

# Match Point: Predicting Outcomes of Hypothetical Tennis Matches Between Top 10 Ranked Players

Cayden J. Tu

Saint Ignatius College Preparatory High School, 2001 37th Ave. San Francisco, CA 94116, U.S.A.; cayden.y.tu@gmail.com

**ABSTRACT:** Tennis' popularity, along with its global viewership, is increasing. However, there has not been an abundance of analysis surrounding predictions of player matchups or a focus on significant statistics affecting game play in tennis. This paper describes a program that can determine the probability of a player winning against his opponent based upon the list of selected top-performing players. Moreover, this paper provides the logistic regression model used to calculate these winning probabilities; this summary demonstrates which variables were significant in predicting the outcome of a tennis match. For data, players were selected from the top 10 performers in men's tennis from 2015-2019. The individual and match statistics for the top ten players were utilized in developing this model; data was taken from the Grand Slam Tournaments.

The model created is a logistic regression model. A logistic regression model was deemed to be the most optimal method as it is designed to predict probabilities for a binary outcome. This model was successfully able to create predictions for the percentage that any of these players was able to beat any other player that was in the dataset. Overall, this model was successfully able to gather information about which players would have the highest chance of winning a tennis match and the most important factors that lead to this outcome.

**KEYWORDS:** Data Modeling; Probability and Statistics; Logistic Regression; Sports Analytics; Tennis Modeling.

## ■ Introduction

From different styles of play to ever changing stroke techniques, tennis is a sport that has evolved over time. Despite the various approaches throughout the past century, the rules of tennis have remained consistent. Although tennis has been a sport played for over 120 years, there has not been extensive research analyzing tennis statistics, especially when compared to research in other major American sports that are surrounded with an abundance of data science.<sup>1</sup> Even though tennis can be a team sport, competitive tennis is more often played as a single player rather than as a doubles partnership. The analysis proposed may more greatly benefit single players as many factors will not be affected by a teammate and are solely caused by their individual performance. The results may individually show the importance of each statistic to a tennis player.

In tennis, performance statistics are one of the most important aspects of a player's game. Strategy develops an individual's performance in various categories and contributes to the win or loss of a match. While statistical prediction models for tennis outcomes have been conducted for a few decades, each model has varying levels of accuracy.<sup>2</sup> The goal of this paper is to determine how strategy affects the outcome of a tennis match. A logistic regression model using Python was developed to analyze match data and determine the probability of a win for each individual player when matched against other players.

The current data science research surrounding tennis strategy highlights which skills are necessary to help an athlete win a match and are largely used by sports bettors. In 2019, researchers in China developed data-driven models of "point-by-point performance for male tennis players"

in Grand Slam tournaments to determine how different contextual variables could be used to predict point outcomes of players based on their opposition.<sup>3</sup> Point-by-point datasets give more detailed information about a single match, as it shows the scoring progression and the player who served.<sup>2</sup> Gollub matched 12,000 point-by-point strings of the players who won and the points scored to generate a prediction and the match progression.<sup>2</sup> This logistic regression model had a 76.2% accuracy.<sup>2</sup>

Additionally, Barnett and Clark demonstrated that it is possible to develop models to show how important certain advantages are to a player.<sup>4</sup> Cornman and colleagues built on this to compare logistic regression, random forests, support vector machine, and neural networks using data from online tennis datasets.<sup>5</sup> Their research indicated that logistic regression and support vector machine have the highest predictability with almost 70% accuracy.<sup>5</sup> Gu and Saaty compared subjective judgments by tennis experts concerning who was most likely to win 94 US Open tennis matches.<sup>6</sup> Their data modeling predicted a 15% greater accuracy rate than that of tennis experts, indicating the benefits to statistical modeling in this sport.<sup>5</sup>

A variety of methods are utilized in sports analytics, especially when analyzing large amounts of data to assist with predicting outcomes. These research studies demonstrate the benefits and drawbacks of various analytical methods. The method selected for determining the potential outcomes in top 10 player matchups was a logistic regression. This type of analysis was utilized to determine the probability of a player winning his match. This paper displays the process for determining the probability of top ranked players winning against other top ranked players. It also shows which statistics

are most important to determine who will win a particular tennis match.

## ■ Methods

### *Data:*

The data that was utilized to perform these regressions was from match statistics of top ranked players from recent years. Data was obtained through “Ultimate Tennis Statistics,” which is a repository for tennis data compiled by tennis statisticians.<sup>7</sup> The specific data utilized to develop this model was taken from the top 10 male tennis athletes during each of the five years between 2015–2019. Due to the limitations of the website design, data was individually collected from each player. Since this process was manualized, only the top ten players’ data was selected. This five-year period was selected as a purposive sample to examine the most recent five-year span of tennis data immediately prior to the COVID-19 pandemic (as 2019 was the last full tennis season). Data was not selected post 2019 to avoid non-sampling errors (or potentially skewed data sets) resulting from crowd capacity, limited opponents, and other COVID-19 related game restrictions.

