JOURNAL OF
EMERGING INVESTIGATORS

Article

Optimizing tennis strategy: a data-driven analysis of point importance

Aditya Singla[1], Anurag Singla[2]
[1] Cupertino High School, Cupertino, California
[2] Software Engineering, Intuit, Mountain View, California

SUMMARY

While many tennis players focus on developing technical skills such as strokes, footwork patterns, and serving/receiving, it's equally important to place enough attention on the strategic and mental aspects of the sport. It's crucial to develop a comprehensive approach to the game that encompasses both technical proficiency and strategic decision-making. This research project focused on winning and losing probabilities from different scores to analyze the importance of winning individual points in a tennis game. The guiding research topic for this study addresses how winning or losing at different scores of a tennis game affect one's chance of winning the game, and how the importance of different scores relate to one another. We hypothesized that the most important score of a tennis game would be 30:30 (server score: receiver score), and that scores which are closer and occur later in the game have higher importance relative to other scores. In the analysis, 30:30 turned out to be second most important point, and the overall data moderately supported the hypothesis as points which were representative of close and later-occurring scores were generally more influential. However, the main takeaway was a new insight, that points which demonstrated the receiver to be winning, such as 15:40 and 30:40, generally had the greatest effect on the outcome of the game. Using the results of this project, players can realize which points are scientifically most important to win so that they can save essential tactics such as a wide serve or defensive receive for crucial points.

INTRODUCTION

In the world of tennis, aspiring professionals engage in rigorous daily practice, striving for continuous improvement and aiming to ascend to higher levels of play. In addition, it is essential to consider the strategic aspects of the sport. Learning about the importance of winning specific points relative to the importance of other points can be pivotal in improving the performance of players around the world. Using this quantitative information, players can make informed decisions about which points are most beneficial to win and call for essential tactics such as a wide serve or defensive receive which has been working for them throughout the match, based on the playing style of the opponent (1, 2). This will lead players towards making the most value out of strategies. Furthermore, players can learn which points are less important and more suitable for trying new things and varying tactics.

During a tennis game, a player reaches a score of 15 after winning their first point, a score of 30 after winning their second point, a score of 40 after winning their third point, and wins the game after winning their fourth point unless they have not won by two points. For example, a player cannot directly win the game after winning a point from a score of 40:40 but can directly win if they are leading 40:30 (all scores are stated in the format of server score:receiver score throughout this manuscript). The player who serves alternates each game, with the winner of each game receiving a point in the set. Sets continue until one player reaches six games and wins by at least two games, with tiebreaks occurring at 6-6 if necessary. Usually, professional men's matches are played in a best of three (first player to win two sets) or a best of five (first player to win three sets) format (3).

This research explored the question of which points are mathematically most important towards winning a tennis game at the professional level by studying the change in expected win/loss percentages based on the outcomes of the given points. The purpose of this study is to provide players with statistical insight regarding which points are most important to win so that players can center their strategies based on this knowledge. We also analyzed the impact of a player serving or receiving, as having the serve is generally thought of as an advantage, especially at a professional level (4). Although there has been very little academic research on this topic, it is generally and reasonably inferred that points which come up later in a game such as 30:30 and 30:40 have more weight than earlier points such as 15:0 (5). This is because there are still many points to be played which can affect the overall winner at the start of the game, whereas points towards the end of the game have a more defined impact on the outcome of a game. We hypothesized that 30:30 would be the most important score, that scores which occur later will have a greater impact on the result, and that the role of serving can influence the ranking of scores. For the purposes of this research, we solely focused on the points which occur in each game. Ultimately, the results demonstrated that points which occur later in a game tend to have a somewhat greater overall impact on the outcome of the game. We also found that points that represent a close score are partially more important than points that represent a lopsided score. Finally, the data signifies the immense extent to which the serve is a powerful tool, as we saw exceedingly higher probabilities of winning from all points if a player had a serve than if they didn't. Overall, this research provides a novel analysis of tennis scores, which can be used effectively by players, and serves as an example of how to apply such analyses to sports in general.

https://doi.org/10.59720/24-370

RESULTS

We formed our results through analyzing matches with individual game-scoring granularity, provided in Flashscore—a website with detailed score records for many sports including tennis—by evaluating outcomes of the game given the outcomes of individual scores (6). Through these results, we quantified the importance of scores in winning a game. The importance of a score is measured through the delta percentage calculation — the delta between the probability of winning the game given winning at that score and the probability of winning the game given losing at that score (**Figure 1**). When referring to "importance", the points are equally important to both players as the delta percentage is equivalent for both players; for example, for the score 40:30 which has a delta percentage of 32.48%, both players have a 32.48% greater probability of winning the game if they win the given point than if they lose it.

