

# 32146 DATA VISUALISATION AND VISUAL ANALYTICS AUTUMN 2024

## Assessment Task 2: Advanced Data Visualization

SIDDHARTH ABHIRAM 25037730

# Table of Contents

- 1. Overview of the Australian Open**
- 2. Summary of the Australian Open Tennis Dataset**
  - Dataset Characteristics
  - Data Formats and Visualization Techniques
  - Integration in Tableau
  - Advantages and Disadvantages of Tableau
- 3. Analysis of Winners by Country**
  - Dominance by Australia and the United States
  - European Representation and Success
  - Geographic and Cultural Implications
- 4. Emerging Talents and Early Career Successes**
  - Early Career Successes of Recent Players
  - Discussion on High Potential Indicators and Career Growth Trajectories
- 5. Resilience in Competition**
  - Statistical Analysis of Match Recoveries
- 6. Female Winners and Global Trends (Bubble Chart Analysis)**
  - Emerging Trends and Historical Implications
- 7. Decadal Dominance in Men's Tennis**
  - Analysis of Player Success Over Different Decades
  - Insights into the Evolution of Men's Tennis and Player Strategies
- 8. Longevity and Performance Trends**
  - Career Trajectories and Peak Performance Periods
- 9. Consistency of Top Players Over the Years**
  - Win Rates and Career Phases of Legendary Players
- 10. Legacies of Success**
  - Visual Representation of Tournament Victories
- 11. Trends in Player Performances Over Time**
  - Analysis of Evolving Competition and Win Rates
  - Historical Context and Player Performance
- 12. Executive Summary**
  - Key Findings

## **1.Overview of the Australian Open**

The Australian Open is one of the four Grand Slam tennis tournaments, considered the most prestigious events in the sport. As the first Grand Slam of the calendar year, the Australian Open holds a significant place in the global tennis landscape.

The tournament has a rich history, dating back to 1905 when it was first held as the Australasian Championships. Over the past 120 years, the Australian Open has evolved to become a world-class event that attracts the top tennis players from around the globe.

## **2.Summary of the Australian Open Tennis Dataset**

The Australian Open tennis dataset has been meticulously organized to provide a detailed look at the performance metrics of participants over its storied history. This dataset includes separate sheets for men's and women's data, enhancing the specificity of analyses that can be conducted. Key computations and transformations within the dataset facilitate a nuanced exploration of trends, player performance, and the impact of game dynamics on outcomes.

### **2.1 Dataset Characteristics:**

- **Data Segregation:** The dataset is divided into two main sheets: 'Men's Data' and 'Women's Data'. This segregation allows for targeted analysis of gender-specific trends and comparisons within the Australian Open.
- **Total Wins:** Calculated using the formula `=COUNTIF($C$2:$C$1000, C3)`, this metric is derived from the 'Champion' column, effectively quantifying the number of times a player has won. This serves as a primary indicator of success and dominance in the tournament.
- **Debut Year and Years Since Debut:** The debut year is extracted directly from the 'Year' column and is transformed into 'Years Since Debut' by subtracting the debut year from the current year (2024), providing a measure of experience and longevity in the competition.
- **Bad Start Analysis:** A custom formula `=(I2+K2)-(J2+H2)>-2` utilizes data from columns representing early match losses and wins to calculate instances of a 'Bad Start', where a player loses the initial sets but potentially recovers later in the match. This offers insights into resilience and recovery during match play.
- **Final Appearances:** The number of times a player has reached the final is computed with `=COUNTIF(C:C,C4)+COUNTIF(Q:Q,C4)`, using the 'Champion' and 'Runner-up' columns. This metric underscores consistency and peak performance in the tournament.

- **Win Rate:** Defined as  $\text{=(SUM(H2,J2,L2,N2,P2)/(SUM(H2,J2,L2,N2,P2) + SUM(I2,K2,M2,O2,Q2)))*100}$ , this calculation takes into account all match outcomes to produce a percentage that reflects a player's success rate, pivotal for assessing skill and effectiveness in the tournament.

## **2.2 Data Formats and Visualization Techniques:**

The dataset's structured format with numerical and categorical data supports various types of visual analyses:

- **Time Series Analysis:** Using 'Year' in a numerical format allows for time series visualizations to track changes and developments over time.
- **Categorical Comparisons:** Segregated data sheets facilitate comparisons across gender lines, revealing trends specific to men's and women's performances.
- **Correlation Analysis:** Calculations like 'Bad Start' and 'Win Rate' provide data for scatter plots and correlation analyses, showing how different aspects of game play influence overall success.
- **Categorical Comparisons:** Segregated data sheets facilitate comparisons across gender lines, revealing trends specific to men's and women's performances.
- **Performance Metrics:** Total wins and final appearances create opportunities for bar or line charts that highlight career achievements and tournament impacts.

