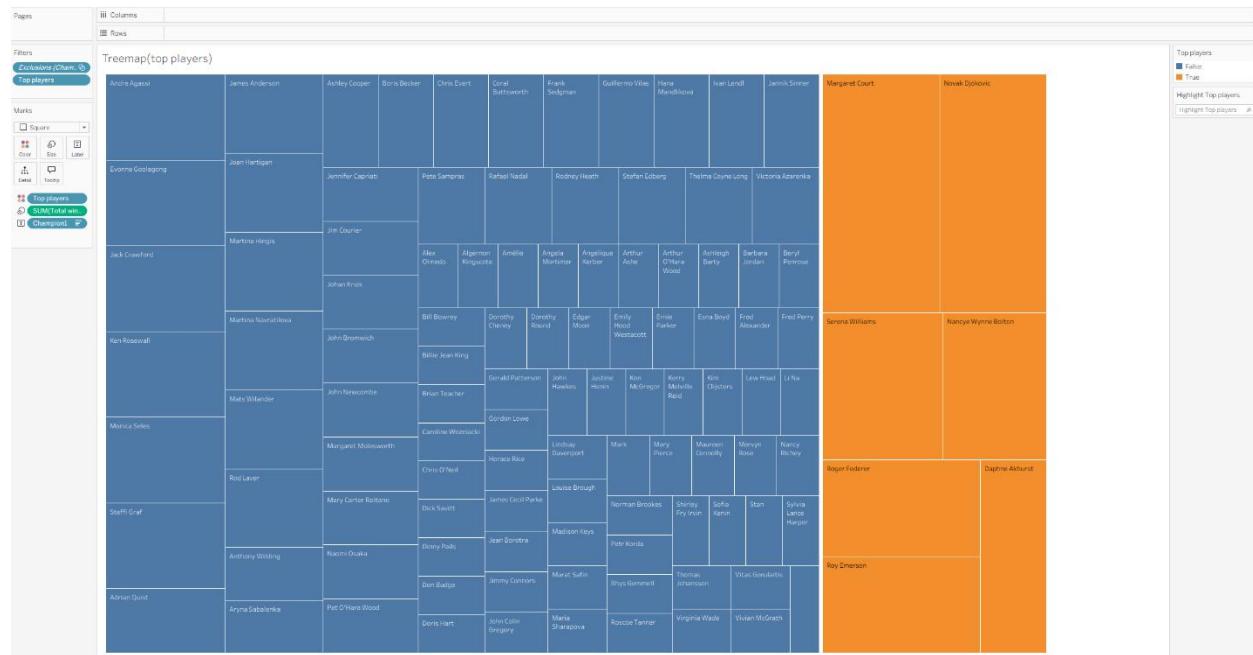


Figure 3: Treemap – Top Players

The second treemap (refer to figure 3) isolates only those champions who have secured five or more Australian Open titles. These top players are highlighted in orange, while all other champions are shown in blue. This design choice draws attention to the select group of players who have consistently dominated the tournament.

The treemap confirms that 7 players belong to this category: Margaret Court, Novak Djokovic, Serena Williams, Nancye Wynne Bolton, Roger Federer, Roy Emerson, and Daphne Akhurst and are considered as the most successful players in Australian Open history. However, Djokovic's and Court's large blocks reflect their record-setting achievements with visibly more titles than other top players, showing their dominance.

This chart is crucial for addressing the focus on “top players,” as it not only identifies them but also visually contrasts their achievements against the wider pool of champions. The balance of men and women among the top players also provides material for gender comparison: Court and Serena highlight women’s dominance across different eras, while Djokovic and Federer highlight men’s dominance in the modern Open Era.