If we consider two players to have an equal chance of winning any given point, then winning at 40:30 gives a 100% probability of winning a game while losing at that score gives a 50% probability. Therefore, winning or losing at this score affects your probability of winning the game by 50%. If we consider the same scenario but look at the point 0:0, winning at this score gives a 65.625% probability of winning the game while losing at this score gives a 34.375% probability, resulting in a lower delta of 31.25%. The following calculations can be done through binomial probability by counting all possible outcomes of picking a win for the player leading 15:0 or for the other player for the next 5 points. It can be seen that 16/32 cases represent a win for the leading player while 10/32 cases represent an equal position at 40:40 at the end of these points. We defined "importance" by this delta percentage, calculated as the difference in a player's chance of winning the game if they win the point versus if they lose the point. Logically, points with greater delta percentages are regarded as more important.

This study was focused on closely contended matches because we believed that strategic decision-making based on score analysis has a greater probability of changing the outcome of the match if both players' skill level is similar on the day of the match.

We ordered all the scores, from least to most important, as 40:0, 30:0, 40:15, 15:0, 0:0, 0:40, 30:15, 40:30, 15:15, 0:15, 15:30, 0:30, 15:40, 30:30, and 30:40 (server score:receiver score) (**Figure 2**). This ranking illustrates how scores are generally more important when a server is losing as opposed to when a server is winning. This can be seen as the four least important scores are cases in which the server is leading and five of the six most important scores are cases in which the receiver is leading.

Additionally, early game scores such as 0:0, 15:0, and 0:15 do indeed have less importance than later scores which are close such as 30:30 and 30:40 although this correlation is weak. Across all these scores, the server has around a 60-65% probability of winning the given point. Another hypothesis we had was that winning at close scores would be more important than winning at less close scores because it is more likely for scores to go in either direction when they are close. This is moderately shown but not to the extent which was expected due to the effect of points in which the server has a losing score (**Figure 1**).

Interestingly, 0:15 ranks amongst the more crucial points despite it being one of the earliest scores. Also 0:30 was proven to be a very influential point despite its bias towards the receiver. On the other hand, a very close score which occurs at a later point but depicts the server to be winning, 40:30, only ranks at the middle in terms of importance. Meanwhile, if the score is flipped to 30:40, depicting the server to be losing, the criticality of the score is drastically increased to the most important score.

It can be seen through scores such as 0:15 and 15:30 that even though the server may be at a lower score, they are