## **2.3 Integration in Tableau:**

The dataset's format, with clearly defined columns and calculated metrics, fits seamlessly into Tableau's visualization capabilities. Linking sheets based on the 'Year' column enables the creation of dynamic dashboards that can display interconnected trends and perform cohort analyses across different time periods.

## **2.4 Advantages of Tableau:**

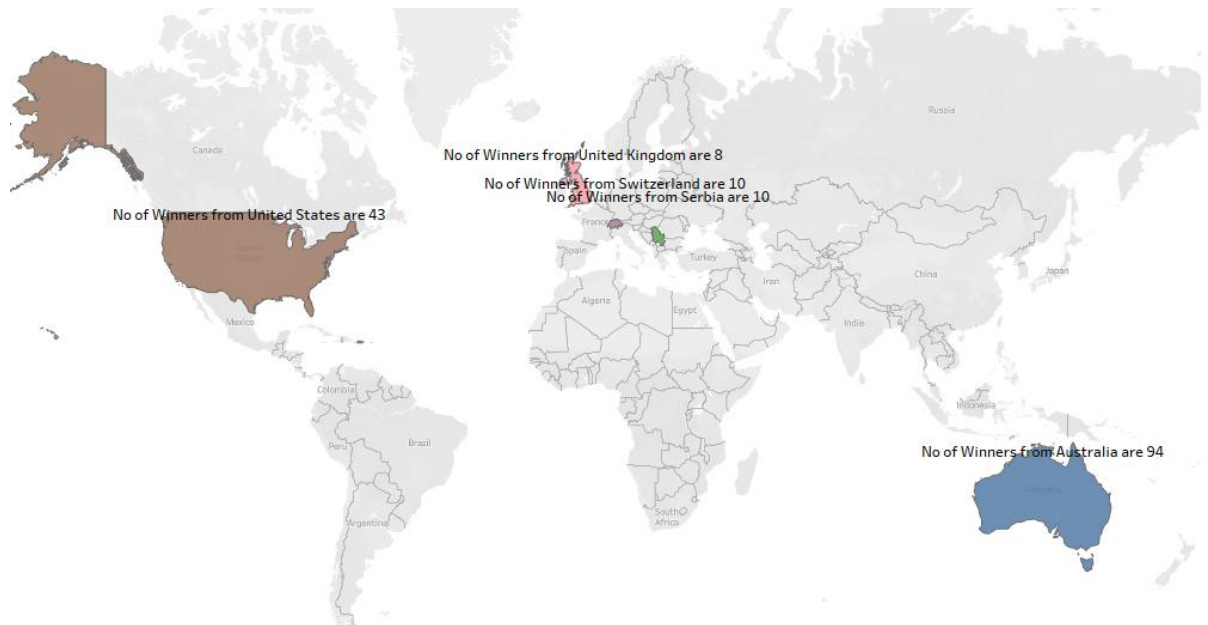
- **Interactive Visualizations:** Tableau's robust interactive features allowed users to explore data layers in more detail, providing a deeper understanding of trends over time.
- **Ease of Use:** With user-friendly interfaces, Tableau enables quick transformations of raw data into visually appealing and informative graphics.
- **Dynamic Data Handling:** Tableau efficiently handles large datasets, making it ideal for the comprehensive and varied data involved in this analysis.

## **2.5 Disadvantages of Tableau:**

- **Cost Concerns:** Tableau can be expensive, which might limit accessibility for smaller organizations or individual researchers.

- **Learning Curve:** While Tableau is user-friendly, mastering its more advanced features requires significant time and training.
- **Performance Issues:** Handling extremely large datasets or very complex visualizations can sometimes lead to performance lags.

### **3. Analysis of Winners by Country (Map Visualization)**



*Fig (1)*

#### **3.1 Dominance by Australia and the United States**

The map visualization highlights the clear dominance of Australia and the United States in producing Australian Open champions. As the host country, Australia has the highest number of winners, with 94 titles. This could be attributed to a combination of factors, including a home advantage, greater participation due to proximity and national interest in tennis, and well-developed sports infrastructure and training programs.

The United States also stands out with a significant number of 43 winners, reflecting the country's strong tennis programs and its ability to cultivate internationally competitive players. The success of American players at the Australian Open suggests that the United States has been a major force in global tennis development and talent production.

#### **3.2 European Representation and Success**

While Australia and the United States lead in the total number of champions, the map also reveals notable representation and success from European countries.

Switzerland and Serbia, each with 10 winners, have produced world-class players who have performed exceptionally well at the Australian Open and other major international competitions.

