Relationship between ATP Ranking and the Men's Tennis Grand Slam Winner in the Last 15 years

Fangyi Su

University of California, Davis CA，95618，USA

Abstract. In this project, I will analysis about the relationship between their real performance on the court and the ATP ranking, which ranks each player at the end of the year by calculating scores of players on the same scale, the records made by the Top3 players, and how the best tennis players are geographically spread worldwide in the last 15 years. In order to clarify the data related to the studies and conclude a clear conclusion, I use a data visualization tool called Tableau. After integrating the dataset, it enables me to conclude that there is a positive correlation between the players' performance and their real ATP ranking (The players that have higher ATP ranking are more likely to win the championship in Grand Slam). According to this comprehensive conclusion, we can predict the future winners of a Grand Slam according to the given ATP ranking for that year.

Keywords: Data Visualization, Sports Statistics, Tennis, ATP Ranking.

1. Introduction

In recent years, tennis has become one of the most popular sports around the world. Each year, the official tennis organization holds four Grand Slam tournaments worldwide. They are Australia Open (be held in January in Melbourne, Australia), the French Open (be held in the Mid of June in Paris, France), Wimbledon(be held in the Mid of July in Wimbledon, UK), and the U.S. Open (be held at the end of August, in New York, US). The increasing number of audiences gather together to witness the birth of a new winner of the Grand Slam. In order to evaluate the players' performance better, the ATP association published a crucial index to quantify the vacillation of their performance which is the ATP ranking. This ranking is also known as the world ranking, which ranks each player at the end of the year by calculating scores of players on the same scale. The ATP ranking will vary often depending on the players' performance because it uses the points system which counts on their best 16 results during a calendar year, for a total of 52 weeks following the result (Sell,2022). So, with the curiosity about how the ATP ranking can reflect the actual performance of players on the court, I started to collect data from Kaggle (one of the biggest online databases), searched for the data information of all of the records of the participants in Grand Slam and their ATP ranking in last 15 years (Wonduk,2022), downloaded them, and began with the data visualization.

In order to make panel data series (observing many observations multiple times by following the same indicators) to be more straightforward to the audience, I chose Tableau to do data visualization creating different graphs that helped to understand. Tableau always owns a wide range of chart types and methods to represent data that enable big data to be simpler and more insightful (Mitisha, 2021). By using this tool, I can easily draw an unbiased conclusion from the graph between two variables that can be easily recognized. Besides that, I can choose among different types of data (mean, median, standard deviation, etc) in order to represent the relationship between data in the best way. For example, when I want to choose the median of the ATP ranking of a group of players, Tableau helped me to calculate the mean of that series of data automatically which eases a ton of workload for me.

After reading the conclusion conducted from six graphs, the audiences will learn more about the relationships between the ATP ranking and the real performance of players in order to make a better prediction about the future winners in the Grand Slam. This project also provided a new point of view which led people to stay doubtful about the preciseness of some official index (like the ATP ranking in this study). Can the ATP ranking correctly reflect all the players' information on the court in that calendar year? Of course not. Many other factors may influence the players'

performance, like long-time suspension because of bad weather (external factors), players' injuries (internal factors), etc. So, those ignorances cause one of the biggest limitations in my studies. I need to improve this study to conduct more precise conclusions when more comprehensive datasets become available.