For each of these players, match statistics were gathered from each of the selected player’s matches in a Grand Slam tournament for that certain year. The Grand Slam Tournaments data from these performers was solely used because players were placed into the brackets of greatest difficulty during these tournaments compared to other tournaments. The statistics that Ultimate Tennis Statistics provide for individual matches include first serve percentage, first serve won percentage, first serve return won percentage, second serve won percentage, second serve return won percentage, ace percentage, break points saved percentage, break points won percentage, and double fault percentage.<sup>7</sup> For a given match, these statistics were given for both players. The difference was calculated between the values for the top ranked player and their opponent to serve as the independent variables in the logistic regression models. Therefore, each of these statistics in the dataset represent the difference between the top ranked player and their opponent. In total, 786 tennis matches were gathered across all of the top 10 tennis players’ matches.

In the dataset, there were statistics of a match that were direct inverses of each other. For instance, a player’s first serve win percentage could be found by subtracting the opponent’s first serve return win percentage from one. Therefore, a player’s first serve win percentage and his opponent’s first serve return won percentage would have to equal one. Similarly, the same is true for second serve won percentage and opponent’s second serve return won percentage, as well as break points saved percentage and opponent’s break points won percentage. Due to the nature of these factors, only first serve won percentage, second serve won percentage, and break points won percentage were used.

Data was available for these top ten male players for both individual matches and season-wide aggregates. In addition to grabbing the matches, the season-wide data was also gathered. This data included the same independent variables

that were used above. The season-wide stats were used to demonstrate an average estimate of the player’s performance relative to other players during that season. After the logistic regressions were created, the season-wide averages were used to predict the probability of each player winning a certain matchup.

### *Statistical Design:*

After gathering the data for each individual match, this data was used to create a logistic regression to predict a winner of a specific tennis match. Analyzing this data and observing the significance of each aspect of the game (i.e., first serve percentage won as it relates to the game outcome) eventually led to a model that uses the previously studied data to see the outcome of any match between top 10 players in the range that was discussed earlier. When listing the outcome, it showed the probability of one player beating the other.

Logistic regression was appropriate for several reasons. First, a logistic regression model was able to describe the statistics in a clear way. Through the regression model, prediction charts may be developed to demonstrate the likelihood that one player will win against another specific player as long as both player’s individual data sets have been included in the modeling. By using a logistic regression model, it was evident which variables were statistically significant. By knowing this, one can produce probabilities for which players would win in future matchups, as one player might be better than the other player in a category that is essential for a win. The logistic regression model fits around this concept of significant data and uses a player’s success in this essential data against given opponents to determine future outcomes of these matches. For the model to be a good fit, it is important to experiment with different combinations of independent variables to produce the best result.

This model was extremely helpful for predicting outcomes as it gave a precise probability of who would win in each match based on an individual’s previous matches. Not only does this allow outside individuals to predict winners of matches, but it also allows individual players to focus on which skills or strategies require improvement to win more matches. If tennis players could see how these statistics can change the outcome of their game, they would benefit in their game strategies.

## ■ Results and Discussion

A logistic regression model was developed to determine the correlative effect between a player’s individual game statistics (like first serve won percentage) and the likelihood of winning the match against a similarly skilled opponent. To develop this model, the dependent and independent variables were carefully defined so that the model would not confound data between a match win and the statistics that led to past wins.

First, the dependent variable in this model is a win variable where one represents a win while a zero represents a loss. Since the model is meant to be predictive, win was a logical dependent variable to select because this type of model anticipates outcomes of events which may only have two outcomes; in the instance of tennis, the outcomes are either

wins or losses.

To create the program, the outcome of the match was recorded as a binary number, either being one for a win or zero for a loss. For instance, if Roger Federer defeated Andy Murray during a match in 2019, the dependent variable on Federer's data would have an output of the number one to indicate that he had won, whereas Murray would have an output of the number zero to indicate a loss. The statsmodel.api package in Python was used to create this logistic regression model. Using the "Logit" function associated with this package and incorporating the data as outlined, a regression was developed.

The independent variables in this program were the performance statistics that were discussed in the previous section (and defined in the Appendix): First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, Break Points Won Percentage, Ace Percentage, and Double Fault Percentage. These are independent variables because their values are used in order to determine the probability that a given tennis player will win a hypothetical match against another player.

Table 1 demonstrates the results of the logistic regression for the matches of the top 10 ranked male players in Grand Slam tournaments from 2015-2019. The table is organized to emphasize the coefficients for independent variables. The significant independent variables were First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, and Break Points Won Percentage (Table 1). The variable's significance is based on P values, where the lower the P value is, the more significant it is in predicting the outcome of the match. The coefficient shows the direction and magnitude of the effect of the variable.