The United Kingdom, with 8 winners, also shows considerable success, although slightly less than Switzerland and Serbia. This could reflect varying levels of participation and development of tennis talent across different European nations over the years.

**3.3 Geographic and Cultural Implications**

The geographic distribution of winners highlighted on the map suggests that North America, Europe, and Australia have historically been the most influential regions in tennis. These areas have consistently produced a large proportion of the Australian Open champions, indicating the presence of well-established tennis traditions, infrastructure, and cultural support for the sport.

In contrast, the map reveals a visible lack of representation from African, Asian, and South American countries. This could suggest that these regions have experienced relatively lower levels of tennis development or participation, potentially due to a range of factors, such as limited access to resources, facilities, and coaching, as well as cultural differences in sports prioritization. The concentration of winners in specific countries and regions may reflect broader cultural and infrastructural factors that contribute to nurturing top-tier tennis talent. This includes investment in sports facilities, coaching programs, and the overall promotion of competitive sports within a country's culture and society.

**4. Emerging Talents (Bar Chart Analysis)**

The bar chart titled "Players with highest potential" provides valuable insights into the early career successes of recent players at the Australian Open, offering clues about their potential future trajectories in the sport.

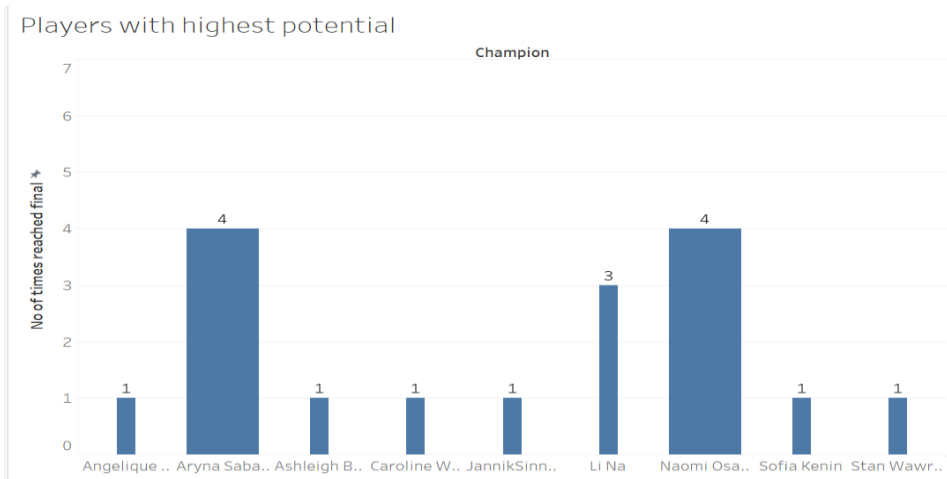


Fig (2)

## **4.1 Early Career Successes of Recent Players**

Two players stand out as dominant performers in the early stages of their careers: Aryna Sabalenka and Naomi Osaka. Both have reached the Australian Open finals on 4 occasions within the first 10 years of their professional debuts. This level of consistency and success at the highest level of the sport, at such a young age, is a strong indicator of their exceptional talent and potential. Another emerging talent highlighted in the chart is Li na, who has reached the Australian Open finals 3 times in her young career.

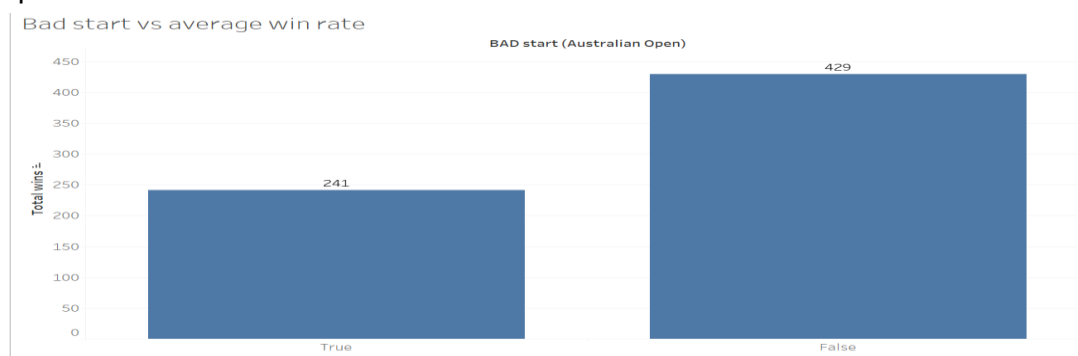
## **4.2 Discussion on High Potential Indicators and Career Growth Trajectories**

The players with multiple finals appearances in their first 10 years, such as Sabalenka, Osaka, and Li na, can be considered to have high potential for future success. Reaching multiple finals in a short span often correlates with exceptional skill, mental toughness, and the ability to perform under the immense pressure of Grand Slam events.