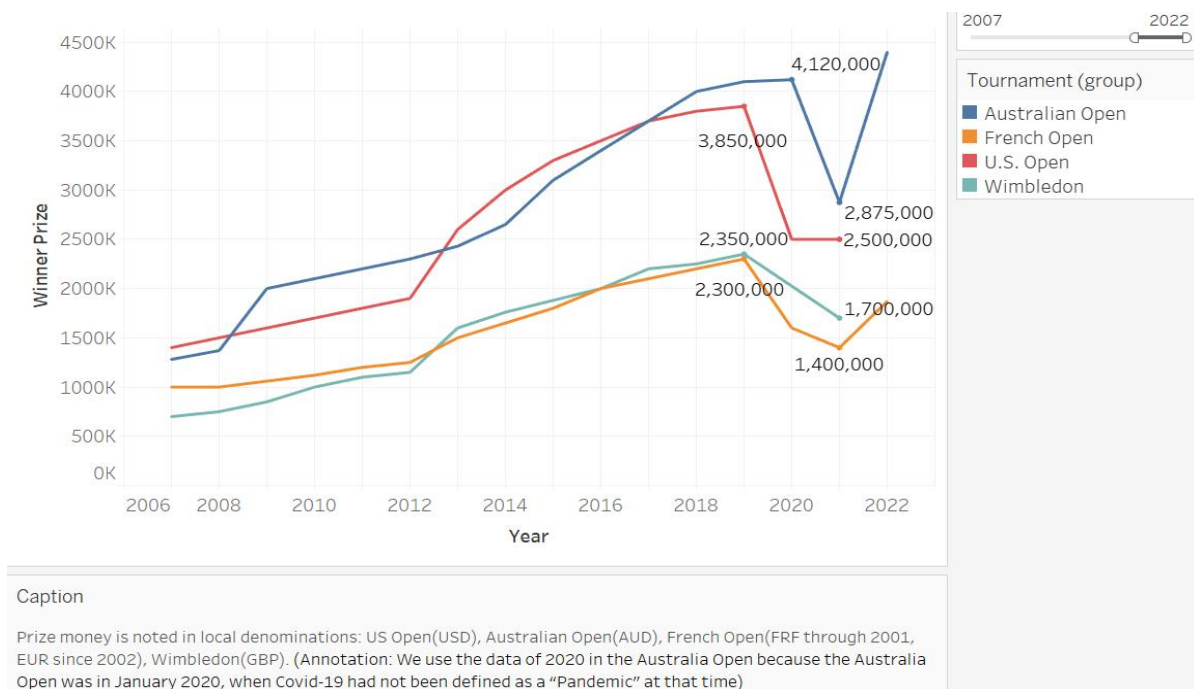## 2. Content

2.1 Change in Winner Prize

Fig 1. Annotation: We use the data of 2020 in the Australian Open because the Australian Open was in January 2020, when Covid-19 had not been defined as a "Pandemic" at that time

As shown in Figure 1.1, from 2007-2019, the winning prize for all four tournaments (Australia Open, French Open, Wimbledon, and U.S. Open) has an increasing trend. It reflects the increasing number of sponsors that are more willing to invest in the tennis tournaments which attracts more competitors to join the tournaments. In 2019, the winner's prize of all four tournaments reached their peaks respectively. During Covid-19 (2020.3-2021), the winning prize for each tournament decreased in large percentages.

- For the Australian Open, the winner prize decreased from 4,120,000 AUD to 2,875,000 AUD (decreased 30.2% after Covid-19).
- For the French Open, the winner's prize decreased from 2,300,000 EUR to 1,400,000 EUR (a decrease of 39.1% after Covid-19).
- For Wimbledon, the winner prize decreased from 2,350,000 GBP to 1,700,000AUD (decreased in 27.7% after the Covid-19).
- For the U.S. Open, the winner prize decreased from 3,850,000 USD to 2,500,000 USD (decreased 35.1% after Covid-19).

So we can conclude that the French Open was mostly influenced by the pandemic because of the largest reduction in the winning prize.

Somehow, the winner's prize reflects the values of that tournament because of the higher number of sponsorships. The amateurs can conclude which tournament is growing and more widely received worldwide by observing the increased rate of winner prizes of that. From this graph, we also can tell which tournament was mostly affected by Covid-19, and which lost the most.