**Table 1:** Logistic regression results to determine a win or loss.

Variable	Coefficients	Standard Error	z	P value
1 <sup>st</sup> Serve Won %	33.71	4.31	7.83	<.001 <sup>a</sup>
2 <sup>nd</sup> Serve Won %	18.14	2.70	6.72	<.001 <sup>a</sup>
1 <sup>st</sup> Serve %	10.75	3.31	3.25	.001 <sup>a</sup>
Break Points Won %	4.38	1.01	4.34	<.001 <sup>a</sup>
Double Fault %	-2.38	8.42	-0.28	.78
Ace %	-3.20	3.10	-1.03	.30

<sup>a</sup> Indicates significant value.

Many of these variables had extremely low *P* values, acceptable at the .001 level, meaning that they were all significant towards the outcome of the game. The sign (either positive or negative) of the coefficient explains how it affects the dependent variable. A positive coefficient would mean that increasing the value of the independent variable will also greatly increase the probability of a win. The magnitude of the coefficient shows how big of an effect it has. The larger the coefficient, the greater the effect of a one unit increase or decrease of the independent variable on the dependent variable. Moreover, the coefficients of several of these statistics were relatively high. The lowest coefficient was "Break Points Won Percentage" at 4.38.

Ultimately, the ace percentage and double fault percentage were both deemed insignificant in predicting the outcome of

a tennis match between top players. One reason why these factors might not prove essential to match outcome could be because aces and double faults do not make up a large portion of points in a tennis match. When professionals play against each other, double faults are rare, and aces are uncommon. Although the ace percentage was insignificant, it may have had a negative coefficient because it could have been correlated with double faulting and, therefore, aggressive serving. When a tennis player serves more aggressively, he or she is more likely to have aces but also more likely to have double faults. For tennis professionals, the serve and return are necessary for the game play, so players always focus on these parts of the point. Professional players make sure they do not "give" their opponents a free point by double faulting. Further, aces are rare because players are trained not to miss the ball completely on a serve or to let the ball bounce multiple times on his side of the net; both of these are relatively simplistic errors and somewhat upcoming to professional play.

To produce prediction probabilities for hypothetical matches of all the top 10 players against each other, another logistic regression model was created using only the four significant variables with LogisticRegression from the sklearn.linear\_model package in Python. After creating this model, the "predict\_proba" function was used to produce prediction probabilities. This package was used to predict probabilities due to a lack of function in the statsmodel.api package to perform the same functions.

The data that was used to predict the outcome of the matches were the season statistics. To help calculate the predictions, the program used the differences between the season-wide statistics across the whole range of top players. The difference of the season-wide statistics were the data that was used in predicting outcomes of all hypothetical matches among the top 10 players from 2015-2019. Furthermore, only the significant independent variables discussed earlier were put into the regression model. Including the insignificant statistics would skew the results of this model if the program used independent variables that were not significant towards the outcome of the match.

By focusing on the study findings, readers can see the importance of individual strategy on a player's game outcomes. For example, by discovering how significant serves are to a game's outcome, along with the importance of aces, players can determine how aggressive to be when serving and which type might yield the best outcome in the game. Also, the "break points won" percentage shows the level of importance in winning games when the tennis player has a high advantage where he needs to win that point to secure the win in that game, giving a perception of the importance of "clutchness." It is important for players to be able to win breakpoints as every time they have a lead, they are able to finish it so they can save energy and the momentum swing of an opponent's comeback.

After this model was created, there were many compelling findings. The model makes player rankings more evident. One such finding was that Milos Raonic had a greater than 50% win probability against all players during the 2018 season.

However, the player who had less than a 50% probability to win against all players was Martin Klizan in 2018. Raonic in 2018 has one of the highest first serve win percentages out of all the players while Klizan in 2018 had one of the lowest. This shows how the model recognizes the importance of first serve won percentage and how much it takes part in determining the outcome of the match. In addition to finding the best and worst player with the average predicted probability against all the other players, it was also found that the top 10 players with the highest average predicted probability against other players and the bottom 10 (Table 2). The probability matrix in the Appendix contains a player's probability of beating each player in the dataset. The probabilities come from the player, whose name is located in rows, and his chance of beating each player, whose name is also in the columns. However, due to efficiency and readability, the matrix uses a number in place of each name. The number correlates to the row number the player was on in the dataset. For instance, in the table, when looking at Rafael Nadal in 2019, the number that takes its place in the matrix table is the number 1 as he was the best performer in 2019, leading him to be the highest in the dataset. After building the matrix, the mean probability was calculated by averaging the values of each row (Appendix).

From these results, there were also some interesting cases that despite the player performing better in the statistics that were significant towards the outcome, they still lost. This can occur many times as tennis is also a game of luck and there can be situations where a player is performing better than usual but is not able to pull off a win. For example, since each set is a race to 6, a player could win the first two sets 6-0, then lose the next three 7-5 but overall have better performance statistics.