By comparing the current crop of emerging talents to the career trajectories of past players at similar stages, analysts and coaches may be able to better predict which of these young stars are most likely to dominate the sport in the coming years. Identifying the key traits and strategies shared by successful players can help guide the development of future tennis champions.

## **5. Resilience in Competition (Bad Start vs. Average Win Rate Bar Chart)**

The bar chart titled "Bad start vs average win rate" provides valuable insights into the resilience and recovery capabilities of players competing at the Australian Open.



*Fig (3)*

## **5.1 Statistical Analysis of Match Recoveries**

The data presented in the chart reveals a stark contrast between the performance of players who experienced a "bad start" (defined as losing the first two sets of a match) and those who did not.

Players who did not have a bad start accumulated a significantly higher number of total wins, 429, compared to those who did experience a bad start, with only 241 wins. This suggests a strong correlation between not losing the first two sets and achieving a greater overall success rate at the tournament.

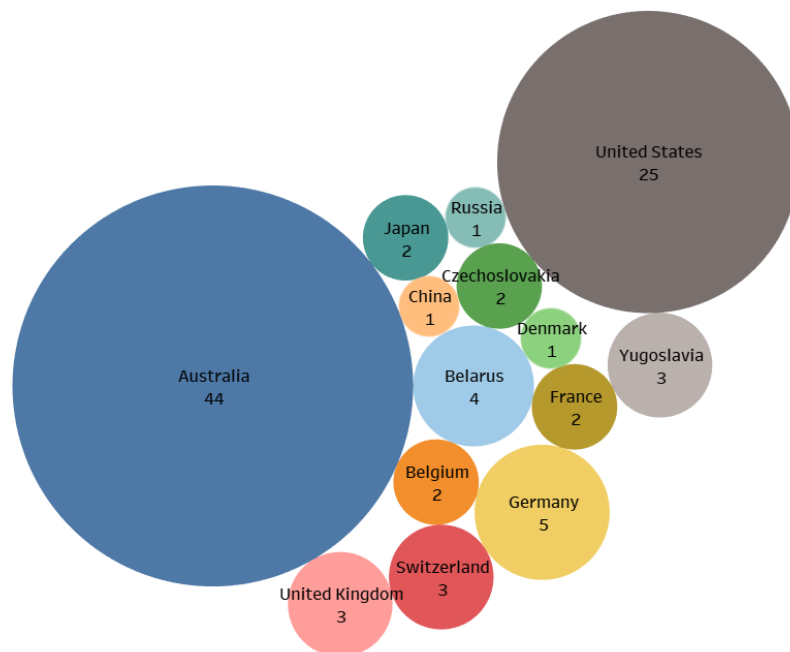
However, the fact that players with a "bad start" were still able to win 241 matches is a testament to their resilience and ability to recover from early setbacks.

## **6. Female Winners and Global Trends (Bubble Chart Analysis)**

The provided bubble chart offers valuable insights into the global distribution of female winners at the Australian Open tennis championships.

Dominance of Certain Regions in Female Tennis

Countries with highest female winners



*Fig (4)*

The chart clearly highlights the dominance of Australia and the United States in producing female champions at the Australian Open. Australia, as the host country, has the highest number of female winners, with 44 titles. The United



States also stands out with 25 female champions, reflecting the strong infrastructure and support systems for women's sports in these countries. This dominance is likely a result of a combination of factors, including effective training programs, access to high-quality facilities and resources, and a cultural emphasis on supporting and promoting women in competitive sports. The investment and development of women's tennis in these regions have paid dividends in the form of sustained success at the Australian Open.

### **6.1 Emerging Trends and Historical Implications**

While Australia and the United States lead in the total number of female champions, the chart also reveals an interesting diversity of representation from European countries. Nations like Germany, Belarus, and Switzerland have each produced multiple female winners, suggesting that tennis enjoys broad support and development across the European continent.

The inclusion of historical countries, such as Czechoslovakia and Yugoslavia, adds an intriguing layer of depth to the data. This reflects how geopolitical changes, and the dissolution of certain nations can impact the sports landscape, both in terms of participation and the recognition of achievements.

## **7. Decadal Dominance in Men's Tennis (Year on Year Dominance Bar Chart)**

The bar chart titled "Year on year dominance (men's)" provides a compelling visualization of the success of male tennis players at the Australian Open, highlighting the dominance of certain individuals over specific decades.