3.3 Geographic Map – Champion Countries by Gender

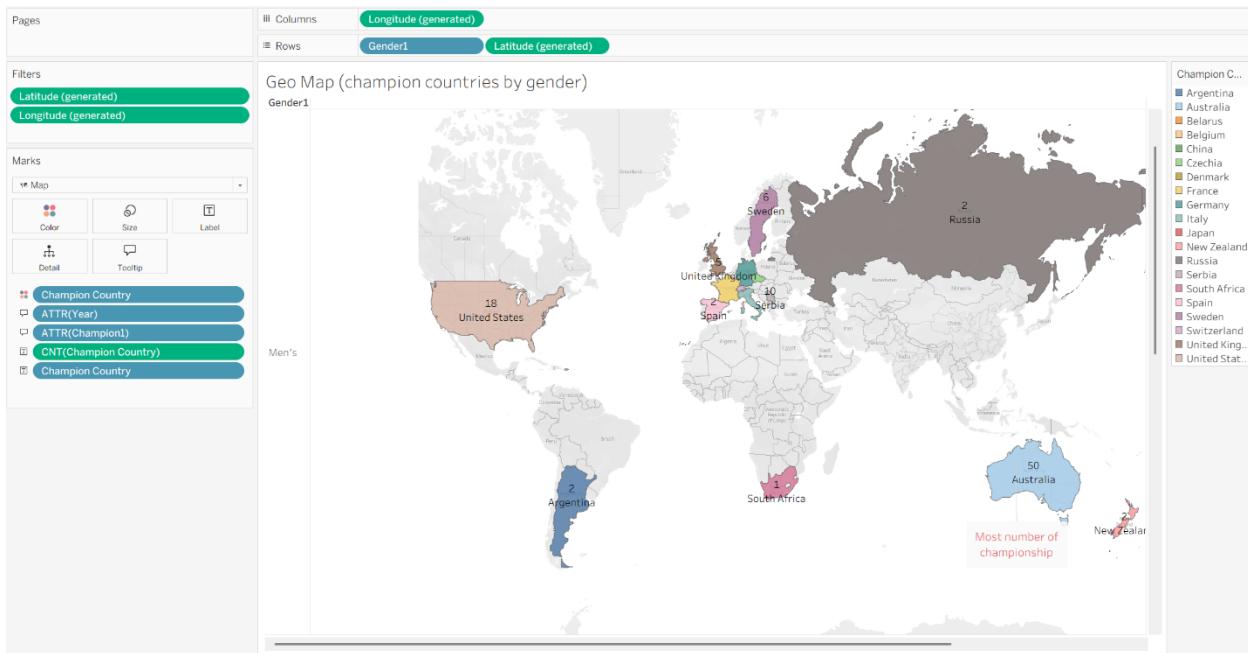


Figure 4: Men champion countries

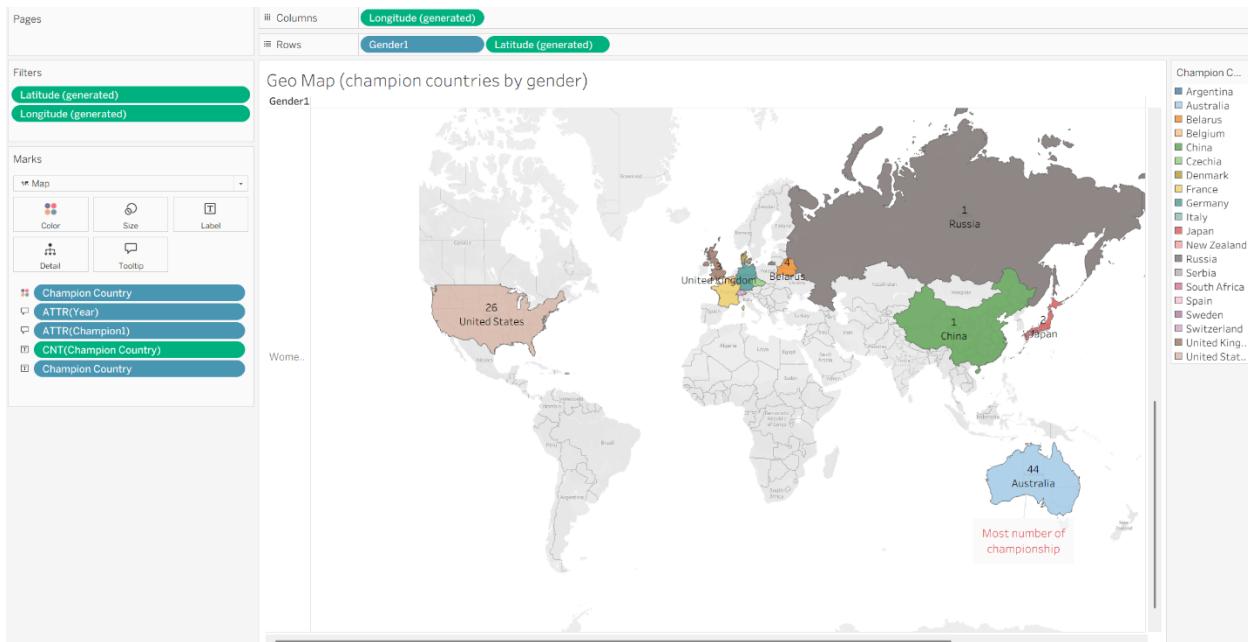


Figure 5: women champion countries

These maps (refer to figure 4, 5) display the distribution of Australian Open champions by country, separated into men's and women's finals. Australia dominates the men's map

with 50 titles, reflecting its historical strength in the early decades of the tournament (1905 – 1976). The United States also appears prominently, particularly from the 1970s onwards, while European nations such as Sweden, Switzerland, and Spain mark important periods of global expansion. Serbia's rise in the 2000s, through Novak Djokovic, highlights the shift in power to Eastern Europe.

On the women's side, Australia again holds a significant share with 44 titles, largely through Margaret Court, Nancye Wynne Bolton, and other early champions. However, the United States leads with 26 titles, showcasing dominance from stars like Serena Williams and Chris Evert. The women's map also shows broader diversity in continents, with champions from Russia, Japan, China, and Belarus, reflecting the globalisation of the women's game in the modern era. These maps demonstrate how Australia's early dominance has given way to a more international field of champions in both men's and women's competitions.