**Table 2:** Top 10 best and worst players used in statistical analysis.

Rank	Player and Year	Average Probability
1	Milos Raonic 2018	87.0%
2	Roger Federer 2018	85.4%
3	Milos Raonic 2015	84.2%
4	Roger Federer 2015	83.1%
5	Roger Federer 2017	80.8%
6	Rafael Nadal 2019	78.9%
7	Rafael Nadal 2017	78.6%
8	Novak Djokovic 2015	77.6%
9	Roger Federer 2019	77.5%
10	Roger Federer 2016	76.9%
41	Rafael Nadal 2015	29.9%
42	Gael Monfilis 2019	28.1%
43	Dominic Thiem 2019	27.1%
44	Kei Nishikori 2016	24.7%
45	Alexander Zverev 2017	24.1%
46	David Ferrer 2015	21.3%
47	Kei Nishikori 2019	20.7%
48	Alex De Minaur 2019	19.2%
49	Andrey Rublev 2019	11.8%
50	Martin Klizan 2018	6.8%

## ■ Conclusion

This model was successful in its attempt to display which statistics were significant out of the list of selected independent variables. The regression showed that the significant variables with a low P value were First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, and Break Points Won Percentage. There were many possible reasons why these were the significant values as the other insignificant values are mostly uncommon occurrences in professional tennis games. After examining the predictions created from the regression model, there were several intriguing results. One such result was that a player could have a better performance in the statistics of the significant values than his opponent but would still lose on occasion. One reason for this could be a player dominating in the first few of his sets but losing in total as his opponent would win all the other sets by a small amount. Even though the player still won more points in total, the opponent still won all the other sets which secured him a win.

This paper was written to help tennis athletes and avid tennis-watchers to understand which statistics are significant towards determining the outcome of a match. Also, this paper provides another perspective on the use of Python to simulate tennis matches. As tennis popularity trends upwards, there still has not been much research on data science surrounding this sport. Nevertheless, tennis enthusiasts have been curious about the creation of a program that could determine which professional tennis player would beat another. Many game strategies are based upon which shots matter for a win. After this research, it is clear how important it is to win games starting with the first serve as it makes up for a large portion of the game. This emphasizes the importance of being ready for a return right after a strong first serve. If a player hits a strong first serve, returning another strong shot will be essential for starting the match. Tennis attracts many players, and it is a game of necessary strategy, making data science an excellent choice for studying this field and how players can perform their best.

### *Future Directions :*

While the results from the initial analysis seem promising to predict tennis outcomes, repeating this modeling with other populations would increase the reliability. In the future, this model could be used on a larger sample size. Having more match data would demonstrate whether this model is applicable to a greater population. Further, including female tennis players in the sample to see if the model can still accurately predict outcomes in more than just male athletes would produce an interesting complement to this research.

Additionally, the data used in this sample was from 2015 to 2019—before the COVID-19 pandemic. It would be interesting to see how the model fits games played during 2020 and beyond, in which there would be different conditions such as little to no crowds, increased safety protocols, and possible psychological stressors for the players.

If the model described here is found to be consistent and generalizable with a larger population, creating a program in which a user-interface allows someone to pick two players

within a range and have it output their probabilities would produce a practical data model.

### ■ Acknowledgements

The work represented in this paper was not funded by any outside source. The author would like to thank and acknowledge Nick Riccardi, Adjunct Professor and Data Scientist at Syracuse University, for his mentorship during this project. Further, the author acknowledges his gratitude to his parents for their immense support during late nights spent writing.

### ■ References

1. Bruce, M. G. L.; Aberdare, B.; Lorge, B. S. Tennis, 2021. Encyclopedia Britannica. <https://www.britannica.com/sports/tennis> (accessed 2021-08-23).
2. Gollub, J. Producing Win Probabilities for Professional Tennis Matches from any Score. Bachelor's Thesis, Harvard College, Cambridge, MA, 2017. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:41024787> (accessed 2021-08-23).
3. Cui, Y.; Liu, H.; Liu, H.; Gómez, M. A. Data-Driven Analysis of Point-by-Point Performance for Male Tennis Player in Grand Slams. *Motricidade*. 2019, 15 (1), 49-61. DOI: 10.6063/motricidade.16370
4. Barnett, T.; Clarke, S. R. Combining Player Statistics to Predict Outcomes of Tennis Matches. *IMA Journal of Management Mathematics*. 2005, 16, 113-120. DOI: 10.1093/imaman/dpi001
5. Cornman, A.; Spellman, G.; Wright, D. Machine Learning for Professional Tennis Match Prediction and Betting. *Computer Science*. 2017.
6. Gu, W.; Saaty, T. L. Predicting the Outcome of a Tennis Tournament: Based on Both Data and Judgments. *Journal of Systems Science and Systems Engineering*. 2019, 28 (3), 317-343. DOI: 10.1007/s11518-018-5395-3
7. Ultimate Tennis Statistics. <https://ultimatetennisstatistics.com/> (accessed 2021-08-23)

### ■ Authors

Cayden Tu is is the class of 2024 at Saint Ignatius High School in San Francisco. At school, he serves in leadership roles in various STEM organizations and the newspaper. His previous work includes coding, robotics, and creating artificial intelligence models focusing on disaster response efficiency and combating misogynistic verbal abuse. He enjoys playing tennis, sports, and data science and conceived this project to merge his passions.

