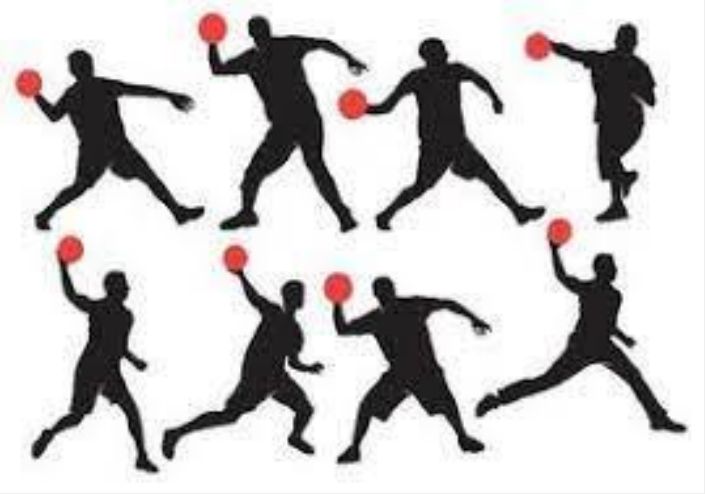
Analyzing the Secrets to Success in Competitive Dodgeball

Casey Dulaney, Conor Fagan, Neil Vakharia, Nidhi Boinapally, Sameer Rao, Vivian Moore



Introduction

Perhaps one of the most widely shared experiences amongst students around the world is playing dodgeball in their physical education class growing up. Dodgeball is a sport composed of two opposing teams that try to eliminate each other by throwing cloth balls and hitting opponents whilst simultaneously avoiding being hit themselves. There are minute variations to the rules in dodgeball–thus we adopted the World Dodgeball Federation (WDBF) ruleset for this project.

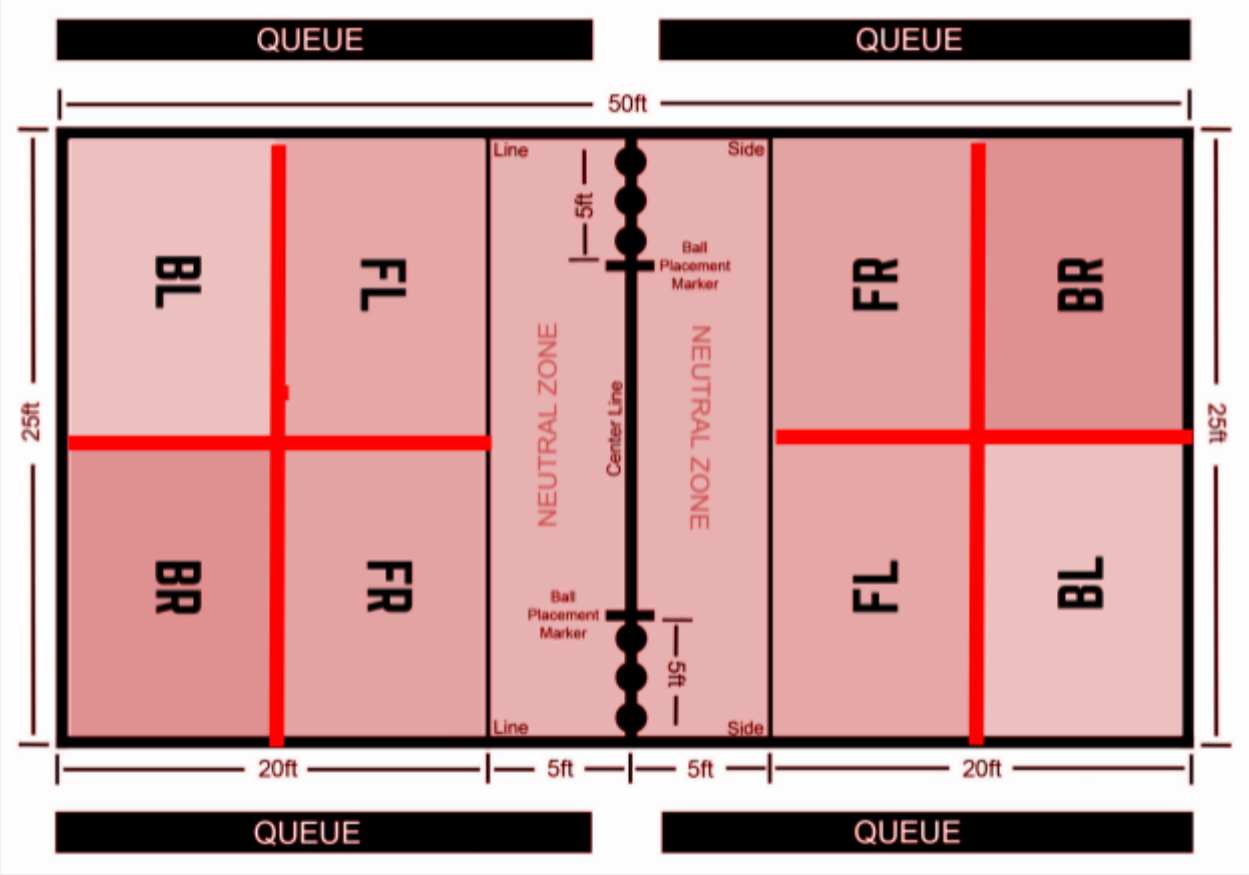
At the beginning of each set, two teams of 6 are encouraged to grab balls from the centerline, which is in the "neutral zone" (Rules). If a player is hit by an opposing player's throw, they are considered "out" and must exit the court. However, if a player's throw is caught, they themselves are eliminated and one player from the targeted team can return to play (Rules). Players are restricted to their own side and the neutral zone (a space where both teams can stand (Rules)). Stepping out of bounds results in an out (Rules). Teams play a set until completion: either time runs out or all players on one side are out (Rules). Teams win 2 points for winning a set (most players standing), 1 point for tying a set (same number of players standing), and 0 points for losing a set (no players or less players standing) (Rules).

