Roundnet Success:

An Analysis on Competitive Spikeball and Factors that Contribute to Scoring

Authors: Grant McGrew, Kush Patel, Nirbhay Sutaria, Prabhath Kotha, William Wu

Introduction

For our project, we decided to observe Spikeball, an easygoing, backyard sport growing in competitive play. Spikeball consists of two teams of two players, where each team is allowed three hits to spike the ball off a round net. A point is scored if the opposing team fails to return the ball off the net. Games are played to 21 points and a match typically consists of three games. While the sport is growing, there is an absence in both professional and amateur coverage. Due to the lack of coverage, we thought it would be insightful to conduct observations on several games with the purpose of analyzing box score statistics to find predictors for point differential and the most common tactics for winning points. In the beginning, we hypothesized that spikes (a hard hit ball into the net) would be the most important indicator for winning games, and we also theorized that it would be important for players to be able to use both of their hands to hit the ball. In order to investigate this, we decided to record 15 statistical variables in an attempt to understand more about the way spikeball is played and if there were surprising correlations or predictors for scoring points.

Game ID	Team ID	Player Name	Left-Hand Serves	Right-Hand Serves	Serving Faults	Aces	Kills	Spikes
0	0	PJ Showalter	0	6	7	1	7	8
0	0	Tyler Cisek	0	8	4	0	5	7
0	1	Caleb Heck	2	8	6	1	6	8
0	1	Preston Bies	2	5	4	1	5	6
1	0	PJ Showalter	0	10	8	2	7	9
1	0	Tyler Cisek	0	7	2	0	4	4
1	1	Caleb Heck	0	5	10	0	7	9
1	1	Preston Bies	2	4	5	0	4	5
2	0	PJ Showalter	0	7	9	2	5	5
2	0	Tyler Cisek	0	10	4	1	5	7
2	1	Caleb Heck	0	5	9	2	6	6
2	1	Preston Bies	4	7	4	1	2	5

For our observations, we primarily watched games hosted by the Spikeball Roundnet Association (SRA). We found videos of these games that were uploaded by players and the SRA on YouTube. Each group member was in charge of viewing several games while also recording box-score statistics on each player throughout the course of each game. Each member had equivalent roles in gathering data, as each member was in charge of finding, viewing and recording their specific games. We recorded our findings in a shared Google Sheets file with a 57x19 table that included all observations and variables pertaining to our study. For our sample, we observed 34 different players and recorded 19 variables for each observed game containing a player. Each observation represents a player's overall statistics in a particular game. We recorded 4 such observations per game. In our raw data, our values were in whole numbers because statistics in spikeball cannot be fractions or decimals. From our preview, we will discuss the five variables of right-hand serves, serving faults, aces, kills and spikes on what they represent and how they were measured. We recorded right-hand serves as successful serves with the player's right hand. Each point could have a maximum of one right-hand serve and a minimum of zero right-hand serves. The second variable, serving faults, were unsuccessful serves that either missed the net, hit the rim of the net, bounced too high or were served towards the wrong spot. There could be a maximum of two faults per rally, and depending on the tournament, either one or two faults could lead to a point for the opposing team. The third variable, aces, were serves that could not be returned, whether by the opponent missing the net or their inability to reach the ball. Our fourth variable, kills, were hits that the opposing team could not return from whether their inability to hit it back into the net, or it flew too high or far. Our fifth variable, spikes, were defined as "a successful forceful return" to the opponent. If a spike also resulted in a kill, a value was recorded for both variables. For the rest of the variables, their names are relatively straightforward and we added a value of one to each cell for each occurrence (by a player) during the game.

In addition to recording these player statistics, we made two smaller datasets. One of these datasets, named "Games," lists each game (Game ID), teams (Team ID), scores and league type for each observation. Our other dataset, "Teams," lists TeamID, team names, player names and their SRA rankings. These smaller datasets helped us organize all of our player observations and group them with teams and games while providing some background information regarding the games and teams we observed.