### ■ Appendix

**Ace Percentage:** The percentage of a player's serve, typically the first serve striking with maximum force, that is so powerful to which the opponent is not able to return the serve at all by missing the ball completely or letting it bounce twice.

**Double Fault Percentage:** In tennis, a tennis player has two opportunities to start a point when it is his turn to serve. Typically, professional tennis players do not miss these two opportunities as that would be a free point awarded to the opponent. The double fault percentage shows out of all the points the athlete played when he was serving, how many of those times did he miss the second opportunity, which is supposed to be much more conservative than the first serve, where he hits the ball into the net or out of the box.

**Break Points Won Percentage:** Break points describe the points in a tennis game where a player is one point away from

winning the game. In typical scoring, the player will be at the highest point number, 40, and if he wins this final point, then he secures the game. This is especially important for players to be able to finish their games when in the lead.

**First Serve Won Percentage:** The percentage of a player being able to win the point off of the first serve. This indicates that a player is able to win most of his points during the first serve.

**Second Serve Won Percentage:** The percentage of a player being able to win the point off of the second serve. This is usually much lower than the first serve win percentage as a

Player Code Reference:

PlayerYear	Player_Code
Rafael Nadal 2019	1
Novak Djokovic 2019	2
Roger Federer 2019	3
Daniil Medvedev 2019	4
Dominic Thiem 2019	5
Stefanos Tsitsipas 2019	6
Kei Nishikori 2019	7
Andrey Rublev 2019	8
Alex De Minaur 2019	9
Gael Monfils 2019	10
Rafael Nadal 2018	11
Roger Federer 2018	12
Novak Djokovic 2018	13
Juan Martin Del Potro 2018	14
Alexander Zverev 2018	15
Dominic Thiem 2018	16
Kevin Anderson 2018	17
Martin Klizan 2018	18
Marin Cilic 2018	19
Milos Raonic 2018	20
Roger Federer 2017	21
Rafael Nadal 2017	22
Novak Djokovic 2017	23
Grigor Dimitrov 2017	24
Jo Wilfried Tsonga 2017	25
Alexander Zverev 2017	26
Andy Murray 2017	27
David Goffin 2017	28
Milos Raonic 2017	29
Juan Martin Del Potro 2017	30
Andy Murray 2016	31
Novak Djokovic 2016	32
Milos Raonic 2016	33
Roger Federer 2016	34
Rafael Nadal 2016	35
Kei Nishikori 2016	36
Juan Martin Del Potro 2016	37
Nick Kyrgios 2016	38
Gael Monfils 2016	39
Stan Wawrinka 2016	40
Novak Djokovic 2015	41
Roger Federer 2015	42
Andy Murray 2015	43
David Ferrer 2015	44
Kei Nishikori 2015	45
Stan Wawrinka 2015	46
Rafael Nadal 2015	47
Tomas Berdych 2015	48
Richard Gasquet 2015	49
Milos Raonic 2015	50

Probability Matrix:

Player Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	50.0%	53.9%	52.5%	93.3%	95.6%	89.4%	97.2%	98.8%	97.5%	95.3%	86.6%	36.5%	84.7%	83.7%	94.2%	84.3%	69.3%	99.4%	84.6%	33.0%	46.2%	50.6%	93.5%	88.8%	76.6%
2	46.1%	50.0%	48.7%	92.2%	94.9%	87.8%	96.7%	98.6%	97.1%	94.6%	84.7%	33.0%	82.5%	81.4%	93.3%	82.1%	65.9%	99.3%	82.4%	29.6%	42.4%	46.7%	92.4%	87.1%	73.7%
3	47.5%	51.3%	50.0%	92.6%	95.1%	88.3%	96.9%	98.6%	97.2%	94.9%	85.4%	34.2%	83.3%	82.2%	93.6%	82.9%	67.1%	99.3%	83.2%	30.8%	43.7%	48.1%	92.8%	87.7%	74.7%
4	6.7%	7.8%	7.4%	50.0%	60.5%	38.0%	70.4%	84.4%	72.8%	59.1%	32.2%	3.9%	28.9%	27.4%	53.6%	28.4%	14.2%	91.6%	28.8%	3.3%	5.8%	6.8%	50.6%	36.6%	19.5%
5	4.4%	5.1%	4.9%	39.8%	50.0%	28.6%	60.8%	77.8%	63.5%	48.5%	23.6%	2.5%	20.9%	19.7%	43.0%	20.5%	9.7%	87.6%	20.8%	2.1%	3.8%	4.5%	40.2%	27.4%	13.8%
6	10.6%	12.2%	11.7%	62.0%	71.4%	50.0%	79.6%	89.9%	81.4%	70.2%	43.7%	6.3%	39.9%	38.1%	65.3%	39.3%	21.5%	94.8%	39.8%	5.4%	9.2%	10.8%	62.6%	48.5%	28.4%
7	2.8%	3.3%	3.1%	29.6%	39.2%	20.4%	50.0%	69.2%	52.9%	37.8%	16.5%	1.6%	14.4%	13.5%	32.7%	14.1%	6.4%	81.8%	14.4%	1.4%	2.4%	2.9%	30.2%	19.4%	9.0%
8	1.2%	1.4%	1.4%	15.6%	22.2%	10.1%	30.8%	50.0%	33.3%	21.2%	7.9%	0.7%	6.8%	6.3%	17.7%	6.7%	2.8%	66.5%	6.8%	0.6%	1.1%	1.3%	16.0%	9.5%	4.1%
9	2.5%	2.9%	2.8%	27.2%	36.5%	18.6%	47.1%	66.7%	50.0%	35.1%	14.9%	1.4%	13.0%	12.2%	30.2%	12.7%	5.7%	80.0%	13.0%	1.2%	2.2%	2.6%	27.8%	17.7%	8.1%
10	4.7%	5.4%	5.1%	40.9%	51.5%	29.8%	62.2%	78.8%	64.9%	50.0%	24.7%	2.7%	21.9%	20.6%	44.5%	21.5%	10.2%	88.2%	21.8%	2.3%	4.0%	4.8%	41.6%	28.5%	14.3%
11	13.4%	15.3%	14.6%	67.8%	76.4%	56.3%	83.5%	92.1%	85.1%	75.3%	50.0%	8.0%	46.1%	44.3%	70.9%	45.5%	26.1%	95.9%	46.0%	6.9%	11.7%	13.7%	68.4%	54.9%	33.9%
12	63.5%	67.0%	65.8%	96.1%	97.5%	93.7%	98.4%	99.3%	98.6%	97.3%	92.0%	50.0%	90.7%	90.0%	96.7%	90.5%	79.8%	99.7%	90.7%	46.1%	59.9%	64.0%	96.2%	93.3%	85.2%
13	15.3%	17.5%	16.7%	71.1%	79.1%	60.1%	85.6%	93.2%	87.0%	78.1%	53.9%	9.3%	50.0%	48.2%	74.0%	49.4%	29.3%	96.5%	49.9%	8.0%	13.4%	15.7%	71.7%	58.7%	37.5%
14	16.3%	18.6%	17.8%	72.6%	80.3%	61.9%	86.5%	93.7%	87.8%	79.4%	55.7%	10.0%	51.8%	50.0%	75.4%	51.2%	30.8%	96.8%	51.7%	8.6%	14.3%	16.7%	73.1%	60.4%	39.3%
15	5.8%	6.7%	6.4%	46.4%	57.0%	34.7%	67.3%	82.3%	69.8%	55.5%	29.1%	3.3%	26.0%	24.6%	50.0%	25.5%	12.5%	90.4%	25.9%	2.9%	5.0%	5.9%	47.0%	33.3%	17.3%
16	15.7%	17.9%	17.1%	71.6%	79.5%	60.7%	85.9%	93.3%	87.3%	78.5%	54.5%	9.5%	50.6%	48.8%	74.5%	50.0%	29.8%	96.6%	50.8%	8.2%	13.7%	16.0%	72.2%	59.3%	38.1%
17	30.7%	34.1%	32.9%	85.8%	90.3%	78.5%	93.6%	97.2%	94.3%	89.8%	73.9%	20.2%	70.7%	69.2%	87.5%	70.2%	50.0%	98.6%	70.6%	17.8%	27.6%	31.2%	81.5%	77.5%	59.1%