Our initial hypothesis was that the easiest way to win would be to throw the most from as close as you can get to the opponent. In theory, by simple probability, reducing the distance and increasing the amount of throws intuitively should yield more hits. However, professional dodgeball has many zones, options for player positioning, and means of attack and defense coordination. Therefore, we want to investigate if this is true - are there any strategies or trends that lead to winning a dodgeball set. More specifically, do certain areas of the court lend to more hits, more outs, and such, do certain defensive arrangements or stances make it easier to survive attacks, and how can we identify a victory-oriented game script as opposed to a loss-oriented one. Dodgeball is not just a simple "throw and don't get hit" gym class game; it is complex, and our major purpose is to understand how players can use statistics to get a leg up.

To do this, we collected play-by-play data on various mixed (co-ed), cloth league dodgeball games from the 2022 WDBF Championships. The data is organized by set with each row representing a singular play. Here is an example of all the plays in one set:

ID	Game	Set	Timestamp	Attacking Country	# of Players on Side 1	# of Players on Side 2	Outcome	Attacker Zone	Attacker Position	Defender Zone	Defender Position	Defender Action
78	CAN vs. SWE	1	3:10:46	CAN	6	6	Miss	Neutral Zone	Standing	BR	Standing	None
79	CAN vs. SWE	1	3:10:47	CAN	6	6	Catch	Neutral Zone	Standing	BR	Standing	None
80	CAN vs. SWE	1	3:10:57	SWE	4	6	Miss	Neutral Zone	Standing	BL	Standing	None
81	CAN vs. SWE	1	3:10:57	SWE	4	6	Miss	Neutral Zone	Standing	BL	Standing	None
82	CAN vs. SWE	1	3:10:57	SWE	4	6	Miss	Neutral Zone	Standing	BL	Ducking	None
83	CAN vs. SWE	1	3:10:57	SWE	4	6	Hit	Neutral Zone	Standing	BR	Standing	None
84	CAN vs. SWE	1	3:11:09	SWE	3	6	Hit	Neutral Zone	Standing	BL	Standing	None
85	CAN vs. SWE	1	3:11:09	SWE	3	6	Miss	Neutral Zone	Standing	BL	Standing	None
86	CAN vs. SWE	1	3:11:22	SWE	2	6	Hit	BR	Standing	FR	Standing	Throwing
87	CAN vs. SWE	1	3:11:34	SWE	1	6	Hit	Neutral Zone	Standing	BL	Standing	None
88	CAN vs. SWE	1	3:11:34	SWE	1	6	Hit	Neutral Zone	Standing	BL	Standing	None
89	CAN vs. SWE	1	3:11:34	SWE	1	6	Hit	Neutral Zone	Standing	BL	Standing	None
90	CAN vs. SWE	1	3:11:34	SWE	1	6	Hit	Neutral Zone	Standing	BL	Standing	None

By looking at the data this way, we can see the progression of a set of dodgeball. Over the course of the past few weeks, each group member was responsible for (1) watching many sets of dodgeball (2) understanding important strategies and common trends (3) recording data for 1 or 2 sets. Every set from the championship tournament was posted on YouTube. At every throw, the video was paused and the variables were recorded. Between 11 sets of gameplay, we compiled 218 total plays. The values observed were ID (we generated), Game: the game being played, Set: whichever set was observed, Timestamp: the timestamp of the YouTube video that the observation was paused at, Attacking Country: whichever country threw the ball, # of Players on Side 1/Side 2: at the time of observation, how many players were on side 1 (the side that corresponds with the first country listed in the matchup)/side 2 (the side that corresponds with the second country listen in the matchup), Outcome: what the result of the throw was (hit, catch, miss), Attacker/Defender Zone: whether the ball was thrown from the back left, back right, front left, front right, (BL, BR, FL, FR) or neutral zone and which zone the ball was thrown to, Attacker/Defender Position: the form that the attacker took during the play (ducking, standing, jumping), and Defender Action: what the defender was doing at the time of the throw (None, Rebounding, Throwing). Rebounding is defined as trying to pick up a loose ball. Zones were measured in accordance with the court map below. To keep the data standard, we watched the first set of each game to get the most strategically neutral set per matchup.



Summary

Table 1.1: Statistics of Quantitative Variables

	Min	Q1	Median	Mean	Q3	Max	SD
# of Players on Side 1	1	3	5	4.42	6	6	1.64
# of Players on Side 2	1	3	5	4.10	6	6	1.75

Table 1.1 summarizes the statistics for the two quantitative variables we observed, the number of players on side 1 and the number of players on side 2. These variables represent the number of players that remained on each side of the court as the game progressed. We found that the five

number summary for each variable was the same. However, the mean and standard deviation for the number of players on each side displayed very slight differences.

Table 1.2: Frequencies & Relative Frequencies of Categorical Variables

	Catch	Deflection	Hit	Miss	Jumping	Standing	Ducking
Outcome	8	9	81	120			
	3.67%	4.13%	37.16%	55.05%			
Attacker Position					1	217	0
					.46%	99.54%	
Defender Position					37	101	80
					16.97%	46.33%	36.70%

	BL	BR	FL	FR	Neutral Zone	None	Rebounding	Throwing
Attacker Zone	12	18	8	9	171			
	5.50%	8.26%	3.67%	4.13%	78.44%			
Defender Zone	89	90	5	8	26			
	40.83%	41.28%	2.29%	3.67%	11.93%			
Defender Action						195	15	8
						89.45%	6.88%	3.67%

Table 1.2 displays the frequencies and relative frequencies for each of the categorical variables we observed. The relative frequencies are shown directly below the frequencies for the individual variables. By analyzing these frequencies we noticed that over half of the throws resulted in misses, only one attacker jumped while throwing, a majority of attackers threw from the neutral zone while a majority of defenders stayed in either the back left or back right of the court when they were shot at, and most defenders would not attempt to rebound or throw when they were shot at.