2.2 ATP Ranking

2.2.1. Correlation between ATP Ranking and their Performance in the Tournaments

Caption
Higher ATP ranking indicates the better perfermance of players at that year.
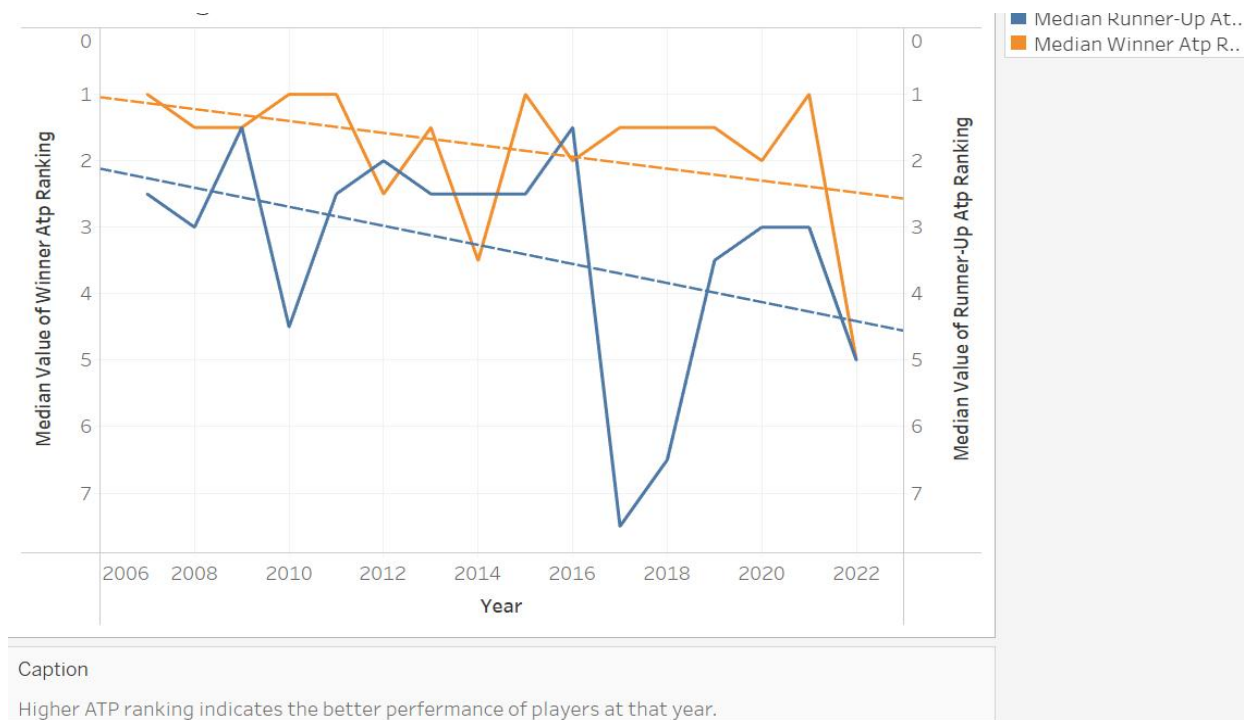
Fig 2. Annotation: The reason why using the median of data instead of average is that the median is less likely to be influenced by the extreme values (e.g: Runner-Up ATP Ranking in 2017)

The ATP ranking was one of the most important indicators to evaluate the tennis players' performance in the tournaments that year. Because it records as scores, more important tournaments (which are more competitive) value more scores. Then the ATP association ranks the players among scores. But if there is a positive correlation between the ATP ranking that can be transferred to the actual performance on the tournament courts is still a debatable question.

The trend line in Figure 2.1 describes the correlation between ATP ranking and the probability of being a winner in the tournaments.

According to the graph above, we can conclude that the median ATP ranking for Winners is higher than that of the Runner-up (the player who lost in the final). Additionally, we can observe that the difference between the median of the ATP ranking of Winners and the median of the ATP-ranking of Runner-up is increasing as the year goes on. Combining these two observations, we can conclude that the players who have higher ATP rankings are more likely to be the winner in the tournaments in future years.

After studying the correlation between ATP scoring and the probability of winning in the tournaments, the amateurs can predict the future winners in the tournaments by comparing their ATP rankings.