3.4 Geographic Map – Runner-up Countries by Gender

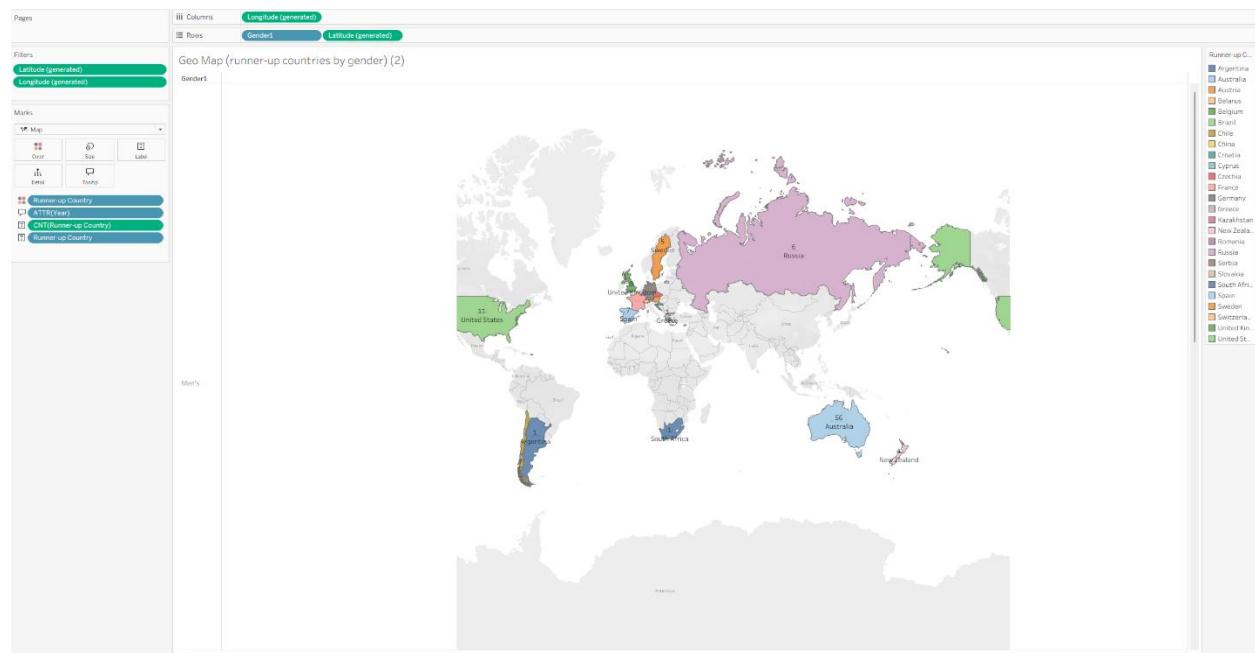


Figure 6: men runner-up countries

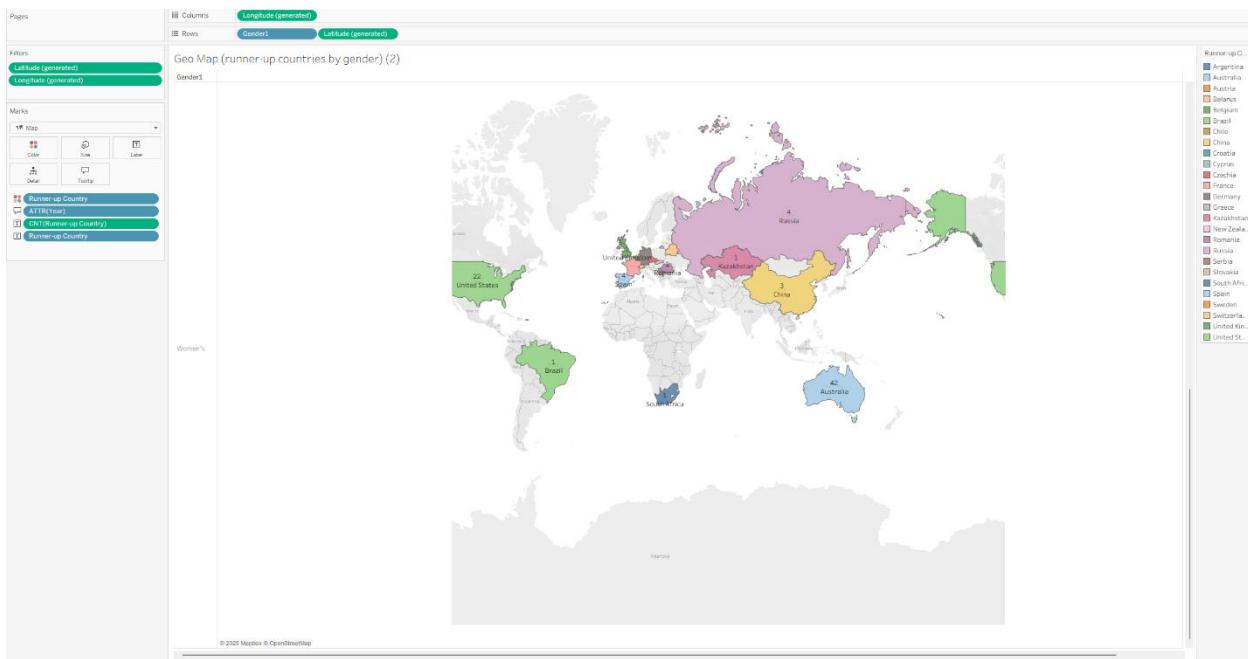


Figure 7: women runner-up countries

The runner-up maps (refer to figure 6, 7) reveal a different distribution from the champions. On the men's side, Australia still leads with 56 runner-up finishes, showing both its strength and the fact that not all Australian finalists converted opportunities into titles. Other frequent finalists include the United States (11), the United Kingdom, and Sweden. Russia also appears with six runner-up finishes, reflecting the presence of strong contenders without consistent championship success.