Figure 2: Multiple Bar Graph of Outcomes Based on Attacker Zone

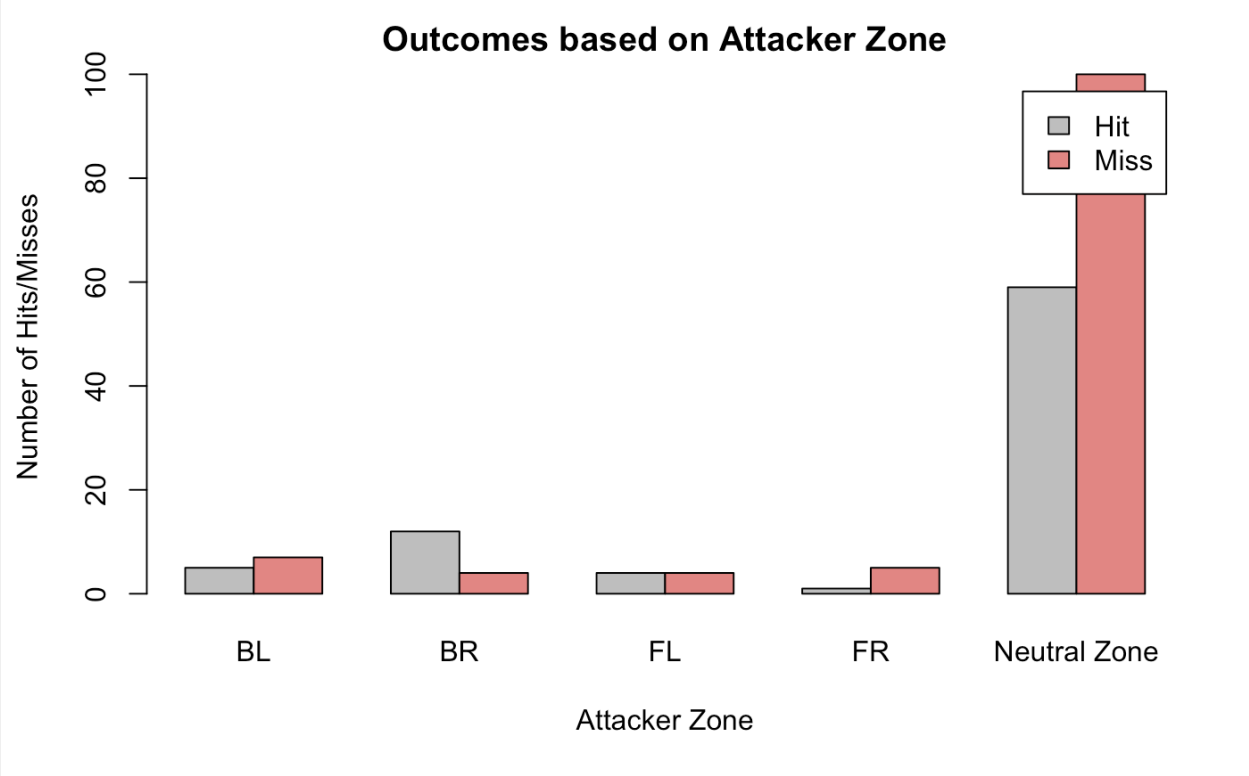


Figure 2 compares hits and misses from each zone. The chart shows that the neutral zone is, by far, the most popular spot to throw from. However, when compared to attacking zones such as FL or BR, attacks from the neutral zone (and BL and especially FR) are much more likely to be misses as opposed to hits.