### **7.1 Analysis of Player Success Over Different Decades**

The most striking feature of the chart is the overwhelming dominance of Novak Djokovic in the 2010s and 2020s, with a substantial bar representing his 10 total wins. This underscores Djokovic's exceptional performance and consistent success in recent years, setting a new benchmark for excellence in men's tennis. In contrast, the chart also reveals the historical peaks of other legendary players. Roy Emerson stands out with 6 wins in the 1960s, cementing his status as a dominant force of his era. Similarly, Jack Crawford's 4 wins in the 1930s highlight his pre-war dominance.

## Year on year dominance(mens)

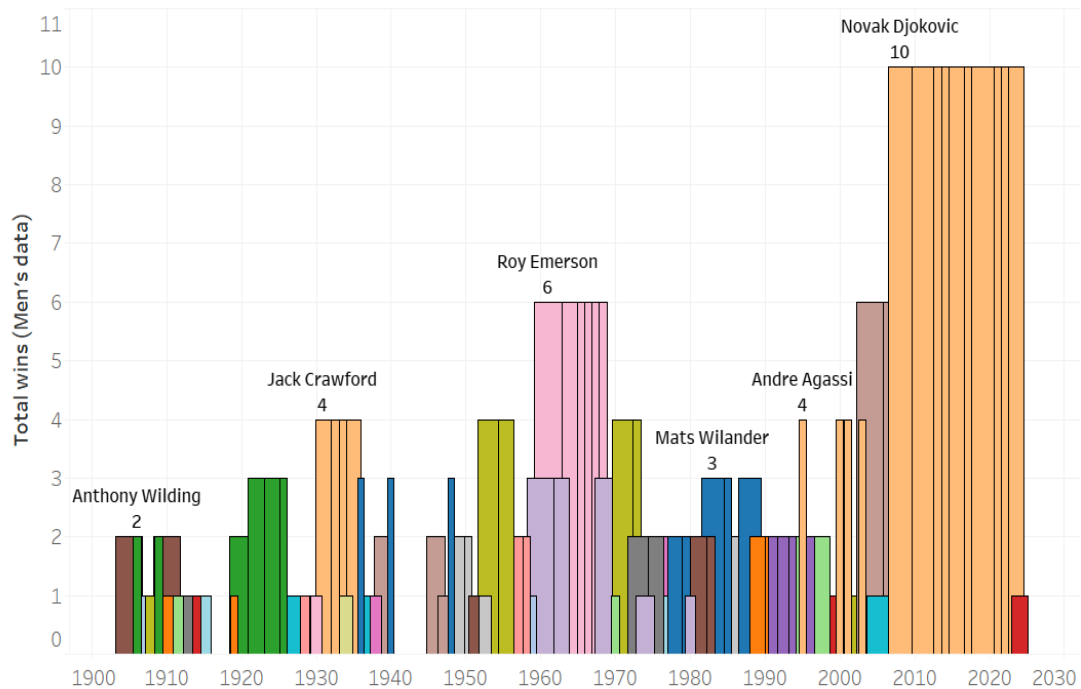


Fig (5)

## 7.2 Insights into the Evolution of Men's Tennis and Player Strategies

The chart effectively visualizes the shifting patterns of player dominance across different decades. From the early 1900s with Anthony Wilding, to the 1930s with Jack Crawford, the 1960s with Roy Emerson, and the sporadic peaks of players like Andre Agassi and Mats Wilander, the data illustrates the evolving landscape of men's tennis.

## 8. Longevity and Performance Trends (Years Since Debut vs. Total Wins Chart)

The chart titled "Years since debut vs total wins" provides a view of player performance at the Australian Open, highlighting the trajectory of success over relatively short-term careers (30 years) .