For women's finals, the United States again features prominently with 22 runner-up positions, showing the depth of its players across eras. Australia follows closely with 42 runner-up finishes, underlining its early role as a central hub of the tournament. The map also highlights contributions from Russia (4), China (3), and even Kazakhstan (1), indicating the increasing spread of international finalists. These runner-up maps indicate that while some countries produced frequent champions, others produced consistent finalists who fell just short, underscoring the competitive nature of the tournament across eras. These maps provide a champions comparison by nationality, showing how dominance and finalist appearances have shifted geographically between men's and women's competitions.

3.5 Parallel Coordinates - Top Players' Set Win Rate

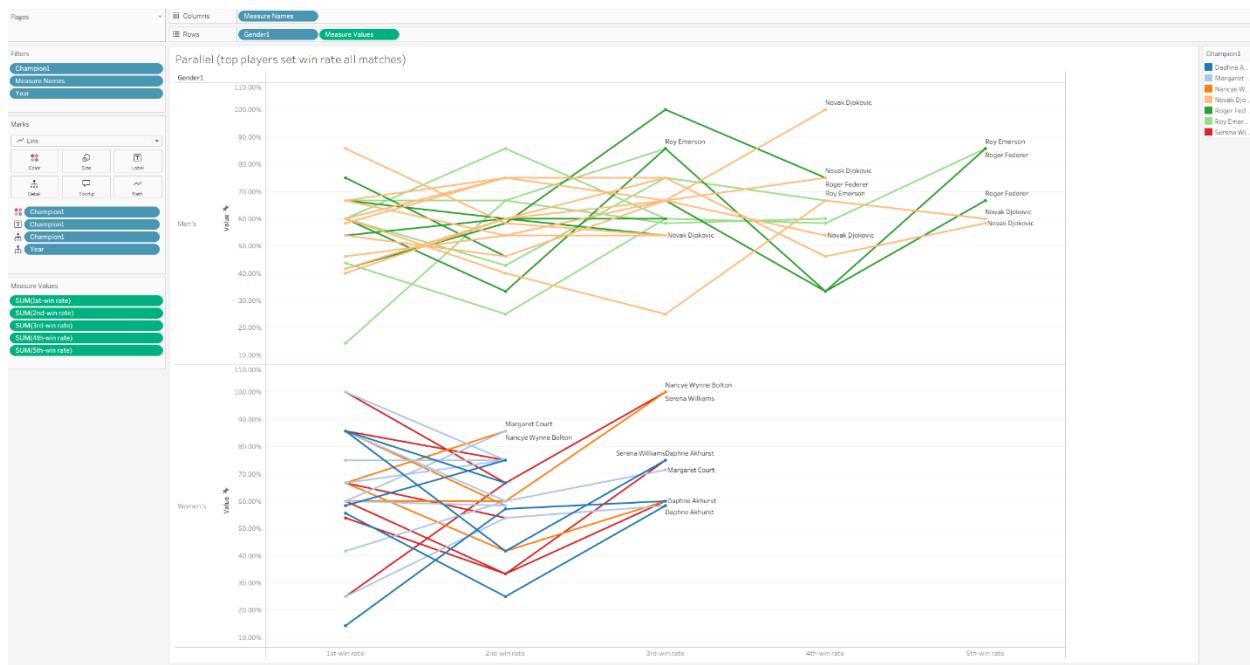


Figure 8: Parallel Coordinates - Top Players' Set Win Rate Performance (All Matches)

This parallel coordinate chart (refer to figure 8) displays the set win rates of top players across all their Australian Open finals. Each line represents a champion, with men's and women's shown separately. The chart shows how often these players won the 1st, 2nd, 3rd, 4th, and 5th sets in their matches. Parallel coordinates often require normalisation when variables use different scales. In this case, each axis represents win rate (%) data already bounded between 0 and 100, so additional normalisation was unnecessary. This ensured fair comparison without distortion.

The visual highlights that top players generally maintain strong set win rates across their careers. For example, Novak Djokovic and Roger Federer show consistently high win rates across multiple sets, reflecting their ability to perform well in long matches. Among women, Serena Williams and Margaret Court stand out for winning a very high percentage of opening sets, often establishing control early in finals. The chart also makes clear that even among top champions, set win rates vary: some excel at fast starts, while others are stronger in later sets, showing differences in match strategy and resilience.

