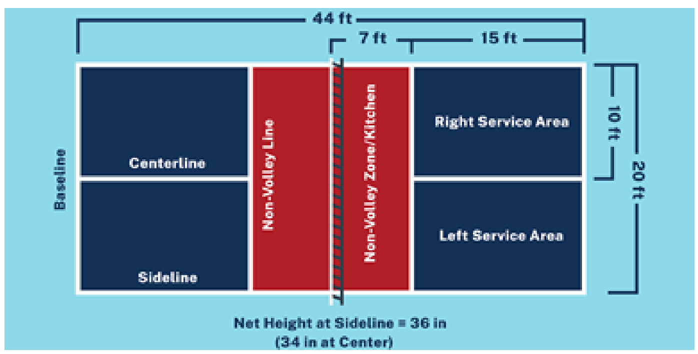
Staying Out of the Kitchen: An Analysis of Winners, Errors, and Other Factors that Lead to Success in Pickleball

Authors: Lucy Gray, Eric Miller, Alexa Penwell, Diane Youngstrom

Introduction

For our project, we decided to observe and analyze one of the fastest growing American sports, pickleball. In pickleball, players use wooden or plastic paddles to hit a perforated ball over a central net, similar to other racket sports like tennis or badminton. Both singles and doubles games are played to 11 points, win by two, and most competitive matches are played to the best of three games. Serves rotate in a circle between players, and a player will continue to serve until their team loses a rally, "earning" the other team the serve. The first hit for either team following the serve must not be a volley (hit directly out of the air). Points are only awarded to the serving team, so holding the serve is a crucial aspect of success. The most defining feature of any pickleball court is the non-volley zone, otherwise known as the "kitchen," a seven-foot zone on either side of the net (Figure 1). Players are not allowed to volley the ball when any part of them is touching the kitchen, and most in-game techniques are based around clever kitchen play.

Figure 1:



Labeled pickleball court diagram (image from: usapickleball.org)

Our first goal was to explore the types of shots taken across rallies and observe how teams would adjust play style to account for streaks of winning or losing consecutive points. We also looked to better understand the ability of players, and how different variables or metrics may predict good or bad players or affect gameplay. In order to determine relevant factors that may contribute to success in pickleball, we recorded the number of forehands, backhands, volleys, and dinks hit by each player during a rally as well as the match score, who served, and what the deciding shot was (i.e. the type of winner or error that ended the rally).

To collect our observations, we watched Pro-Level Mixed Doubles Pickleball matches on YouTube. All videos were sourced from either "The Pickleball Channel" or "Association of Pickleball Professionals (APP) TV". Alexa, Eric, and Lucy watched two games each and collected raw statistics on all four players on the court. We initially recorded this data in individual files (February 24-March 3) within a shared Google Sheets file, and Diane assisted in organizing and compiling our observations into one single Google Sheets document, as well as cleaning the data. We finished collecting data with 1032 observations.

For our sample, we observed 6 pickleball games and a total of 15 different players. Each observation represents the statistics for one player for one point of a game. We measured 7 variables for each point: Serve, Forehand, Backhand, Volley, Dink, Deciding_Shot, and Won_Rally. The Serve variable is binary; 1 if that player served the point, 0 otherwise. The Forehand, Backhand, Volley, and Dink variables count the number of that type of hit during the rally. We defined these four variables as the following: a forehand is a stroke played with the palm of the hand facing in the direction of the stroke; backhand is a stroke played with the arm across the body and the back of the hand facing in the direction of the stroke; a volley is any stroke played directly out of the air without letting the ball touch the ground; a dink is any stroke played into the kitchen of the opposing team. These four variables are numerical, and forehands and backhands are the only two that cannot overlap. For example, a player may hit a backhand volley that is also a dink, in which case it will be recorded in all three columns.

Our Deciding_Shot variable is categorical, with 8 options. The error categories are E-Net (Forced), E-Net (Unforced), E-OB (Forced), E-OB (Unforced), and E-Other (Unforced). Errors can be forced or unforced. When an error is recorded as forced, a "winning shot" was awarded to a player on the opposing team. In order to avoid double-counting a "winning shot" as both a "winner" and a "forced error", we largely excluded forced errors from our analysis. E-Net (Unforced) represents when a player hits the ball into the net, E-OB (Unforced) represents when a player hits the ball out-of-bounds, and E-Other represents a player making some other technical violation, such as hitting a volley while standing in the kitchen. E-Other are always unforced. The winning point categories represent when a player hits a good shot that results in their team winning the rally. They are: W-Spike, W-Forehand Drive, and W-Backhand Drive. A spike is a rally won by a forceful volley hit high in the air and directed downwards at a sharp angle onto the opponent's side. A forehand drive is a forceful forehand shot, and a backhand drive is a forceful backhand shot. For the Won_Rally variable, 1 is recorded for the two players whose team won the rally, and 0 is recorded for losers of the rally.