Years since debut vs total wins

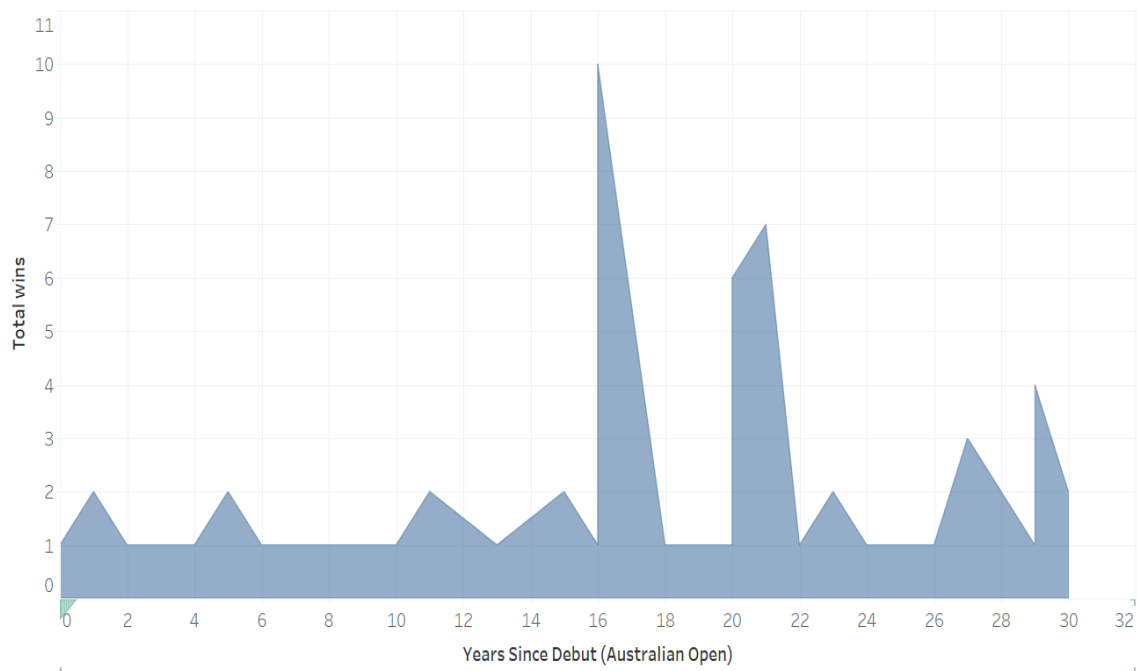


Fig (6)

### **8.1 Career Trajectories and Peak Performance Periods**

The data presented in this visualization reveals several key insights into the career trajectories of top tennis players:

**Early Career Peaks:** The chart shows prominent peaks in the total number of wins around the 6th and 16th years since a player's debut. This suggests that some athletes are able to achieve high levels of success relatively early in their careers, potentially due to strong initial training, early career momentum, or exceptional talent.

**Sustained Success:** Significant peaks later in players' careers, notably around the 16th year since debut, showcase their ability to maintain a high level of performance over many years. This demonstrates exceptional skill and longevity, which are particularly commendable in the physically demanding sport of tennis. These players have not only reached the pinnacle of success but have also managed to stay competitive for an extended period.

**Declines and Resurgences:** The visualization also reveals periods of decline in a player's win totals, followed by subsequent resurgences, such as those observed around the 10th and 20th years since debut. These fluctuations could be indicative of changes in competitive strategy or recovery from injuries.

**Late Career Wins:** Remarkably, the chart shows the presence of wins recorded up to the 30th year since a player's debut. This ability to maintain high

performance levels well into the later stages of a career is unusual in tennis and highlights the exceptional endurance and adaptability of these elite athletes.

## 9. Consistency of Top Players Over the Years

The "Consistency of Top Players Over the Years" chart provides a compelling visualization of the win rates of some of the most significant players in the history of the Australian Open. This data-driven analysis offers valuable insights into the patterns of success and the consistency of top performers across different eras and genders.

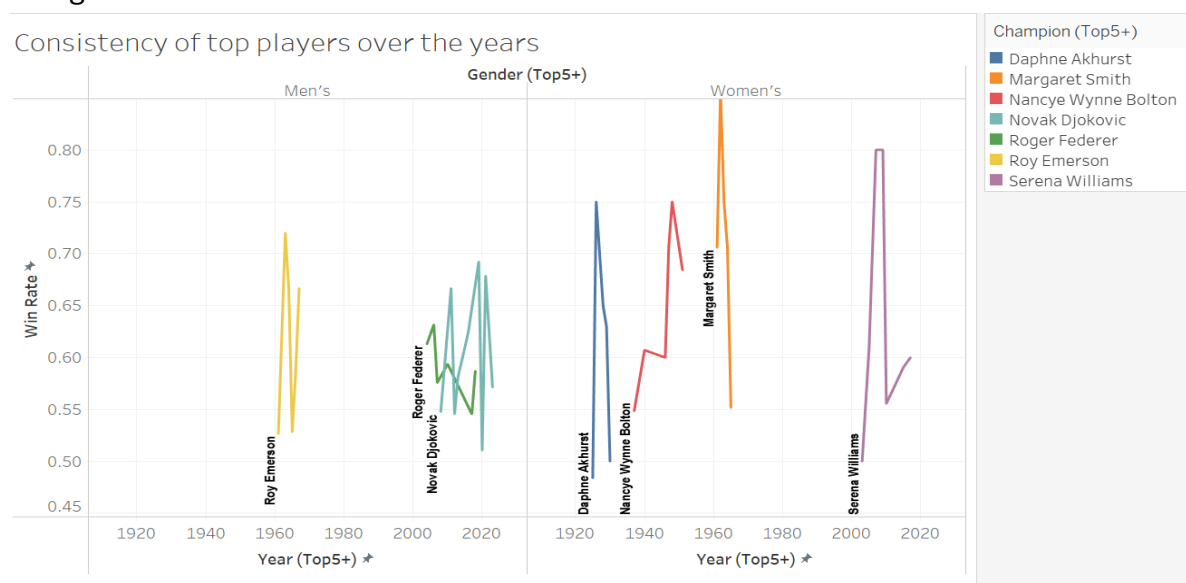


Fig (7)