Figure 3.1: Mosaic Plot of Count of Different Defender Positions by Outcome

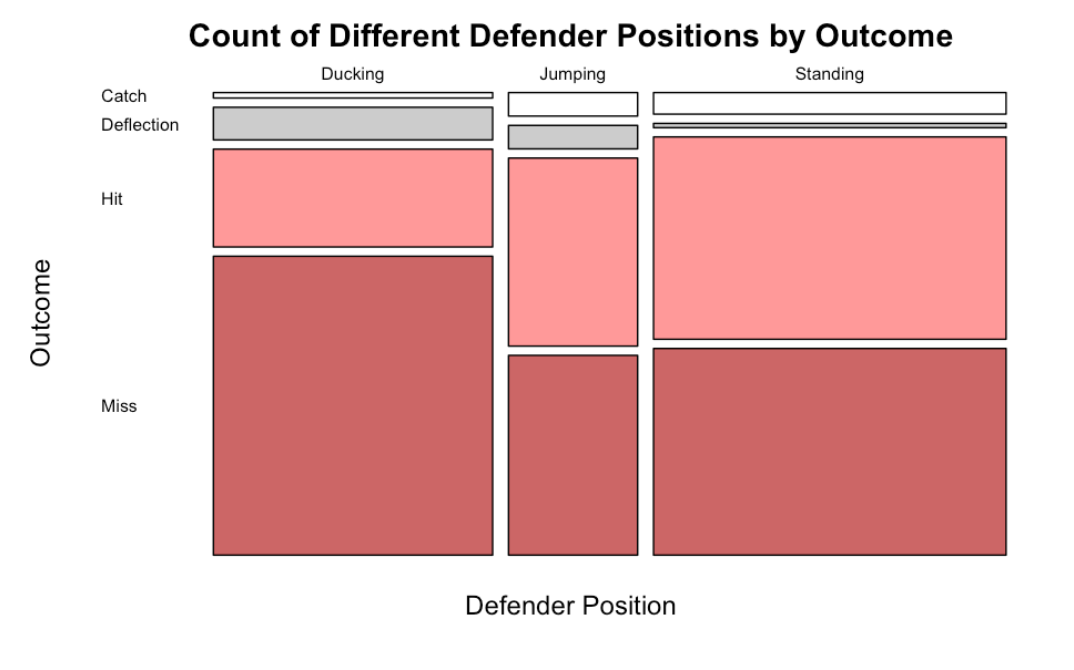


Figure 3.1 is a mosaic plot that looks at the relationship between defender position and outcome. The mosaic plot displays percentages of data broken down by groups. In our case, it takes the count of each defensive position (ducking, jumping, standing) and separates the count of each defensive position by the count of each outcome (e.g. the proportion of jumps by catches, deflections, hits, and misses). This results in a stacked bar chart representing the proportions of all 12 combinations (jump/catch, duck/hit, etc.).

Table 3.2: Frequency Table and Chi Square Test of Defender Positions by Outcome

	Defender Position		
Outcome	Ducking	Jumping	Standing
Catch	1	2	5
Deflection	6	2	1
Hit	18	16	47
Miss	55	17	48

X-squared	18.499
df	6
p-value	0.005099

Table 3.2 displays the results of a contingency table and chi-square test between the defender position and outcome variables. The contingency table displays the frequencies, and the resulting p-value from the chi-square test is 0.005099.