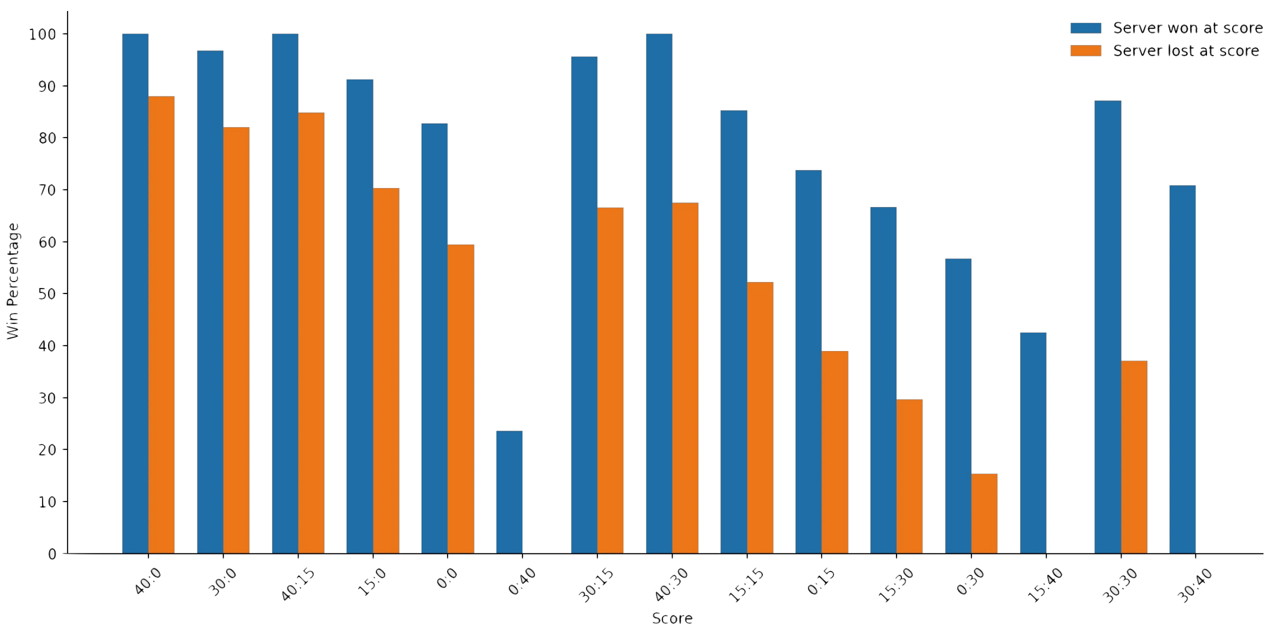

Figure 1: Probabilities that server wins game given winning versus losing at given score. Scores are presented in the form (server score:receiver score) and ordered in terms of delta percentage. The blue bars demonstrate the probability that the server wins the game if they win at the given score, while the orange bars demonstrate the probability that the server wins the game if they lose at the given score.

https://doi.org/10.59720/24-370

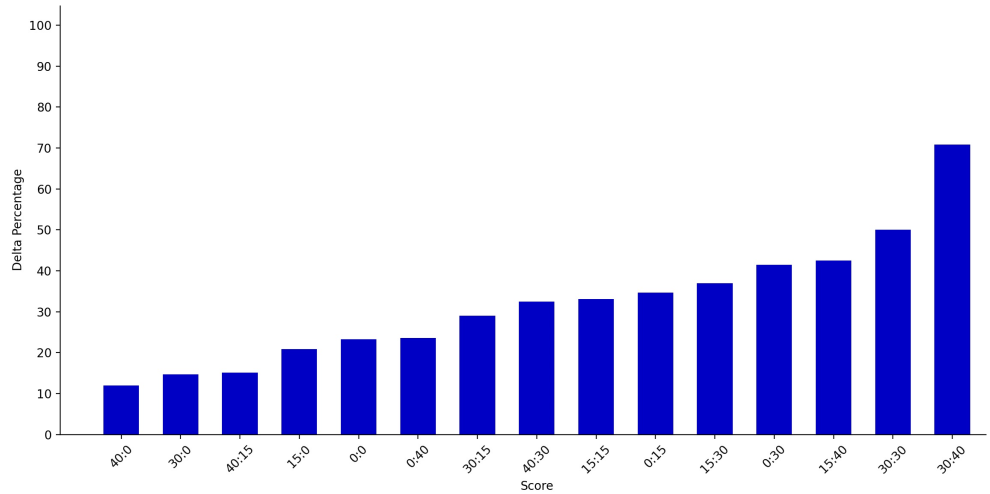

Figure 2: Delta percentages between the two win probabilities for given scores. Scores are presented in the form (server score:receiver score). The bars represent the corresponding delta percentage value for each of the scores. The delta percentage is calculated as the absolute difference between the two win probabilities; this is equivalent to the difference between the orange bars and blue bars in Figure 1.

still the predicted winner of the game. Then, if these points become 15:15 and 30:30 respectively, the probability for the server to win increases to 65-70%. Scores such as 40:0, 30:0, and 40:15, in which the server is winning, are relatively unimportant; meanwhile, 0:30, 30:30, and 30:40, in which the server is losing, are very important.

DISCUSSION

The study's central question is to determine the points that have the greatest impact on whether one wins or loses a game. We hypothesized that closer scores and later-occurring scores would have the greatest importance; these claims were evidenced to some extent by the results, but our main finding was that points in which the receiver was winning generally have exceeding importance than the points in which the receiver didn't have a greater score.

The ranking of scores from least important to most important: 40:0, 30:0, 40:15, 15:0, 0:0, 0:40, 30:15, 40:30, 15:15, 0:15, 15:30, 0:30, 15:40, 30:30, and 30:40 shows that closer scores and later-occurring scores are generally higher in importance but especially, that points in which the receiver is winning are most important. From these statistics, we can see that it is very beneficial to win points when the server is losing. This is likely because the versatility, control, and quality of the serves at the top level usually make the server the expected winner of the given game; if the server is already winning, it seems unlikely that the server will lose regardless of the outcome of the next point because they are not in a difficult situation. Because of the drastic advantage of the serve, there are many points which are not considered close in score or late in the game that have a relatively high importance because they are cases where the server is losing, creating a greater opportunity for the receiver to win the game.