### 8.1 Win Rates and Career Phases of Legendary Players

The chart reveals distinct trends in the performance of individual players over time:

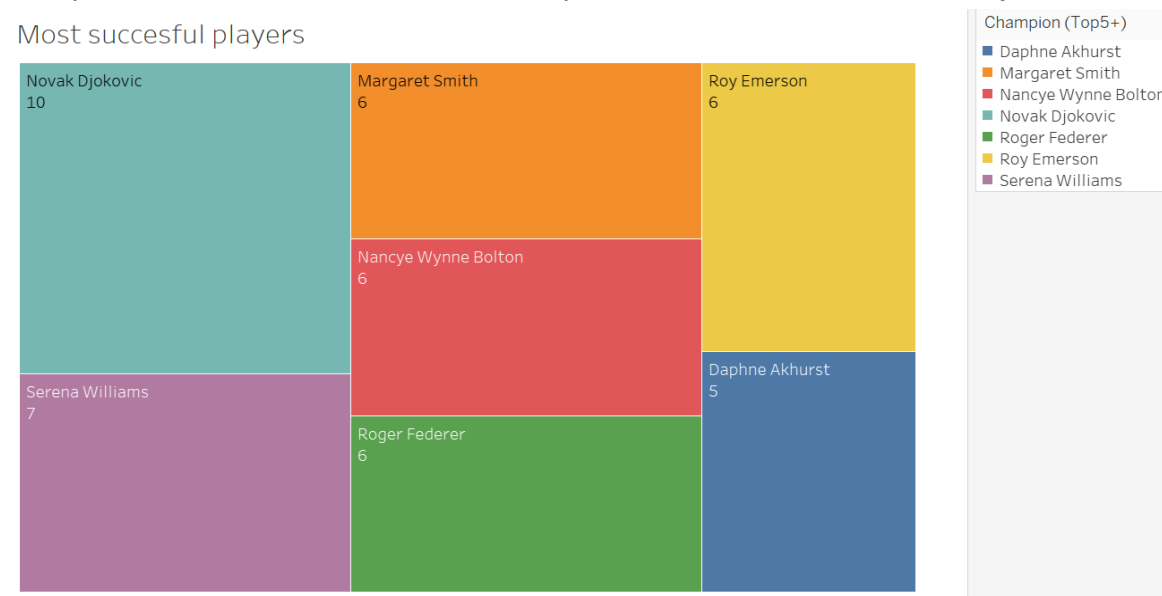
**Roy Emerson:** Emerson's win rates show a sharp peak in the early 1960s, indicating a period of dominance in the men's category. His exceptional performance during this time suggests he was able to achieve a level of excellence that was unmatched by his contemporaries.

**Roger Federer and Novak Djokovic:** These modern-era players display more fluctuating win rates from the early 2000s onwards. The multiple peaks for both Federer and Djokovic suggest their ability to achieve repeated successes and make comebacks at different stages of their careers. Djokovic shows a resurgence in win rate towards the latest years, reflecting his continued excellence and adaptability in the sport.

**Serena Williams:** Williams exhibits a striking spike in win rate towards the late 2010s, underscoring her dominant phase during this period. This is particularly noteworthy given the challenges of competing at an older age compared to many of her peers, highlighting her exceptional longevity and competitive spirit.

## **10. Legacies of Success (Most Successful Players Treemap)**

The treemap titled "Most successful players" provides a compelling visual representation of the total number of wins at the Australian Open for players who have won the tournament five times or more. This visualization effectively uses size and color to represent each player's success, offering an at-a-glance comparison of the most dominant champions in the tournament's history.