Insights

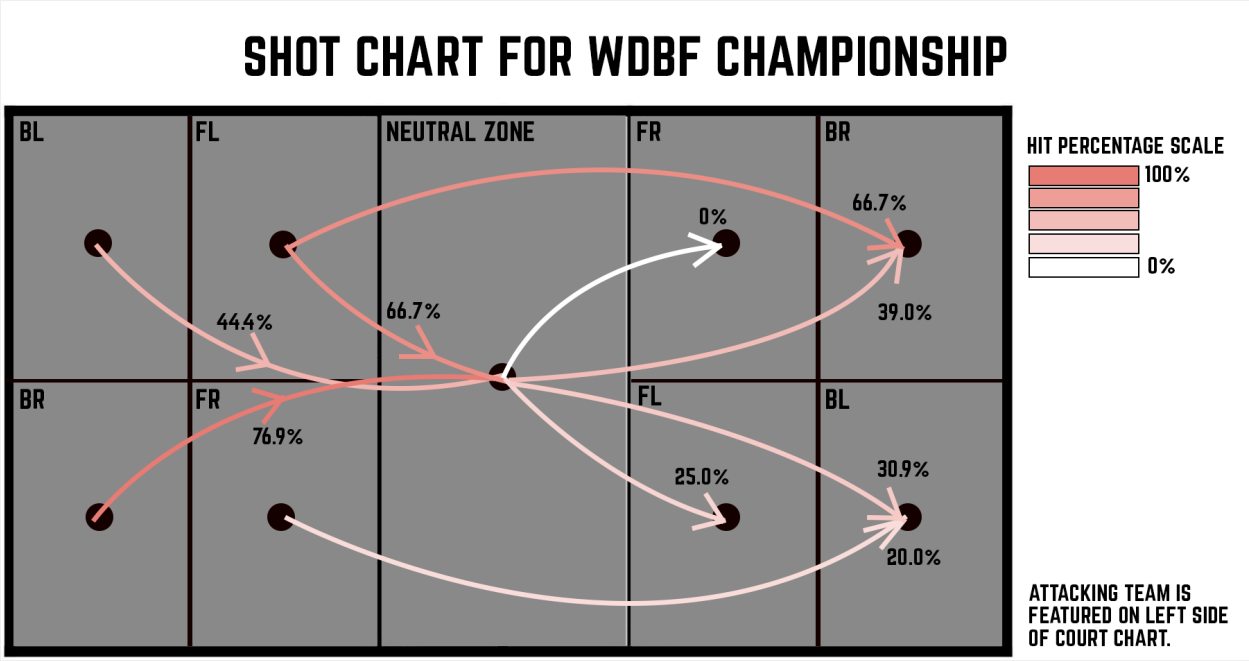
Understanding the Impact of Location on Hits:

As seen in **Table 1.2 and Figure 2**, there is a large disparity between attacks to and from the neutral zone when compared to other zones. The neutral zone is the largest zone by area, so it makes sense as to why so many attackers and defenders end up in the neutral zone. But from an efficiency standpoint, we wanted to understand whether packing the neutral zone on both sides of the ball was smarter or would it be better to increase the occupancy percentages of the other zones. **Figure 2** suggests that just because there is a high frequency of throws from a certain zone, that does not mean there is a higher success rate for getting hits. It is also not necessarily the case that high-hit attacking zones such as BR or FL remain high hit zones when isolating a single defensive zone. To answer our question, we must evaluate attacking and defending zone interactions instead of looking at them in isolation.

To quantify the effectiveness of throwing from one zone to another, first we must make a set of all potential zone-to-zone pairs. Each attacking zone {BR, BL, FR, FR, NZ} is paired with an equivalent defensive zone {BR, BL, FR, FR, NZ} to make 25 zone-to-zone pairs. Our primary measure for how successful a throw from the attacking part of the pair to the defensive pair was hit percentage, calculated by dividing the number of hits from the zone-to-zone pair by the number of throws from the zone-to-zone pair. We chose to ignore pairs that had 2 or less total throws. There are only 9 pairs that were not ignored: {BR-NZ, BL-NZ, FR-BL, FL-BR, FL-NZ, NZ-BR, NZ-BL, NZ-FR, NZ-FL}. Interestingly, there was a lack of full court throws (throws

from one back zone to another) and cross court throws (throws from the attacker's left to the defender's right or vice versa). The trend that barely any players attempt these throws implies that these throws rarely end in a hit, so attempting them would be unwise.