The best-case scenario for a receiver if the server has a higher score than the receiver is to tie the score with the next point. Even in this case, the server is still the expected winner. This example illustrates why the delta percentage is low when the server is leading because regardless of the next score, the server has a great likelihood of winning the game. This is further proved by comparing our calculation for delta percentage if the two players are equal in skill level and our experimental data: the delta percentages were 50% and 32.48% for our calculated and experimental data respectively. Neutral scores, which are 0:0, 15:15, and 30:30, are distributed across the ranking based on how early they occur in the game with 0:0, 15:15, and 30:30 ranking 13th, 9th, and 2nd respectively. Neutral points more clearly demonstrate our hypothesis about the relative value of early and late points, as although all these scores are equal, winning at 30:30 is far more valuable than winning at 0:0. Looking at the average placements of the neutral points in the ranking, it is unsurprisingly seen that they rank around the average of all scores. This is because scores in which the server is winning have less importance than neutral scores while scores in which the server is losing have more importance than neutral scores.

Points in which the receiver is winning rank higher for importance. This is likely due to the fact that further extending the receiver's lead due to them winning the next point will put the server in a difficult situation, making it more probable that the receiver will be able to cause an upset against the serve; on the other hand, a receiver losing the next point while having a lead will either lessen or nullify the lead, making it far less probable that the receiver will be able to defend against the serve. This finding may be biased by the characteristics of our data used, given that we used professional men's matches, in which the player who lost the match won at least one set.

One major takeaway for tennis players based on the results of this analysis is that it is vital to win the point at 30:40, given that it is the foremost score and has a delta percentage that is over 20% than those of all other scores (**Figure 2**). The other points which are notable in importance relative to all the scores are 0:30, 15:40 and 30:30. Winning any of these points results in gaining more than a 40% probability of winning the game compared to losing at these points. The least important scores are 40:0, 30:0, and 40:15 each with delta percentages under 20%. Points which have relatively low importance can be strategically used as occasions to test new tactics against the opponent, alter one's playstyle, or try out a new shot. However, relatively important points may deserve the

https://doi.org/10.59720/24-370

implementation of an essential strategy and could be worth saving special tactics for.

It's necessary to address that these results can only be applied to Association of Tennis Professionals (ATP) playing standards because factors such as strategy and effectiveness of serve may vary according to skill level. It is also important to note that because only men's, professional, and close matches were analyzed, the findings may not be directly applicable to other matches which don't meet these criteria. Furthermore, the statistics will not align perfectly with each player on the individual basis as they serve as general averages for all players (7). For example, players' levels may be different and tennis games are heavily reliant on momentum which can mean different scores have a different psychological impact on players. Finally, we determined importance by finding the absolute value between the probability of winning the game if the given point is won versus if it is lost but there are other valid methods which can be used to rank importance. For example, instead of using delta percentage between winning and losing the point, importance can be ranked by measuring how much the probability is increased from winning the point compared to the original situation.

Further research in this topic can look at topics such as which games are most important to win a set, which strategies are most reliable for winning points, and what psychological strategies players can implement to improve their cognitive performance. Different studies may also use different characteristics for their data: for example, women's matches which aren't filtered based on closeness in the score may produce different results. It would also be useful if there were more projects done in this specific topic to cross-check the probabilities found in this study.

MATERIALS AND METHODS

Data collection

To collect the data for individual games, we found the point-by-point scores of professional matches from Flashscore (6). Data was obtained from the Flashscore website because it is one of few reliable websites containing not only match results, but detailed scoring results for individual games. The timeframe of matches used in our analysis was during the middle of 2024. Men's matches were only picked from ATP tournaments, which represents the tournaments conducted by the official, international men's tennis organization.

A total of 1,689 games were collected across 58 matches. To minimize the effect of other variables, only matches which lasted three sets in "best out of three" matches and matches which lasted more than three sets in "best out of five" matches (signifying a close match) were used, an equal distribution of player rankings was used, and only men's matches were used. Other than filtering by player level, gender, or closeness of the match, we didn't consider any other factors including players' countries, individual players, or locations of the tournament.

Because the matches which provide the data are close, and therefore, provide at least 3 sets of data which each contain 6 – 12 games and an average of 9 – 10 games, data for 1,689 games can be obtained through just the 58 chosen matches. To verify that player rankings were evenly distributed, we used a similar number of players rankings in each range of 100 (1 – 100, 101 – 200, ..., 901 – 1000).

Data analysis

The dual bar graph shown demonstrates the probability of winning a game given winning versus losing at a given score (**Figure 1**).