2.2.2 Geographically Spread Worldwide

After watching several tournaments, someone may be curious about why those competitive players all gather on the same continents in the world. The graph listed below shows how competitive players in tennis tournaments are geographically distributed worldwide.
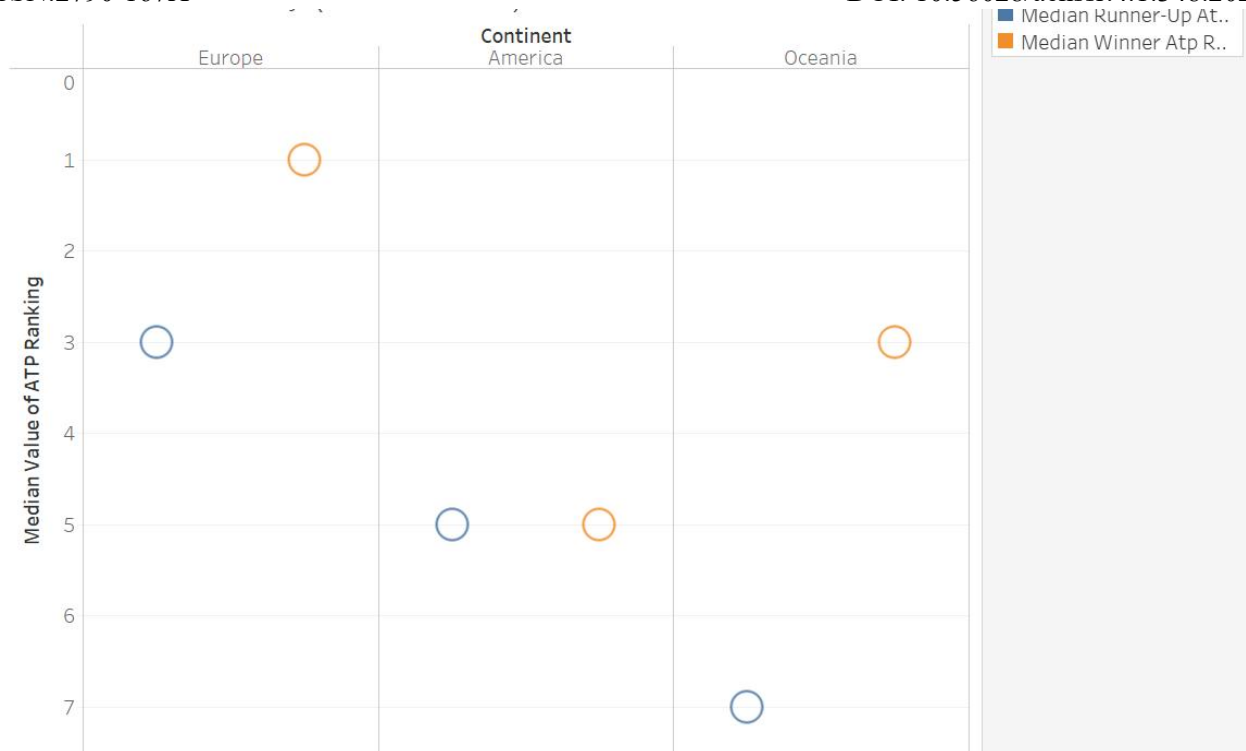
Fig 3. median

From the graph above, we can easily tell that the European players have higher median ATP rankings than the players from other continents, whether for winners or runners-up. So, we can conclude that European tennis players dominate worldwide which means they are more likely to win the championships in the following tournaments (combined with the conclusion from the previous conclusion).

In recent years, most men's winners of four Grand Slam tournaments come from Europe (like Novak Djokovic from Serbia and Rafael Nadal from Spain). Very few of them come from America and Oceania. The potential explanation for that phenomenon might be that higher-level training and higher-quality courts are more likely to be funded by governments of European countries. That helps the players grow better, earn higher ATP rankings, and win more championships in tournaments. So we can conclude that if the player comes from a European country, there is more likely to be the winner than the players from other countries.

3.   The Best Players in the Last 15 Years

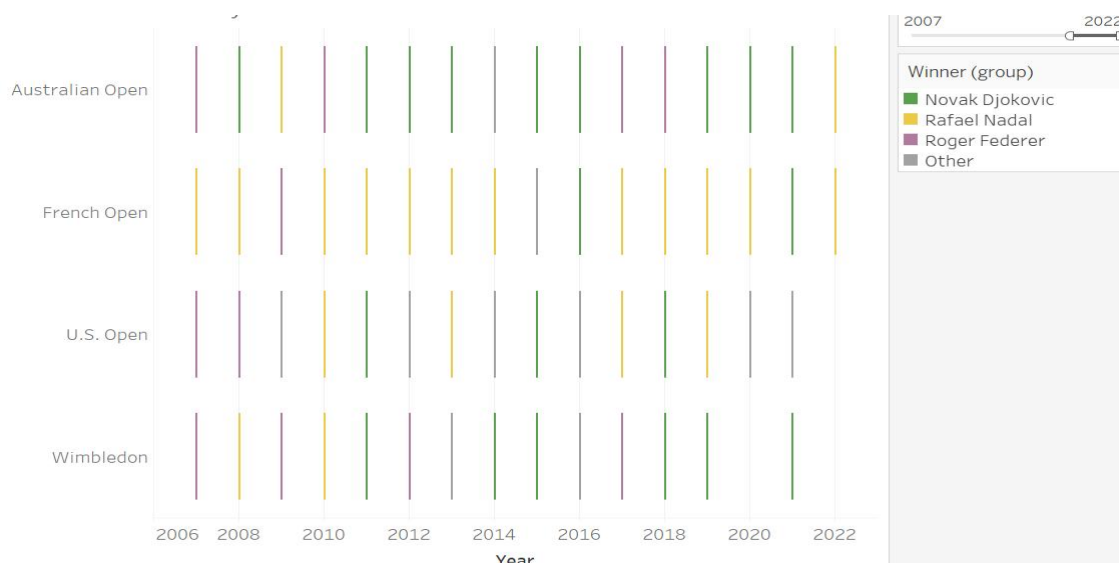### 3.1 The number of Winning in the Different Tournaments