For the throws that were considered, their hit percentages were added on the shot chart seen below:



The shot chart operates similar to a heat map. The arrowed paths represent throws from the attacking side of the ball in their respective zone to the defensive. The color is "the heat" - a redder color means a higher hit percentage. As seen in the chart, there are 4 zone-to-zone pairs that have a high hit percentage - above 40%. These are {BL-NZ, BR-NZ, FL-NZ, FL-BR}. Most evidently, most high hit percentage pairs have the defender situated in the neutral zone. In addition, not one high hit percentage pair has an attacker situated in the neutral zone. Contrary to our initial hypothesis, it is not wise to be in the neutral zone whether you are an attacker or a defender - being in the neutral zone is high risk, low reward. As seen by the 4 pairs, it is much more successful to be on one's own side and throw balls at those who have overextended themselves by putting themselves in the neutral zone. It is worth noting that many high hit percentage pairs start in the leftmost zones of the court. Although this may be mostly circumstantial because there is not as much data from these zones, one possibility is since most dodgeball players are right handed, defenders are expecting throws from the attacker's right side

(the defender's left). Throws from the attacker's left side (the defender's right) might catch them off guard, leading to more hits.

Exploring the Relationship between Defender Positions and Throw Outcomes:

For our second insight, our group wanted to answer the question "how are certain defensive positions that the defender takes (a duck, jump, or some kind of movement standing up) related to the outcome of the throw?" To answer this question, we first created a mosaic plot, which can be observed in **Figure 3.1**. If there is no association between the position that a defender takes and the outcome of a throw, the mosaic plot should form a grid. The plot in **Figure 3.1** clearly does not form a grid, which gives initial evidence to the idea that there is some kind of relationship between defensive positions and outcomes. To further test the validity of the relationship between defensive positions and outcomes, a contingency table was created to perform a chi-square test. As seen in **Table 3.2**, the p-value obtained for the chi-square test between defensive positions and outcomes was significant at the 0.01 level, further building upon the evidence for a relationship between defensive position and outcome. Between the figure and the table, we can determine that there is a statistically significant relationship between defender position and outcome.

According to **Figure 3.1**, it seems that out of the three different defender position types, ducking has the highest proportion of misses, while standing and jumping share near identical proportions of misses. This information may be useful, as it prompts the idea that a defender will be more likely to miss a ball thrown at them if they duck. So, if the objective of the team is to keep their players from getting out, having the player duck when a ball is thrown at them might increase their chances of staying in the game. However as observed in **Figure 3.1**, ducking has a far lower proportion of catches than jumping and standing (which are once again, near identical). If the objective of a coach is for a defensive player to catch the ball in order to get an out in favor of their team, it may be best to tell the player to stand or jump instead of ducking. As evidenced in **Figure 3.1**, ducking had the highest proportion of deflections. If a team wants more balls on their side of the court, it may be best to tell the defensive player to duck. To summarize, ducking may be the best defensive position choice if the objective of a team is to increase the number of balls on their side of the court or miss the most balls. If the objective of a team is to increase the

number of player outs for the other team, standing or jumping may be the best defensive position choice.

Reflection:

While we stand by our analysis, there are a few critiques we have of our project. First, our data collection was sometimes inconsistent. The “defender and attacker position variable” was subjective, which could have led to inconsistencies caused by human error within our dataset. To make our analysis more precise and specific, we could have created more zones. A larger number of zones would allow for a closer examination of where the optimal throwing zones are. We also could have collected positional coordinate data of where exactly the players were on the court. Using numerical variables would have allowed for regression modeling and spatial data on a smaller scale, which also would have been more precise. Lastly, a larger overall sample size would have led to a more precise conclusion. Though we did sample more than 50 throws, our data was based off of a limited number of teams. The small sample size could have led to a conclusion that these team’s throws were more accurate or less accurate than the true average. We would need more teams to validate our conclusions more confidently. Additionally, some zone pairs have only a few samples. A larger sample size would be ideal to validate our findings.

References

"Rules." *World Dodgeball Federation*, https://worlddodgeballfederation.com/rules/.