In addition to recording these necessary player statistics, we also recorded the game number and match score for each observation in order to organize our observations and help calculate winning/losing streaks. Match Score is in the format "serving team points.receiving team points.server number (either 1 or 2)". The output of the first 8 rows of our data can be found on the next page.

Game Number	Match Score	Player Name	Serve	Forehand	Backhand	Volley	Dink	Deciding_Shot	Won_ Rally
1	0.0.2	Yates	0	1	0	1	1		0
1	0.0.2	Jardim	1	7	3	9	1	E: Net (Forced)	0
1	0.0.2	Kovalova	0	7	0	4	1		1
1	0.0.2	Wright	0	1	3	4	0	W: Backhand Drive	1
1	0.0.1	Yates	0	1	0	0	0		1
1	0.0.1	Jardim	0	1	1	1	0		1
1	0.0.1	Kovalova	1	0	0	0	0		0
1	0.0.1	Wright	0	1	2	1	0	E: Net (Unforced)	0

Summary

Table 1A:

Variable	Minimum	Maximum	Total Count	Mean	Standard Deviation
Forehand	0	19	1738	1.684	2.347
Backhand	0	11	924	0.895	1.307
Volley	0	14	1141	1.106	1.605
Dink	0	14	794	.769	1.571

Table 1A summarizes the quantitative variables that we recorded for each observation when collecting data. We included the minimum, maximum, total count (across all players and games), mean, and average per rally. Forehands were the most common shot across the games we collected with a total count of 1738, and dinks were the least common. Since every shot was considered to be either a forehand or backhand, we can calculate the average number of shots per player per rally to be 2.579.

Table 1B:

Deciding_Shot	Count	Relative Frequency (%)
E: Net (Unforced)	91	36.992
E: OB (Unforced)	36	14.634
E: Other	12	4.878
W: Forehand Drive	40	16.260
W: Backhand Drive	28	11.382
W: Spike	38	15.447
W: Other	1	0.407

Table 1B summarizes the Deciding_Shot variable. We elected to remove the forced errors from this table, as this would result in some points being double counted because all forced errors also involved a winning shot. In Table 1B we recorded the count and relative frequency of the events that resulted in a rally being won or lost. One interesting thing to note is that the statistic that "75 percent of rallies are won (or lost) because of errors," (commonly heard and seen in the world of pickleball) seems to be overestimated. We only observed 56.504% of points to be won or lost due to an unforced error.

Figure 2:

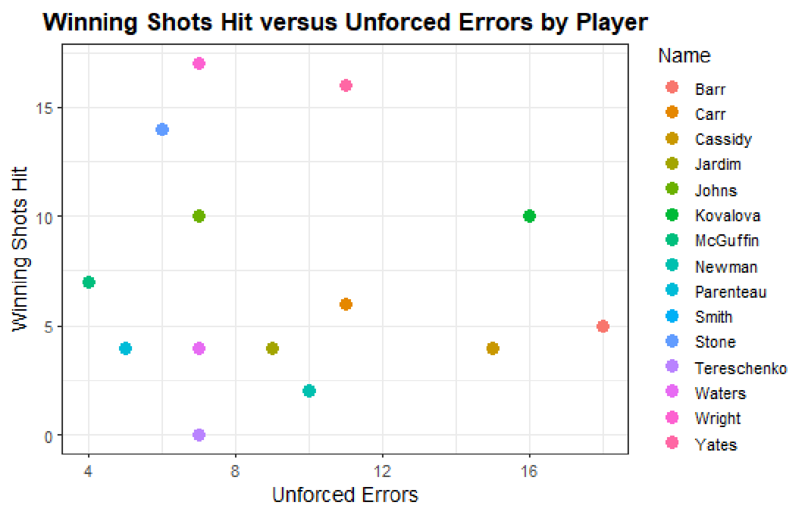


Figure 2 shows a plot comparing the number of winning shots hit and the number of unforced errors committed by each player. Either an unforced error or a winner was hit by one player every rally. Therefore, it seems logical that the number of rallies won (i.e., the outcome of the game) would be related to the number of winning shots and unforced errors. Players near the top-left of the plot are likely better players who hit many winners and few unforced errors, and players near the bottom-right are those likely weaker players, who hit few winning shots and many unforced errors. In our first insight we will develop a metric, Player Point Efficiency, that further characterizes the relationship between the number of winnings shot and unforced errors hit by a given player.