Fig 4. graph

After observing the graph (Figure 4 above, we can see that purple, yellow, and green color (which represents Roger Federer, Rafael Nadal, and Novak Djokovic respectively) occurs in the highest frequency. So we can conclude that Roger Federer, Rafael Nadal, and Novak Djokovic are the Top3 men's tennis players in the recent 15 years.

After we categorize the data in the varieties of tournaments (Australia Open, French Open, Wimbledon, and U.S. Open), we can find that purple and green color (which represents Roger Federer and Novak Djokovic respectively) occurs frequently in different rows as the years pass. So we can conclude that Roger Federer and Novak Djokovic are more all-round players who can adapt to different kinds of tournaments (surfaces). However, the yellow color(which represents Rafael Nadal) occurs quite frequently in the single row(the second row which represents the French Open). It means that Rafael Nadal is more specialist in the French Open than the other two players.

Also, we can see how the frequency of color occurring changes as time passes. Between 2007-2012, the purple color (which represents Roger Federer) occurs the most. But in the recent decade, yellow and green colors (which represent Rafael Nadal and Novak Djokovic respectively) occur more frequently than purple colors(which represent Roger Federer). In the other words, Rafael Nadal and Novak Djokovic did better in Grand Slam Tournaments in the recent decade while Roger Feder did better in the first five years.

We use the frequency of color occurring to show the frequency of players' winnings in the tournaments. We also can make some predictions about future winners of tournaments.

After the observation, we can conclude that Novak Djokovic is a more all-round player than Rafael Nadal in recent years. Under this assumption, we predict that Rafael Nadal has a higher probability to win in the French Open than Novak Djokovic while Novak Djokovic is more likely to beat Rafael Nadal in tournaments other than the French Open in future years.

3.2 The Number of Winning on the Different Tournament Surfaces

Four tournaments also can be categorized in the varieties of tournament surfaces. On the different kinds of courts, the velocity of the ball reflection on the court will be totally different. Therefore, the International Tennis Federation (ITF)   classifies the surfaces as slow, medium-slow, medium, medium-fast, and fast according to the pace of the interaction between the ball and the court (Itftennis, 2022). So, the players may have different performances playing on different courts.

For convenience to study, we mainly classify the type of court into three categories.

The Australian Open and U.S. Open used the **Hardcourt** made of cement and bituminous.

The French Open used the court made by **Clay**.

Wimbledon used a court made of **Grass**.