3.6 Parallel Coordinates – Top Players’ Aggregate Set Win Rate

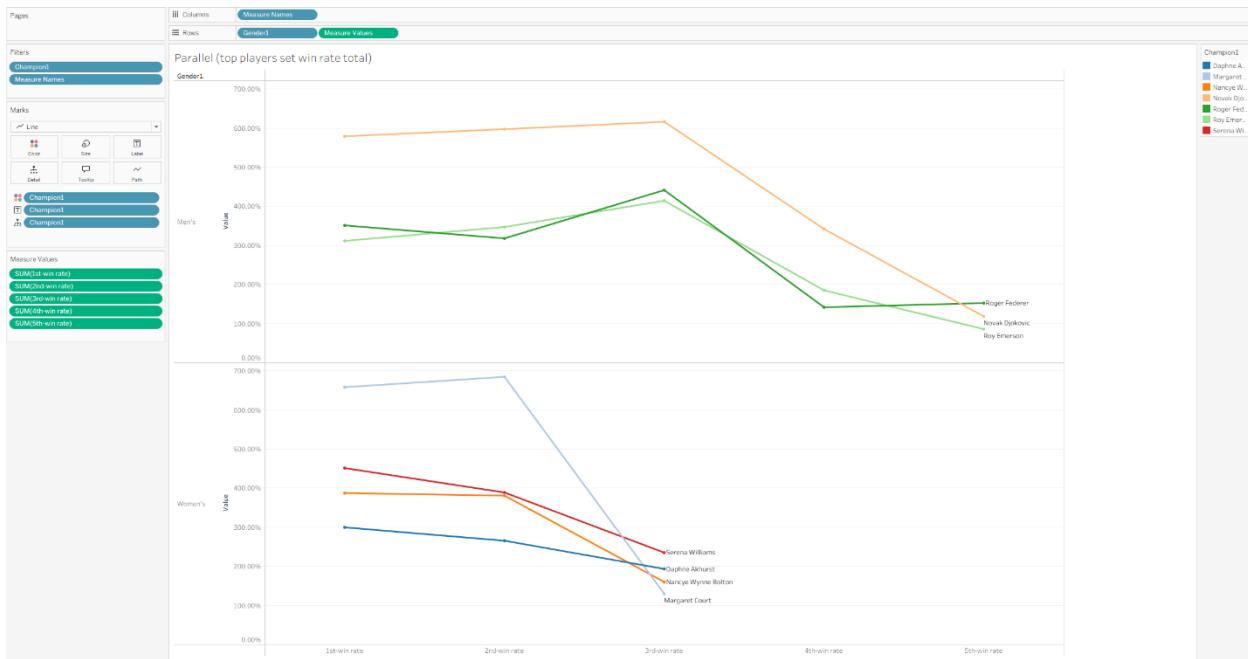


Figure 9: Parallel Coordinates – Top Players’ Aggregate Set Win Rate (Total)

The second chart (refer to figure 9) aggregates win rates across all matches, showing overall trends in how often top players win sets at each stage of a final. Normalisation is also not needed in this case, as the data are on the same scale. The general pattern indicates that win rates are usually higher in the first few sets and gradually decline in later sets. This suggests that while champions often start strongly, opponents are more competitive in the third, fourth, or fifth sets, making these sets closer contests.

For men, Djokovic, Federer, and Emerson all show this downward slope starting from third set, reflecting the intensity of longer men’s finals, where late sets are frequently tighter. In women’s finals, Margaret Court and Nancye still maintain their win rates even in later sets (second sets), but the same downward trend is evident, demonstrating that opponents often push them harder as matches progress.

This chart provides important context: it shows not only dominance but also where matches became more competitive. The declining win rates in later sets emphasise that while top players usually secure victory, their opponents often fought back, making for more balanced finals in the closing stages.