Figure 3:

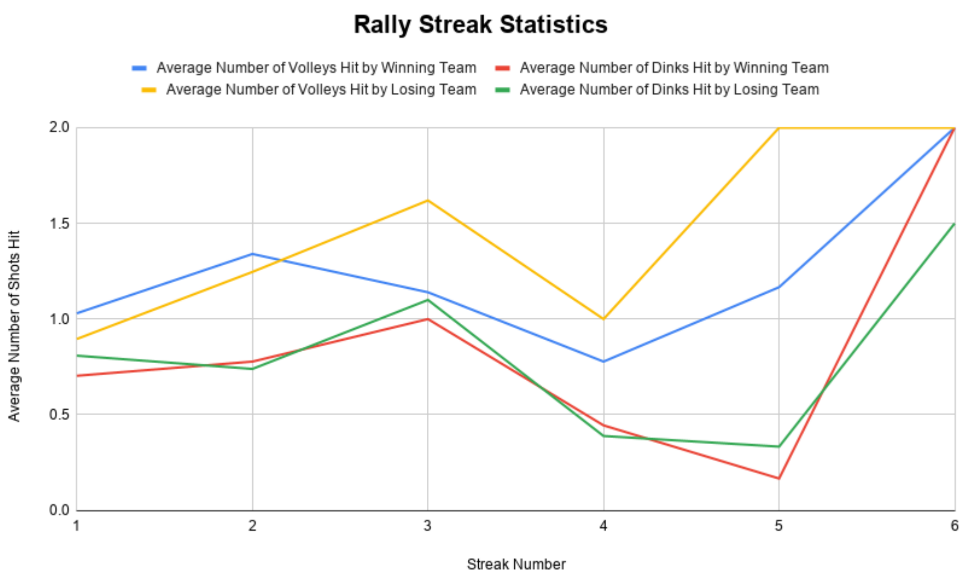


Figure 3 shows a plot of the average number of volleys and dinks hit by the winning and losing team, graphed over rally streak length, with rally streaks being the number of consecutive rallies won by a team, calculated from the Won_Rally variable. As you can see in the figure, when a team has won two rallies in a row they are hitting more volleys than the losing team on average, but after the second rally, the winning team begins to hit fewer volleys on average than the losing team. The average number of dinks, which is typically a more defensive shot, hit by each team are closer in number than the volleys, and the team hitting more dinks alternates after every rally.

Insights

Player Point Efficiency (PPE):

The Player Point Efficiency (PPE) metric was developed to build on the relationship between winning shots and unforced errors shown in Figure 2. We explored how PPE could be used to predict the pickleball game outcome and how gender impacted PPE. Since a couple of players played in multiple games and players didn't always have the same partner (e.g., Stone was on a team with Smith for one game and Carr for another), we calculated PPE for each player for each game they played in. The equation for calculating Player Point Efficiency is shown below:

$$Player\ Point\ Efficiency\ = \frac{(Total\ of\ winners\ hit\ per\ player)^2 - (Total\ unforced\ errors\ per\ player)^2}{Total\ Shots\ per\ player}$$

PPE squares the number of winners (winning shots) and unforced errors that a player hits in a given game to assign those shots more weight since they decide whether their team wins or loses the rally. The difference between the squared number of winners and unforced errors is then divided by the total number of shots for a player in a given game to normalize the metric. The team that won the most rallies always won the game; therefore, a metric that predicts the number of rallies won (PPE) is useful for determining success in pickleball. Since the number of rallies varied across different games, we normalized the number of rallies won (i.e., Won_Rally) to the total number of rallies in a given game (i.e., Player Won_Rally + Opponent Won_Rally), resulting in Adjusted Score. This way the game winning rally ratio could be compared across games. The equation for Adjusted Score is shown below:

$$Adjusted\ Score\ = \frac{Player\ Won\_Rally}{Player\ Won\_Rally + Opponent\ Won\_Rally}$$