The following graph (Fig 5) records how the winners were distributed among different kinds of tournament surfaces in recent 15 years.
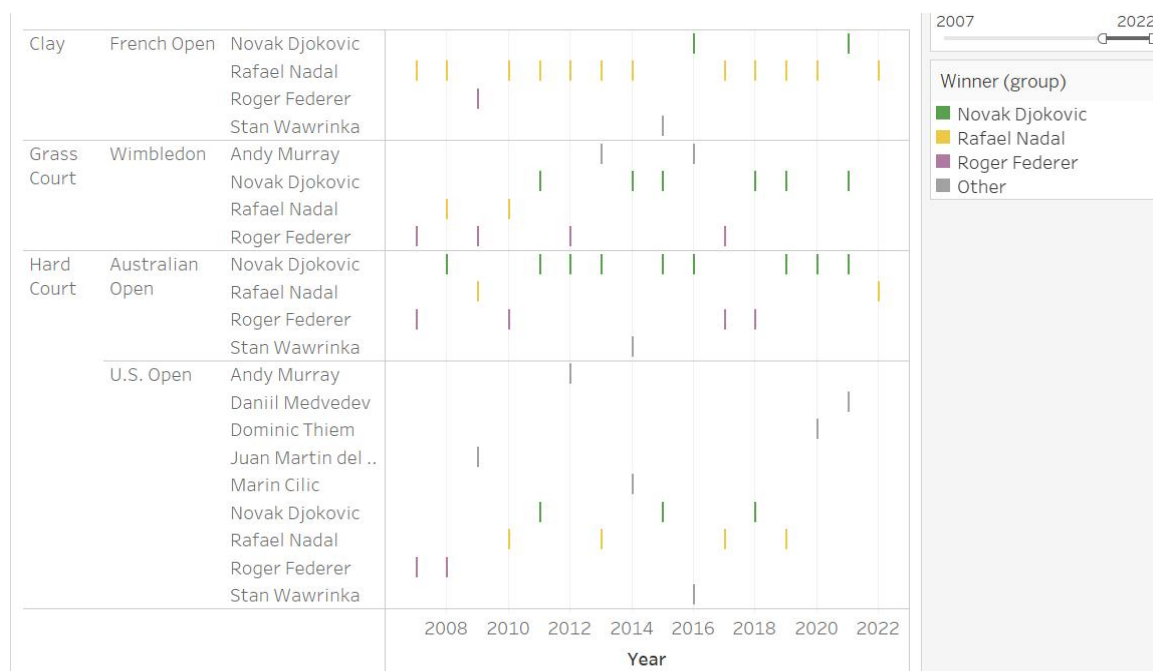

Fig 5. the winners were distributed among different kinds of tournament surfaces in recent 15 years.

On the Hard Court:

The green color occurs in the highest frequency which refers to Novak Djokovic as the best tennis player on the hard court (got 9 championships from the Australian Open and 3 championships from U.S. Open) in the last 15 years.

On the Clay:

The yellow color is quite a high frequency. It indicates that Rafael Nadal is the best tennis player on clay (got 12 championships from the French Open) in the last 15 years.

On the Grass Court:

Green color occurs in the highest frequency which refers to Novak Djokovic is the best tennis player on grass (got 6 championships from Wimbledon) in the last 15 years.

Overall:

Green color occurs in the highest frequency in total which refers to Novak Djokovic as the best tennis player of all of those kinds (got 12 championships on hard court, 2 championships on clay, and 6 championships on grass court) in the last 15 years.

According to the graph above, we can easily tell that there are only 4 different winners produced on both clay and grass in the last 15 years. But there are many different winners produced in hard-court tournaments. So we can conclude that it is more difficult for the players to get the championship on clay and grass courts (from French Open and Wimbledon).

Because of the properties of different kinds of surfaces, the player is adaptable to different velocities and fractions of balls on different courts. According to the reference above, the players who did well on hard courts may not do well on grass and clay courts but the winners of grass and clay court also did well on hard courts. So, when we predict the winner of tournaments, we can infer that the players who did well on the grass and clay court will do better in all tournaments.
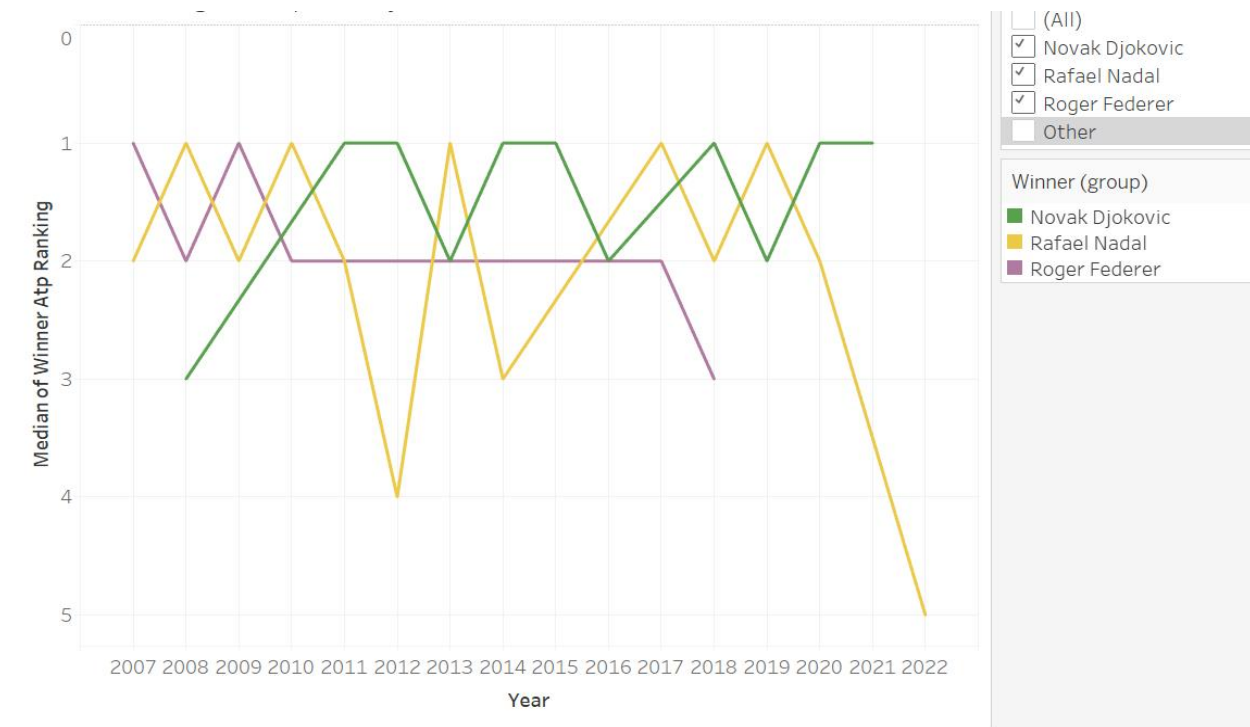
3.3 Comparison of the Top 3 Players