18	0.6%	0.7%	0.7%	8.4%	12.4%	5.2%	18.2%	33.5%	20.0%	11.8%	4.1%	0.3%	3.5%	3.2%	9.6%	3.4%	1.4%	50.0%	3.5%	0.3%	0.5%	0.6%	8.6%	4.9%	2.1%
19	15.4%	17.6%	16.8%	71.2%	79.2%	60.2%	85.6%	93.2%	87.0%	78.2%	54.0%	9.3%	50.1%	48.3%	74.1%	49.5%	29.4%	96.5%	50.0%	8.1%	13.5%	15.7%	58.8%	37.6%	24.6%
20	67.0%	70.4%	69.2%	96.7%	97.9%	94.6%	98.6%	99.4%	98.8%	97.7%	93.1%	53.9%	92.0%	91.4%	97.1%	91.8%	82.2%	99.7%	91.9%	50.0%	63.6%	67.5%	96.8%	94.3%	87.1%
21	53.5%	56.3%	54.2%	94.2%	96.2%	90.8%	97.6%	98.9%	97.8%	96.0%	88.3%	40.1%	86.6%	85.7%	95.0%	86.3%	72.4%	95.6%	86.5%	36.4%	50.0%	54.0%	98.4%	90.7%	79.3%
22	49.4%	53.3%	51.9%	93.2%	95.5%	89.2%	97.1%	98.7%	97.4%	95.2%	86.3%	36.0%	84.3%	83.3%	94.1%	84.0%	68.8%	99.4%	84.3%	32.5%	45.7%	50.0%	93.3%	88.5%	76.2%
23	6.5%	7.6%	7.2%	49.4%	59.8%	37.4%	69.8%	84.0%	72.2%	58.4%	31.6%	3.8%	28.3%	26.9%	53.0%	27.8%	13.9%	91.4%	28.3%	3.2%	5.6%	6.7%	50.0%	36.0%	19.1%
24	11.2%	12.9%	12.3%	63.4%	72.6%	51.5%	80.6%	90.5%	82.3%	71.5%	45.1%	6.7%	41.3%	39.6%	66.7%	40.7%	22.5%	95.1%	41.2%	5.7%	9.8%	11.5%	64.0%	50.0%	29.7%
25	23.4%	26.3%	25.3%	80.5%	86.5%	71.6%	91.0%	95.9%	91.9%	85.7%	66.1%	14.8%	62.5%	60.7%	82.7%	61.9%	40.9%	97.9%	62.4%	12.9%	20.7%	23.8%	80.9%	70.3%	50.0%
26	3.4%	4.2%	4.0%	34.9%	45.1%	24.7%	56.0%	74.1%	58.8%	43.6%	20.2%	2.1%	17.7%	16.7%	38.3%	17.4%	8.0%	85.2%	17.7%	1.8%	3.1%	3.7%	35.5%	23.6%	11.3%
27	5.8%	6.7%	6.4%	46.4%	57.0%	34.7%	67.3%	82.3%	69.8%	55.5%	29.1%	3.3%	26.0%	24.6%	50.0%	25.5%	12.5%	90.4%	25.9%	2.9%	5.0%	5.9%	47.0%	33.3%	17.3%
28	12.8%	14.6%	14.0%	66.6%	75.4%	55.0%	82.8%	91.7%	84.4%	74.3%	48.7%	7.6%	44.8%	43.0%	69.8%	44.2%	25.1%	95.7%	44.7%	6.6%	11.1%	13.0%	67.2%	53.6%	32.8%
29	30.2%	33.6%	32.4%	85.5%	90.1%	78.1%	93.5%	97.1%	94.2%	89.6%	73.4%	19.8%	70.2%	68.7%	87.2%	69.7%	49.4%	98.6%	70.1%	17.4%	27.1%	30.7%	85.8%	77.1%	58.6%
30	7.2%	8.4%	7.9%	52.0%	62.4%	39.9%	72.0%	85.4%	74.3%	61.0%	34.0%	4.2%	30.5%	29.0%	55.6%	30.0%	15.2%	92.2%	30.5%	3.6%	6.2%	7.4%	53.6%	38.5%	20.8%
31	14.2%	16.2%	15.5%	69.2%	77.6%	57.9%	84.4%	92.6%	85.9%	76.5%	51.7%	8.6%	47.8%	45.9%	72.2%	47.2%	27.5%	96.2%	47.7%	7.4%	12.4%	14.5%	69.8%	56.5%	35.4%
32	12.2%	14.1%	13.4%	65.6%	74.5%	53.9%	82.1%	91.3%	83.7%	73.4%	47.5%	7.3%	43.6%	41.9%	68.8%	43.1%	24.2%	95.5%	43.6%	6.3%	10.7%	12.6%	68.2%	62.4%	31.7%
33	35.3%	38.9%	37.7%	88.2%	92.1%	81.9%	94.8%	97.7%	95.4%	91.6%	77.7%	23.8%	74.9%	73.4%	89.7%	74.4%	55.2%	98.9%	74.8%	21.1%	31.9%	35.8%	85.5%	81.0%	64.0%
34	36.3%	50.2%	48.8%	92.3%	94.9%	87.8%	96.7%	98.6%	97.1%	94.6%	84.8%	33.2%	82.6%	81.5%	93.3%	82.2%	66.1%	99.3%	82.5%	29.8%	42.6%	46.9%	92.5%	87.2%	73.8%
35	3.8%	4.6%	4.1%	35.8%	46.0%	25.4%	56.9%	74.9%	59.7%	44.6%	20.8%	2.1%	18.3%	17.2%	39.2%	17.9%	8.3%	85.7%	18.3%	1.8%	3.2%	3.8%	36.4%	24.9%	13.3%