Figure 4:

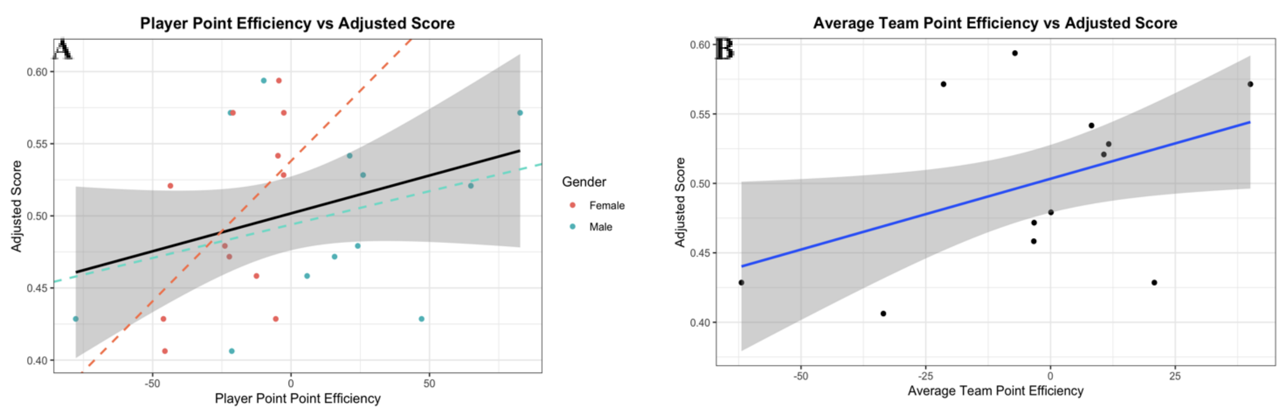


Figure 4A shows how Player Point Efficiency ratings differ between males and females. The colored dotted lines are linear regressions for how well the female/ male Player Point Efficiencies predict Adjusted Score. The black line in Figure 4A is the linear regression for all players (male and female),where the gray shading represents the corresponding confidence interval for the black line. None of the linear regression models in Figure 4A had significant p-values, indicating that gender differences in PPE alone do not predict the Adjusted Score (which is indicative of the game outcome).

The Average Team Point Efficiency is the average of the Player Point Efficiency for the partners on a team for a given game. Since we observed Pro-Level Mixed Doubles Pickleball, teams always have one male and one female. The line in Figure 4B is the linear regression model for how well Average Team Point Efficiency predicts Adjusted Score, where the gray shading around the line is the confidence interval for the linear regression model. The linear regression model for the male players (shown in Figure 4A) closely resembles the linear regression for all male and female players (i.e., the black line in Figure 4A) and the linear regression for Average Team Point Efficiency (i.e., the blue line in Figure 4B). Interestingly, the p-value was 0.0374 for the Average Team Point Efficiency linear regression in Figure 4B, indicating that the Average Team Point Efficiency significantly predicts Adjusted Score despite the lack of significant findings in the linear regression model depicted in Figure 4A. However, the multiple R^2 value was only 0.1825, which means that only 18.25% of the variation is explained by Average Team Point Efficiency.

Rally Streaks:

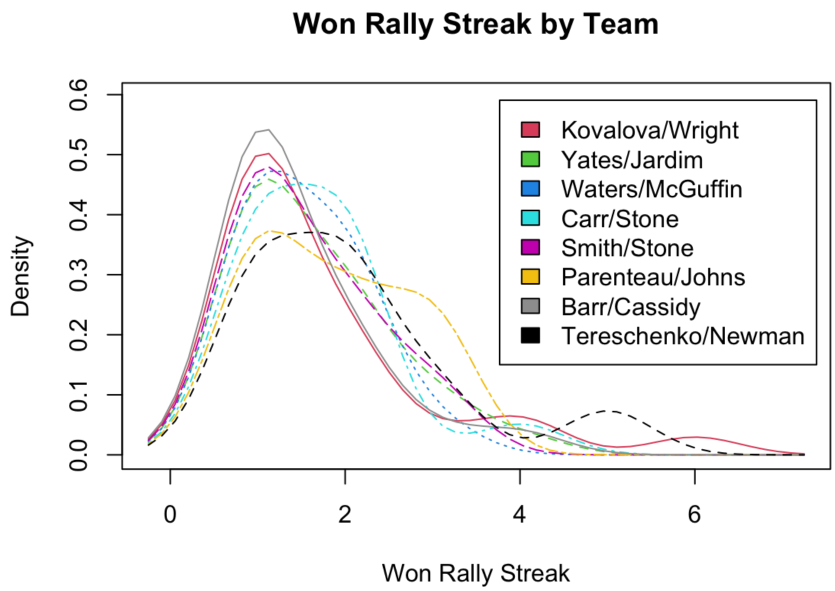
From Figure 3, it is evident that when a team has won two rallies in a row, they are hitting more volleys on average than the losing team. However, after the second rally, Figure 3 shows that the losing team begins to hit more volleys on average than the winning team. In order to determine why this change occurs after the second rally, we decided to look at the impact of serving on a team's chances of winning a rally. After a team wins two rallys in a row, they are guaranteed to now have the serve (if they didn't already) which could explain the change in frequency of different shot types. In order to test if a team is less likely to win the rally if they are serving, we conducted a 1-sample proportions test on the observations in our dataset where Serve == '1'. This test resulted in a confidence interval of [0.381, 0.467] with a p-value of 0.0004. Since the p-value is below 0.05, we reject the null hypothesis that a team's chances of winning the rally when they have the serve is equal to 50%. 50% is also not in our confidence interval, which leads us to draw the same conclusion.

The serving team being less likely to win the rally impacts both the volley and dink frequencies. As stated above, the winning team hitting fewer volleys on average than the losing team after the second rally is likely due to the fact that they are having to maintain their rally streak while serving. Because serving puts them at a disadvantage, they are able to create fewer opportunities to hit volleys, which are an offensive shot. Further, the winning team must attempt the difficult "third shot drop" if they do not want to be put out of position, keeping them on the back foot. This is likely the cause of the difference in volleys hit by the serving and returning teams.

The same reasoning can explain the alternating dink average between the serving and returning teams. Because dinks are typically a conservative, defensive shot, they will be used by the winning team because of their serving disadvantage. However, it is also likely that the losing team will hit dinks because, although the returning team is given the opportunity to be more aggressive, they will also naturally be on the defensive because they do not want to mess up and extend the opponents winning streak.

Below, we created a set of density curves displaying the length of the won rally streaks for each set of partners.

Figure 5:



As you can see from the curves, the vast majority of streaks were between 1 and 3 rallies long, while streaks of lengths 4, 5, and 6 happened much less frequently. As a result, the data in Figure 3 is most accurate for streaks 1, 2, and 3, while the values for the remaining larger streaks may have been influenced by outliers. Given our results, we would advise the winning team to focus on hitting an aggressive serve and first returning shot in order to give the returning team fewer opportunities to come into the net and hit volleys. We think this would make the probability of the serving team winning the point closer to 50%, therefore minimizing their serve disadvantage and making it easier for them to continue winning rallies and gaining momentum.

Critiques

When reflecting on our data collection process, one change we could make is to collect data for only one team per game. This change would hopefully eliminate any collector errors that were made during data collection, since the collector would have fewer players to focus on. By collecting data for both teams, there are also possible issues regarding multicollinearity between observations, since some of the variables recorded necessarily had “opposite” variables recorded for the other team. One example of this overlap is winning shots and forced errors (which we did our best to avoid). There is also some subjectivity in the variables we collected, such as the type of winning shot or difference between forced and unforced errors. Another aspect of the game we could have looked into collecting would be external factors, such as one game where Barr and Cassidy were playing on only 15 minutes of rest since their previous match.

Another change we could make would be to collect more data. Although we greatly exceeded our minimum required number of observations, more data would allow us to expand upon our analysis of Won_Rally Streaks specifically. When looking at the higher number of streaks present in our data, there were only a few observations per streak which limited the amount of data for analyzing the average number of each shot type for longer streaks.

The last change we would likely make would be to make sure we either have enough data to include multiple games for many players or make sure that there is no overlap between teams. Collecting data from multiple games of two players (one partnered team) would have allowed us to develop personalized metrics and statistics to help them hone their pickleball game. On the other hand, collecting data from many non-overlapping players would have facilitated the creation of metrics and statistics for the average pickleball player. However, in the data we collected there was one instance of a player (Stone) having two different partners (Smith and Carr) which is not ideal, as Smith and Carr might both be impacted by Stone in ways we were unable to assess despite evaluating PPE by game.