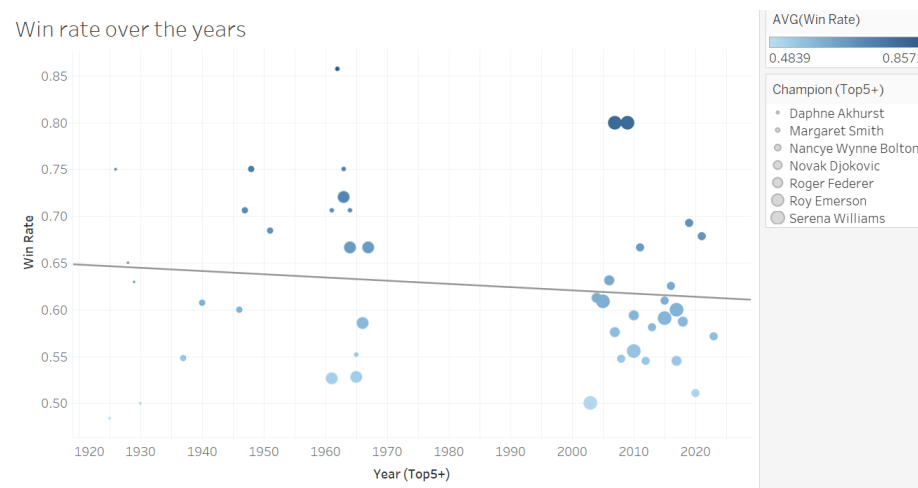
*Fig (8)*

### **10.1 Visual Representation of Tournament Victories**

The standout feature of this treemap is the overwhelming dominance of Novak Djokovic, who stands out with 10 total wins. The sheer size of his block in the visualization highlights his exceptional performance in recent years, cementing his status as one of the most successful players in the modern era of tennis. Alongside Djokovic, the treemap also prominently features historical greats such as Margaret Smith, Nancye Wynne Bolton, and Roy Emerson, each with 6 wins. The significant areas occupied by these players in the visualization underscores their historical importance and sustained success at the Australian Open. Serena Williams, with 7 wins, represents the most successful female player in recent decades, reflecting her outstanding career and profound impact on the sport of women's tennis.

## **11. Trends in Player Performances Over Time (Win Rate Over the Years Scatter Plot)**

The scatter plot titled "Win rate over the years" provides a visual representation of the win rates of top players (those with 5 or more wins) at the Australian Open across different years. This data-driven analysis offers insights into the evolving nature of competition and the strategies required for sustained success in the tournament.



*Fig (9)*

### **11.1 Analysis of Evolving Competition and Win Rates**

The trend line shown in the scatter plot indicates a slight decline in win rates over the years. This downward trend suggests that while the competition at the Australian Open has perhaps become more intense or the field more evenly matched, the top players' dominance may not be as pronounced as in earlier years.

The variability in win rates, ranging from just above 0.50 to nearly 0.85, highlights that while all these players have been successful, their level of dominance in terms of match-winning has varied considerably. The higher win rates, especially those approaching 0.85, suggest periods of extreme dominance by certain players during specific years.

### **11.2 Historical Context and Player Performance**

The scatter plot reveals some interesting historical trends. Earlier years (pre-1980s) show fewer data points, which could imply fewer tournaments or less consistent recording of data. The variability in win rates during these years is also quite pronounced, suggesting that a few players could have had periods of strong dominance.

In more recent years (2000s onwards), the number of data points increases, reflecting possibly better record-keeping and more consistent participation among top players. The clustering of win rates in this era around the 0.60 to 0.70

range suggests a more competitive environment where top players are winning fewer matches relative to earlier decades.

The scatter plot also highlights the consistent performances of specific players, such as Novak Djokovic and Serena Williams, who have multiple entries, indicating their sustained success at the highest levels of the tournament over many years.

## **12. Executive Summary**

The detailed analysis of the Australian Open tennis championships dataset over the past 120 years has offered profound insights into the evolution and historical dynamics of this premier tennis event. Utilizing advanced visualization techniques in Tableau, this report has effectively illustrated the trends and patterns in player nationalities, gender distribution, and the performance trajectories of tennis elites, thereby enriching our understanding of the tournament's impact and evolution.

### **12.2 Key Findings Summary**

**Dominance by Australia and the United States:** Both countries have demonstrated significant success in the tournament's history, with Australia benefiting from the advantage of being the host nation, and the U.S. showcasing the strength of its tennis programs.

**Emerging Talents and Veteran Performances:** Players like Aryna Sabalenka and Naomi Osaka have shown remarkable early career success, hinting at potential future dominance. Conversely, players such as Angelique Kerber and Ashleigh Barty illustrate that less immediate success in finals can still lead to significant overall career achievements.

**Resilience in Competition:** The analysis emphasizes the importance of strong starts in matches. However, it also highlights the resilience displayed by players who manage to recover from early deficits, a key trait in competitive tennis.

**Regional Dominance in Women's Tennis:** The dominance observed in women's tennis by players from Australia and the U.S. likely stems from robust support systems and infrastructure in these regions. Meanwhile, emerging trends in Asia and historical success in Europe reflect shifting dynamics.

**Evolving Landscape of Men's Tennis:** The varying dominance across decades, with recent years marked by Novak Djokovic's exceptional performances, contrasts with earlier periods where players like Roy Emerson and Jack Crawford led.

**Longevity and Performance Trends:** Insights into career trajectories show both early peaks and sustained success, with some players maintaining high performance deep into their careers.