3.7 Scatter chart - Match Duration Over Time

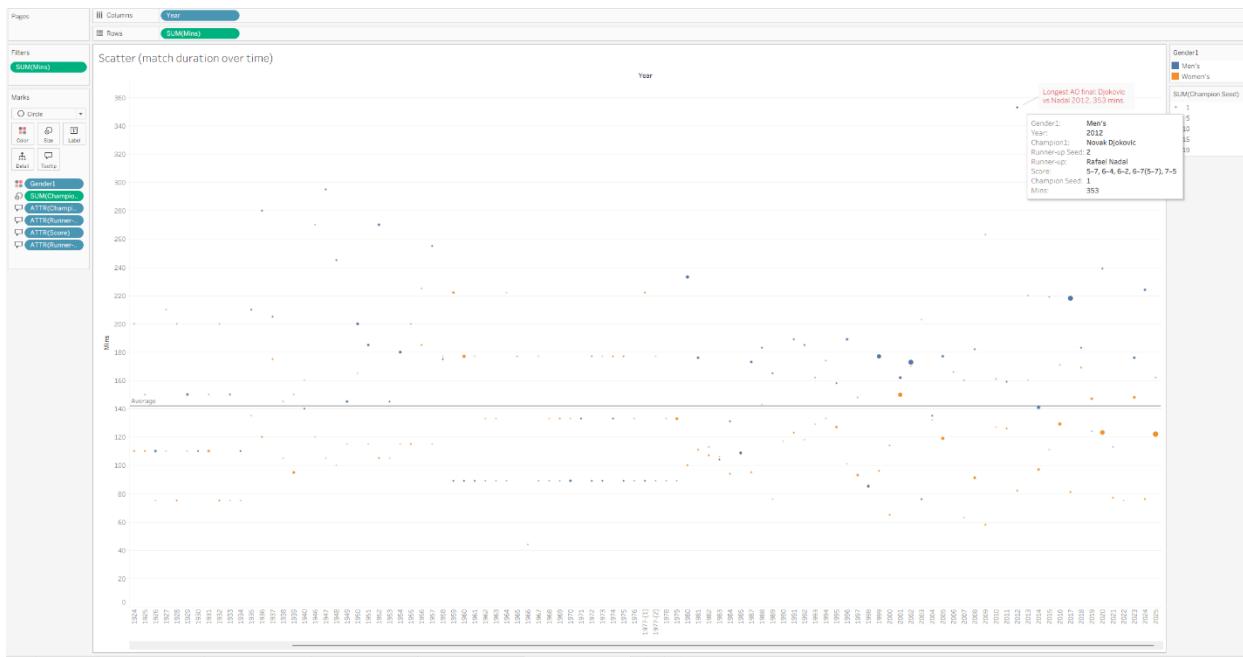


Figure 10: Scatter chart - Match Duration Over Time

This scatter chart plots the duration of Australian Open finals against year, with men’s matches shown in blue and women’s in orange. Each point represents a final, and the size of the marker reflects the champion’s seed. An average line has been added to help compare individual matches with the overall trend.

The chart highlights a clear pattern: men’s finals are generally longer than women’s, with durations often exceeding 180 minutes. A notable outlier is the 2012 men’s final between Novak Djokovic and Rafael Nadal, which lasted 353 minutes — the longest in tournament history. Women’s finals tend to be shorter, averaging closer to 100 minutes, though several extend above 150 minutes, showing that competitive balance exists in some eras.

The scatter also reveals historical trends. Early finals, particularly in the first half of the 20th century, were shorter on average, reflecting differences in playing styles and tournament conditions. Over time, men’s matches in particular have lengthened, especially in the modern era where physicality and baseline rallies dominate play. The inclusion of seeds shows that top-seeded players are not guaranteed quick victories; many long matches involved high seeds battling one another in closely contested finals.

Overall, this visualisation provides insight into both gender differences and the evolution of match intensity over time, while also drawing attention to extreme outliers that shape the tournament’s history.