Delta percentage values are shown in increasing order in the figure based on the difference between the two win probabilities (**Figure 2**). **Equation 1** was used to calculate delta percentage values:

$$\text{Delta Percentage} = p(Win \mid Won\text{-}at\text{-}score) - p(Win \mid Lost\text{-}at\text{-}score) \quad \textbf{(Equation 1)}$$

Delta percentage was used to determine importance because it gives the most direct value of how much the point changes the probability of winning the game based on its result. Points with high delta percentages are those which are most likely to change the outcome of a game while those with low delta percentages are unlikely to change the outcome of a game; therefore, points with high delta percentage values should be treated as more important than those with low delta percentage values.

We calculated the results by finding the presence of a given score and counting how many games from the input were won and lost, programmatically. We ranked the importance by calculating the difference between the probability of a server winning the game given they win at the score being studied and the probability of a server losing the game given they lose at the score being studied.

We used the formula:

$$p = p \pm Z\sqrt{(p(1-p))/n} \quad \textbf{(Equation 2)}$$

with a confidence of 95%, and therefore, a Z-value of 1.96 to calculate the confidence interval ranges. Specific points have around 600 games of data on average and for the sake of simplicity, around 300 games per sub-case (server wins or server losses), giving us a value of n as 300. On average, the absolute difference between p and 50% was 25% for sub-cases without guaranteed results (0% and 100%), giving a value for p(1 - p) to be around 3/16. Plugging these estimated values into the formula gives:

$$p = p \pm 5\% \quad \textbf{(Equation 3)}$$

It is important to note that this confidence interval isn't accurate for all sub-cases and points, as they have a variety of values for p and n based on the data provided. Overall, this estimated of range of tells us that the results found are quite close to what is the true percentage across all matches ever played, at least among games which follow the selection criteria used for this study.

All the lines of scores which represented games were adjusted to fit (server score:receiver score) format by using a program which identifies the presence or absence of the phrase "BP" (the Flashscore website writes "BP" next to points which are game points for the receiver also known as break points) to determine whether or not that line needs to be altered. From this point, the analysis was carried out by using a program that detects which lines contain a given score and a program that counts how many of the inputted lines cause a win and loss for the server. To incorporate the Flashscore data, data from matches were copied into a document and then filtered programmatically, ensuring us that we were

https://doi.org/10.59720/24-370

left with just the raw lines which correlated to data from the games (8). First, we found all occurrences of the given score using the program which detects all lines with the score; then, we inputted those lines back into the same program looking for specifically the score which represents that point won for the server and won for the receiver. This process gave us the lines separated in terms of whether the point was won or lost by the server at the specific score. Finally, we inputted both sets into a program which looks for "BP" at the end to determine if the server or receiver won the set. This gave us a count which was used to calculate the percentage.

Received: October 31, 2024
Accepted: April 27, 2025
Published: December 21, 2025

REFERENCES

1. "Home of the Ultimate Tennis Lesson." *Fuzzy Yellow Balls,* https://www.fuzzyyellowballs.com/. Accessed 4 Apr. 2025.
2. Singles Strategy." *George Mason University,* https://mason.gmu.edu/~cmanni11/375/singles-strategy.pdf. Accessed 6 Apr. 2025
3. "Tennis Explained: Learn the Game: ATP Tour: Tennis." *ATP Tour,* https://www.atptour.com/en/news/tennis-explained-learn-the-game. Accessed 7 Apr. 2025
4. Cazzaro, Andrea. "Understanding the Importance of First Serve in Tennis with Data Analysis." *Medium*, Towards Data Science, 5 July 2020, https://medium.com/data-science/understanding-the-importance-of-first-serve-in-tennis-with-data-analysis-4829ab088d36. Accessed 16 Oct. 2024.
5. Cotet, Cosmin. "Not Every Point Is Equally Important in Tennis." *New York Tennis Magazine*, https://newyorktennismagazine.com/article/not-every-point-equally-important-tennis. Accessed 16 Oct. 2024.
6. "Tennis Live Scores, Tennis Results - ATP, WTA." Tennis Live Scores, Tennis Results - ATP, WTA | Flashscore.com, https://www.flashscore.com/tennis/. Accessed 3 Aug. 2024.
7. Over, Scott and Peter O´Donoughue. "Analysis of Strategy and Tactics in Tennis." *ITF Coaching & Sport Science Review*, https://itfcoachingreview.com/index.php/journal/article/view/624. Accessed 9 Apr. 2025.
8. Singla, Aditya. "GetScores.ipynb." *Cupertino High School*, https://github.com/adityasingla122/tennis-analytics/blob/main/GetScores.ipynb. Accessed 25 Jul. 2024.