Summary

Table 1:

Spikeball Statistic	Min	Q1	Median	Q3	Max	Mean	SD	Total Count
Left-Hand Serves	0	0	0	0	4	0.375	0.945	21
Right-Hand Serves	3	6	7	8	12	7.054	2.101	395
Serving Faults	0	4.75	7	8.25	14	6.679	2.985	374
Aces	0	0	1	1	4	0.893	0.947	50
Kills	1	3	4.5	6	9	4.5	1.916	252
Spikes	2	5	6	8	13	6.286	2.294	352
Under-hand hits	1	6	7	9	13	7.321	2.58	410

Left-hand hits	0	2	3	6	13	3.857	2.673	216
Right-hand hits	8	14	18	22	31	18.625	5.548	1043
No-hand hits	0	0	0	1	3	0.339	0.64	19
Behind Back Hits	0	0	0	1	2	0.304	0.537	17
Dives	0	0	1	1	5	0.893	1.021	50
Out of Position Returns	0	0.75	1	2	3	1.179	0.956	66
Errors	0	0	0	1	2	0.518	0.687	29
Hit Faults	0	0	0	1	3	0.661	0.815	37
Spike Blocks	0	0	0	1	4	0.768	1.044	43

Table 1 summarizes the 16 main statistics we kept track of while watching Spikeball games. We calculated the minimum, first quartile, median, third quartile, maximum, mean, standard deviation, and total count (sum across all players and games) for each of the statistics. Many of the “plays” were uncommon, such as Left-hand serves. This variable has a third quartile of zero, and there were only 21 total occurrences across all 56 player-game combinations. On the other-hand, Right-hand hits were the most common “play” recorded. There were 1,043 total occurrences of a Right-hand hit, and the minimum value of 8 is higher than maximum for many of the other statistics.

Figure 1:

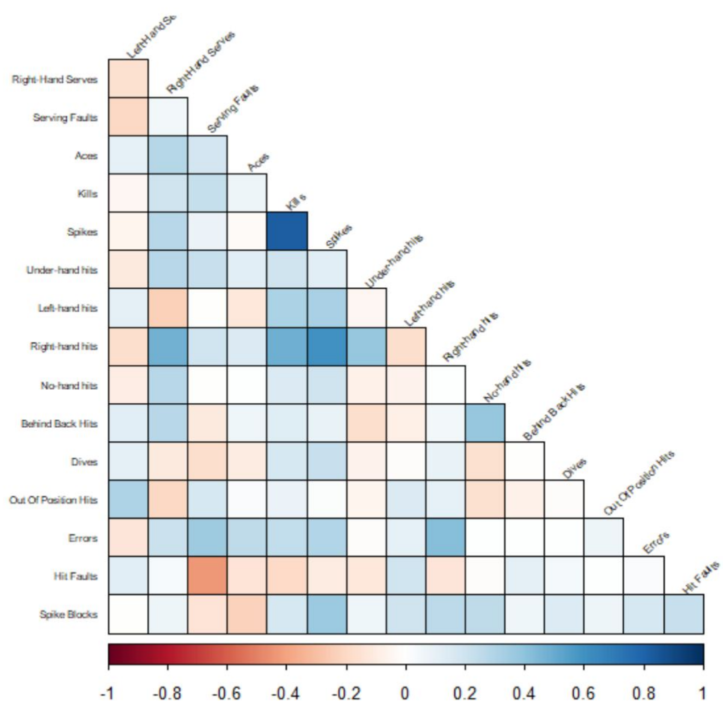


Figure 1 demonstrates a correlation matrix between all of the numeric variables recorded in the data. The most significant positive correlation was found between kills and spikes, while the most significant negative correlation was found between hit faults and serving faults. Other noteworthy correlations were between spikes and right hand hits, and errors and right hand hits.

Figure 2:

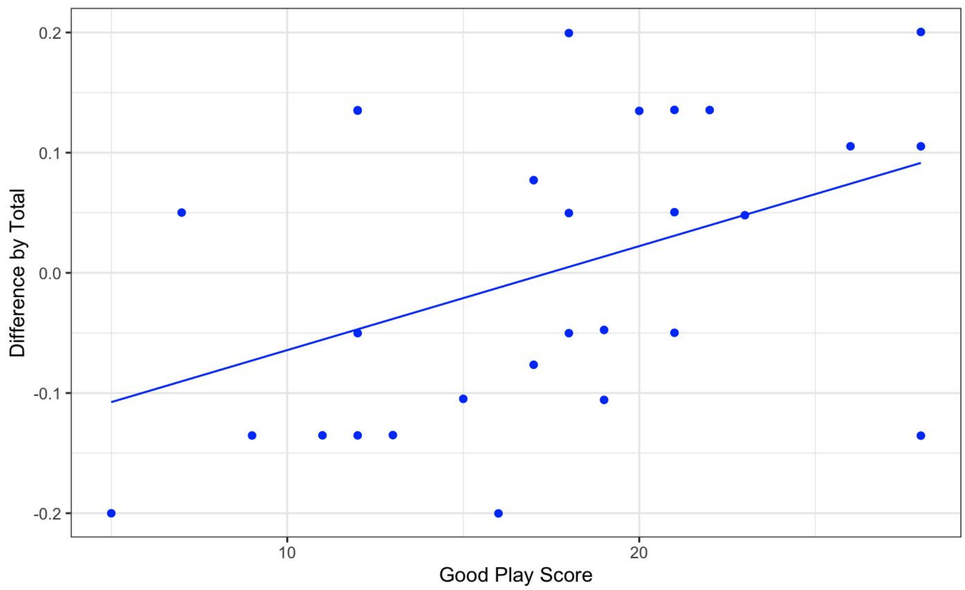


Figure 2 demonstrates a plot between the team's score differential (normalized over the total points scored by both teams that game) by what was defined as 'Good Player Score'. Here are the equations to calculate the two variables:

$$Difference\ by\ Total\ =\ \frac{Team\ Score - Opponent\ Score}{Team\ Score + Opponent\ Score}$$
$$Good\ Play\ Score\ =\ (3 * Aces)\ +\ (2\ *\ Spike\ Blocks)\ +\ Kills$$

The 'Good Play Score' was weighted based on the subjective importance of each metric in the swing of the game. The p-value from the single linear regression done between the two variables was .0224, indicating a strong case for a relationship between the two variables, with a positive correlation factor of .43.

Insights

Measuring Flexibility:

From Table 1 and Figure 1, it is evident that right hand hits correlated well with the other metrics purely because many players lack the flexibility to use both hands. The total count of right hand serves and right hand hits versus left hand serves and left hand hits was staggering, considering we would think that better spikeball players would be adept at using both hands.

As a result, we chose to dive further into the ability of players to use both of their hands. We created a Flexibility metric to quantify this ability. In order to do this, we added the Left-Hand Serves and the Right-Hand Serves to get the Total Serves, and we added the Left-hand hits and the Right-hand hits to get the Total Hits.

$$Total\ Serves\ =\ Left\ Hand\ Serves\ +\ Right\ Hand\ Serves$$
$$Total\ Hits\ =\ Left\ Hand\ Hits\ +\ Right\ Hand\ Hits$$

We then divided each of the individual numbers by their respective totals:

$$Left\ Hand\ Serve\ Ratio\ =\ Left\ Hand\ Serves/Total\ Serves$$
$$Right\ Hand\ Serve\ Ratio\ =\ Right\ Hand\ Serves/Total\ Serves$$
$$Left\ Hand\ Hit\ Ratio\ =\ Left\ Hand\ Hits/Total\ Hits$$
$$Right\ Hand\ Hit\ Ratio\ =\ Right\ Hand\ Hits/Total\ Hits$$

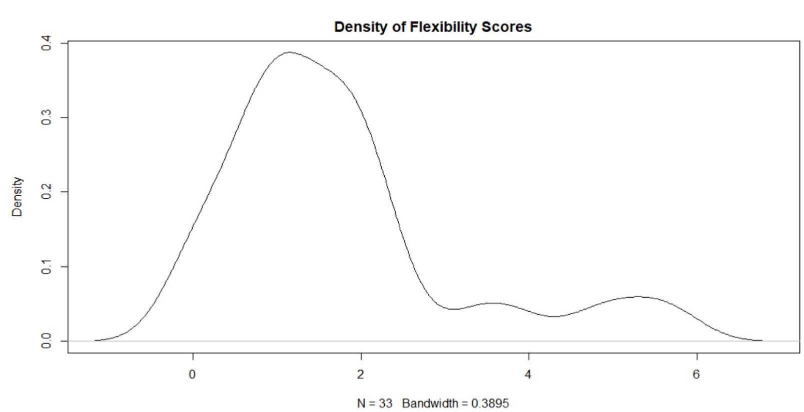
After this, we found the absolute value of the difference between the like ratios, and then subtracted that value from one to get a Serve Flexibility score and a Hit Flexibility score:

$$Serve\ Flexibility\ Score\ =\ 1\ -\ |\,Right\ Hand\ Serve\ Ratio\ -\ Left\ Hand\ Serve\ Ratio\,|$$
$$Hit\ Flexibility\ Score\ =\ 1\ -\ |\,Right\ Hand\ Hit\ Ratio\ -\ Left\ Hand\ Hit\ Ratio\,|$$

The idea behind this was that a totally flexible player would hit half of their shots with each hand, leading to ratios of 0.5 for both, and resulting in Serve and Hit Flexibility scores of one. These scores were then added together and divided by two to get a ratio out of the total possible score. These results were then multiplied by 10 to give the rating a common 10 point scale.

$$Flexibility\ Score\ =\ \frac{Serve\ Flexibility\ Score + Hit\ Flexibility\ Score}{2}\ *\ 10$$

We computed a density curve for the Flexibility Scores:



As we can see in the density curve, the vast majority of values fall in the 0-3 range. This is out of a score of 10, which indicates a player that uses their hands evenly to hit all balls. From this, we can see that most players do not use both of their hands evenly. Based on the general style of the game, we figured that players would be forced to use both of their hands, but we can clearly see that this is not the case.