8. References

- Australian Open. (n.d.). *Australian Open official website*. Tennis Australia. Retrieved September 2025, from <https://ausopen.com/>
- ATP Tour. (n.d.). *ATP Tour – Results archive*. Association of Tennis Professionals. Retrieved September 2025, from <https://www.atptour.com/>
- International Tennis Federation. (n.d.). *ITF Tennis – Tournament records and statistics*. International Tennis Federation. Retrieved September 2025, from <https://www.itftennis.com/>
- Wikipedia contributors. (n.d.). *List of Australian Open men's singles champions*. In *Wikipedia*. Retrieved September 2025, from https://en.wikipedia.org/wiki/List_of_Australian_Open_men%27s_singles_champions
- Wikipedia contributors. (n.d.). *List of Australian Open women's singles champions*. In *Wikipedia*. Retrieved September 2025, from https://en.wikipedia.org/wiki/List_of_Australian_Open_women%27s_singles_champions

- Wikipedia contributors. (n.d.). *Australian Open*. In *Wikipedia*. Retrieved September 2025, from https://en.wikipedia.org/wiki/Australian_Open
- Tableau Software. (2025). *Tableau Desktop (Version 2025.2) [Computer software]*. Salesforce. <https://www.tableau.com/>
- Microsoft. (2025). *Microsoft Excel [Computer software]*. Microsoft Corporation. <https://www.microsoft.com/>

9. Appendix

To supplement the main analysis, additional support was sought using ChatGPT. The tool was used to gather official match durations where available and to estimate missing values based on set scores. This includes:

- ChatGPT conversation (link): <https://chatgpt.com/share/68db72f6-b2b0-8004-90d4-1934b12634c0>
- Screenshots of the prompts used (Figures A1 and A2).

The screenshot shows a ChatGPT interface. The user prompt is:

can you look up the match duration (in minutes) for each australian open tennis final match (for both men and women) during period 1905-2025, look up the internet and give me the result (also the source). I will fill it in the excel file by myself

The AI response is:

I've found match durations for several **men's Australian Open finals**, as well as one **women's final** where duration was reported. Here's what I could gather from reliable web sources:

| Men's Singles Finals – Durations | | | |
|----------------------------------|-----------------------------------|--|------------------------------|
| Year | Final Match-up | Duration | Source |
| 2009 | Rafael Nadal vs. Roger Federer | 4 hours 23 minutes | Wikipedia |
| 2012 | Novak Djokovic vs. Rafael Nadal | 5 hours 53 minutes (longest Grand Slam final ever) | Wikipedia +1 |
| 2022 | Rafael Nadal vs. Daniil Medvedev | 5 hours 24 minutes | Wikipedia |
| 2024 | Jannik Sinner vs. Daniil Medvedev | 3 hours 44 minutes | Wikipedia |

Figure A1: Example ChatGPT prompt used to request match durations

for values that you cannot find, can you fill in with estimated value: (~30 minutes for short sets, ~50 minutes for long/tiebreak sets) and cross-referenced against similar matches. For example, a straight-set 6–2, 6–4, 6–3 final was estimated at approximately 110–120 minutes, while a five-set final with multiple 7–6 sets was estimated closer to 250–300 minutes.

Here's a proposed approach: for finals where I found a confirmed duration, I report that. For others, I estimate based on your rule ("~30 min for a short set, ~50 for a long/tiebreak set") and by comparing to similar scorelines.

I'm only able to find confirmed durations for some recent finals. (Older finals often don't have reliable time records online.) I'll list both confirmed and estimated durations.

Confirmed Durations (where available)

| Year | Gender | Scoreline / Notes | Duration (minutes) | Source |
|------|--------|--|--------------------|------------------------------|
| 2009 | Men | Nadal d. Federer (5 sets, included a tiebreak) | 4 h 23 m = 263 | Wikipedia |
| 2012 | Men | Djokovic d. Nadal (5 sets, two tiebreak sets) | 5 h 53 m = 353 | Wikipedia +1 |
| 2022 | Men | Nadal d. Medvedev (5 sets, one | 5 h 24 m = 324 | Wikipedia |

Figure A2: Example ChatGPT response showing method of estimating match durations.