36	16.2%	18.4%	17.6%	72.4%	80.1%	61.6%	86.3%	93.6%	87.7%	79.2%	55.4%	9.9%	51.5%	49.7%	75.2%	50.9%	30.6%	96.7%	51.4%	8.5%	14.2%	16.5%	72.9%	60.1%	38.9%
37	12.1%	24.9%	23.9%	79.4%	85.6%	70.1%	90.3%	95.6%	91.3%	84.8%	64.5%	13.9%	60.8%	59.0%	81.7%	60.2%	39.2%	97.8%	60.7%	12.1%	19.6%	22.5%	68.8%	68.8%	48.2%
38	9.1%	10.5%	10.0%	58.1%	68.0%	46.0%	76.8%	88.3%	78.8%	66.7%	39.7%	5.4%	36.1%	34.4%	61.6%	35.5%	18.8%	93.9%	36.0%	4.6%	7.9%	9.3%	58.7%	44.5%	25.2%
39	13.5%	15.4%	14.7%	68.0%	76.5%	56.5%	83.6%	92.1%	85.2%	75.4%	50.2%	8.1%	46.3%	44.5%	71.0%	45.7%	26.3%	96.0%	46.2%	7.0%	11.8%	13.5%	65.8%	55.0%	34.1%
40	47.7%	51.6%	50.2%	92.7%	95.2%	88.4%	96.9%	98.6%	97.2%	94.9%	85.5%	34.4%	83.4%	82.3%	93.6%	83.0%	67.3%	99.3%	83.3%	31.0%	43.9%	48.0%	95.9%	87.8%	74.9%
41	52.0%	62.2%	60.9%	95.2%	96.9%	92.3%	98.0%	99.1%	98.2%	96.7%	90.2%	44.8%	88.7%	87.9%	95.9%	88.4%	76.1%	99.6%	88.7%	41.0%	54.8%	59.0%	92.4%	91.0%	82.3%
42	12.9%	14.0%	13.4%	65.5%	74.5%	53.8%	82.0%	91.3%	83.7%	73.3%	47.4%	7.3%	43.6%	41.8%	68.7%	43.0%	24.2%	95.5%	43.5%	6.3%	10.6%	12.5%	66.1%	52.3%	31.6%
43	2.2%	3.4%	3.3%	30.5%	40.2%	21.1%	51.1%	70.1%	53.9%	38.8%	17.1%	1.7%	15.0%	14.0%	33.7%	14.6%	6.6%	82.4%	14.9%	1.4%	2.5%	3.0%	31.1%	20.1%	9.4%
44	11.8%	13.6%	12.9%	64.7%	73.8%	52.9%	81.5%	91.0%	83.2%	72.4%	46.4%	7.0%	42.7%	40.9%	67.9%	42.1%	23.5%	95.4%	42.6%	6.0%	10.3%	12.1%	65.3%	51.4%	30.9%
45	15.1%	17.2%	16.4%	70.7%	78.8%	59.6%	85.3%	93.0%	86.7%	77.7%	53.4%	9.1%	49.5%	47.6%	73.6%	48.9%	28.8%	96.5%	49.4%	7.9%	13.2%	15.1%	68.2%	58.2%	37.0%
46	5.2%	6.1%	5.8%	43.8%	54.3%	32.3%	64.9%	80.7%	67.5%	52.9%	26.9%	3.0%	23.9%	22.6%	47.3%	23.5%	11.3%	89.4%	23.9%	2.6%	4.5%	5.3%	44.4%	31.0%	15.8%
47	19.0%	21.5%	20.6%	76.0%	83.0%	65.9%	88.5%	94.7%	89.6%	82.2%	60.0%	11.7%	56.2%	54.4%	78.6%	55.6%	34.8%	97.3%	56.1%	10.2%	16.7%	19.3%	76.5%	64.6%	43.5%
48	17.1%	19.5%	18.7%	73.7%	81.2%	63.2%	87.2%	94.0%	88.4%	80.3%	57.1%	10.5%	53.2%	51.4%	76.5%	52.6%	32.1%	97.0%	53.1%	9.1%	15.1%	17.5%	72.0%	61.8%	40.6%
49	50.0%	64.5%	63.2%	95.7%	97.2%	93.0%	98.2%	99.2%	98.4%	97.0%	91.1%	47.2%	89.7%	89.0%	96.3%	89.4%	77.9%	99.6%	89.6%	43.4%	57.2%	61.4%	95.8%	92.6%	83.7%
50	96.4%	94.2%	97.2%	69.8%	92.8%	85.8%	87.8%	64.7%	63.7%	91.7%	96.3%	83.8%	77.9%	90.9%	86.5%	52.3%	41.5%	87.8%	97.1%	88.2%	84.9%	94.8%	81.0%	82.9%	39.2%
1	95.8%	93.3%	85.4%	66.4%	91.6%	83.8%	85.9%	61.1%	49.8%	90.4%	95.6%	81.6%	75.1%	89.5%	84.6%	48.4%	37.8%	86.0%	96.6%	86.4%	82.8%	93.9%	78.5%	50.5%	35.5%
2	96.0%	93.6%	86.0%	67.6%	92.1%	84.5%	86.6%	62.3%	51.2%	90.9%	95.9%	82.4%	76.1%	90.0%	85.3%	49.8%	39.1%	86.6%	96.7%</						