Ranking Players:

From Figure 2, we can see that metrics such as aces, kills, and spike blocks were important in predicting the 'difference by total' metric. This difference by total metric takes into account that winning teams will almost always score 21 points, so trying to correlate any metric with team points would not be as beneficial as analyzing the score differential. This was then normalized by the total points scored by both teams. From the p-value being less than .05 and the relatively moderate correlation factor, we can see that having more aces, kills, and spike blocks are important not only in winning the games, but ensuring the difference is larger. Using the relative weights on aces and spike blocks was important because in a game, we found it to be very easy for the team receiving the serve to score off a spike. As a result, good serves and getting aces are crucial in breaking away from the opponent, which is why aces were given the largest weight, and it reflects well in the correlation graph. Also, being able to block spikes allows the serving team to regain the chance to break away with a point, which is why this was given a weight as well.

Since the serve is crucial in breaking away, we decided to come up with metrics to allow us to compare the serving ability of each player. The first metric showed us how often a player had a serve that didn't result in a fault. This was calculated as:

$$\frac{Right\ hand\ serves + Left\ hand\ serves}{Right\ hand\ serves + Left\ hand\ serves + Faults}$$

This statistic resulted in the following players ranking at the top:

Player Name	Save Percentage
Julie Haselton	100.00%
Kenny Ortega	87.50%
Nancy Gougeon	83.33%
Tyler Cisek	77.78%
Becca Graham	73.33%

Another Statistic we used was the percentage of serves that resulted in an ace. This was calculated as:

$$\frac{Ace}{Right\ hand\ serve + Left\ hand\ serve + Faults}$$

This statistic resulted in the following players ranking at the top:

Player Name	Ace Percentage
Shaun Boyer	33.33%
CSW tall	23.53%
Nancy Gougeon	16.67%
Sam Buckman	15.38%
Caleb Heck	14.29%

Finally we created a statistic to calculate the average number of points earned per serve. This was calculated as:

$$\frac{Ace - .5(Faults) + .5(Right\ hand\ serve + Left\ hand\ serve - Ace)}{Right\ hand\ serve + Left\ hand\ serve + Fault}$$

This statistic resulted in the following players ranking at the top:

Player Name	Expected Points / Serve
Julie Haselton	0.50
Nancy Gougeon	0.42
Kenny Ortega	0.38
Shaun Boyer	0.33
Becca Graham	0.30

There are a couple of flaws with this statistic. First, not every fault leads to a second fault and loss of point. Therefore, .5 might not be the correct number to subtract for every fault. In the future if we count double faults then we can subtract 1 for every double fault instead of .5 for every fault. Another flaw with this statistic is the +.5 for every good serve that isn't an ace. We do not know the percentage of points the serving team wins vs the receiving team. This is another variable we can record in the future.

Going Forward:

Reflecting on the data collection that was done, there are some things we could do to not only improve how we collected our data but how we analyzed it as well. To improve how we are collecting our data, it is important for everyone involved to have standard definitions of variables being collected such as a spike or out of position hit. This will lead to more consistency in our data. Our analysis could also be improved by collecting data for more variables. An example of this would be counting the number of double faults.

In the future, it could be interesting to include variables that reflect the power of the player's serves or hits. While watching the videos, there were occasions where a player like PJ Showalter would try a different style of serving, called a flango, where instead of a traditional side-hand serve, he would go over the head for more power, which seemed to have led to a higher chance of an ace. In addition, some players would fake the power they put behind a serve, and because it is a lot softer than the receiving player would expect, the player would have to dive forward to return it, and the difficulty to do this would lead to an ace for the server as well. If this were done, each observation would have to be by point or hit, which would take more time, but give more insight into the certain plays that lead to points.

While we collected our data at the player level it could also be interesting to see that data on a team to team basis. With data at the team level, we can compare our data with power rankings found online and see which variables contribute most to a team's success.