Fig 6 Novak Djokovic's ATP ranking in the last ten years.

From this graph (Fig 6), we can easily see that Novak Djokovic has had a high and stable ATP ranking in the recent decade. By contrast, Rafael Nadal has had a high but unstable ATP ranking in recent 15 years. And Roger Federer's ranking was high in the first ten years but dropped out of the Top5 in 2018.

Thus, after combining observations from this graph and the conclusion from Problem 2 (players who have higher ATP ranking are more likely to win the next tournaments), we can conclude that Novak Djokovic is more likely to win the next tournament than the other two players.

For players, there are many factors (e.g injuries) that may affect their ATP ranking that year. So when we evaluate the ATP ranking, comparing high and stable ATP rankings is a more efficient indicator (than just comparing high) to predict the winner in the next tournament.

4. Conclusion and Application

Based on data in recent 15 years, Winner Prizes of all four tournaments steadily increased before 2019, reduced a lot after that, and recovered recently. There is a positive relationship between Atp Ranking and their actual performance on the court because the winners always get a higher ATP ranking than the Runner-Ups. Among the Top3 players, Rafael Nadal specialized in playing on Clay, while Novak Djokovic is more specialized in playing on surfaces other than Clay. In the Chronicle aspects, Roger Federer dominated in the first five years, while Rafael Nadal and Novak Djokovic did better in the later 10 years. Last but not least, Men's winners of grand slams are geographically gathered in Europe, which has a better environment to train compared to the other regions in the world.

According to all the trends and statistical information listed above, it is easier to make some predictions about the future winners in the tournaments. Additionally, we also know how to train tennis players in the future (e.g sending them to Europe to accept good training resources). All in all, It provides really straightforward graphs to explain the correlation between ATP ranking and players' performance on the court. However, this project just takes into consideration some factors (like the properties of the tournament surface), and it ignores that some of the other factors may also

affect the players' performance like injuries. If it can consider those influential factors, it will be able to conduct more precise and comprehensive conclusions.

References

[1] Itftennis.com.   (n.d.).   Retrieved   September   26,   2022,   from https://www.itftennis.com/en/about-us/tennis-tech/recognised-courts/.

[2] Mitisha. A (2022, May 31). 4 reasons why Tableau is the best tool for Data Visualization. Zehntech. Retrieved   September   26,   2022,   from https://www.zehntech.com/4-reasons-why-tableau-is-the-best-tool-for-data-visualization/.

[3] Sell, K. (2022, February 8). How do tennis rankings work? (easy guide). My Tennis HQ. Retrieved October 31, 2022, from https://mytennishq.com/how-do-tennis-players-earn-ranking-points/

[4] Wonduk. (2022, June 5). Men's Tennis Grand Slam winner dataset. Kaggle. Retrieved September 26, 2022, from https://www.kaggle.com/datasets/wonduk/mens-tennis-grand-slam-winner-dataset.