UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PHYS 497 PROJECT

Using Statcast Data to Study the Baseball Bat Collision

Riku Komatani
ruk2@illinois.edu

supervised by
Alan Nathan

August 24, 2023

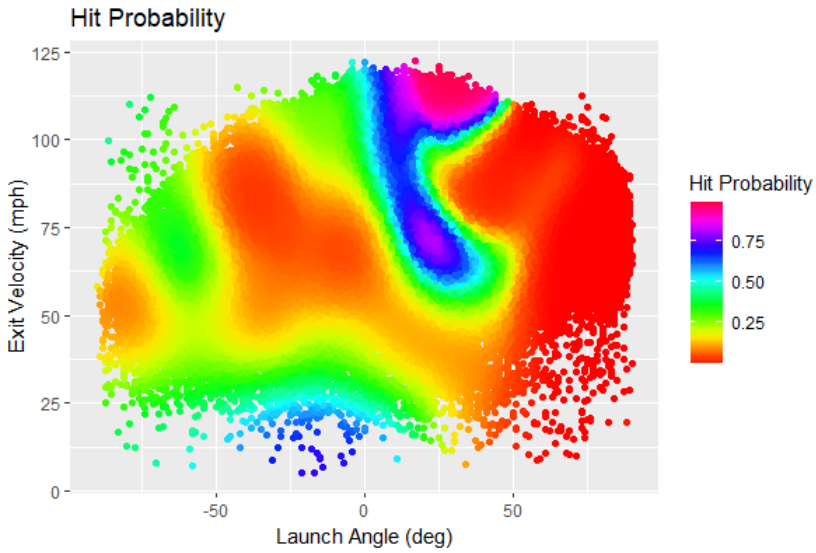

Figure 1: Color heat plot of hit probability for balls in play with respect to launch angle and exit velocity.

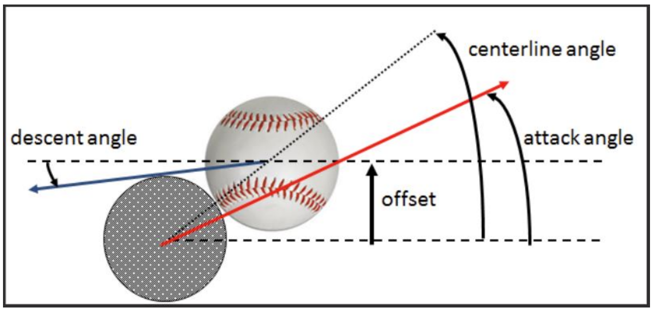

Figure 2: Terminology used in ball-bat collision.

1 Introduction

Baseball players focus on obtaining a good launch angle with high exit velocity when hitting. These two factors play a key role in terms of predicting the outcome of a batted ball. Figure 1 shows a color heat plot of hit probability for balls in play with respect to launch angle and exit velocity. For instance, a batted ball with low launch angle (less than 5 degrees) will likely result in a ground out regardless of its exit velocity, while a batted ball with launch angle of 25 degrees will result in a line drive single or could reach over the fence if the exit velocity is high enough. Of course, there are other factors that could change the outcome of a batted ball including spray angle and ball spin. But since they are hard to control by the hitter and are not as important as launch angle and exit velocity when predicting an outcome, we will only focus on these two for now.

Some might also argue that players like Ichiro Suzuki intentionally hit bloopers with high launch angle and low exit velocity to let the ball drop in between an outfielder and an infielder to earn a single. This effect can be seen from Figure 1 that a ball hit with launch angle of 35 degrees and exit velocity of 65mph has higher hit probability than a ball hit with same launch angle with exit velocity of 80mph. There are only few players in the MLB history that are capable of doing this, and most hitters tries to hit line drives with high exit velocity. So, we will obey the data and focus on the majority.

Looking at Figure 2, notice that the launch angle that gives the maximum exit velocity occurs when the attack angle is parallel to the center line angle, but the resulting launch angle is not always the optimal for producing a home run. However, achieving highest exit velocity results in highest chance of getting a hit. In this paper, we will explore the correlation between players' launch angle that gives the maximum exit velocity and their Homerun percent, wOBA, Slugging, Whiff rate, and Foul ball percent. I will use the data from Statcast, a tool used in Major League Baseball primarily for

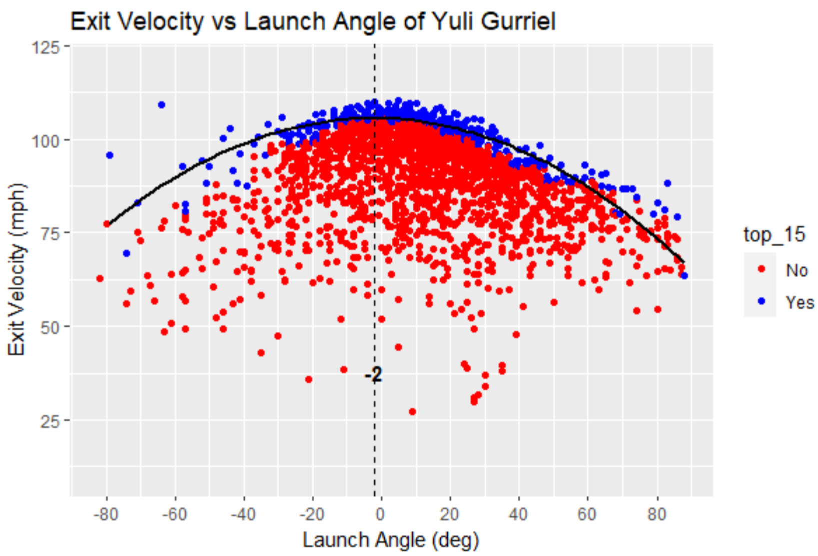

Figure 3: Comparison of Exit Velocity vs Launch Angle of player with lowest peak launch angle.

tracking the baseball. In addition, I will use the software package R to analyze and develop graphs that establishes connections between these factors.

For the sake of this paper, I will call the launch angle that gives the maximum exit velocity, a peak launch angle. Since different players have different attack angle and swing path, this peak launch angle differs across players. After finding the correlation between peak launch angle and various hitting stats, I will explore how each batters adjust to hit home runs with their peak launch angle.

2 Method

First, I imported the 2017 - 2019 & 2021 - 2023 seasons (excluding 2020 season due to the pandemic) MLB data from Statcast by using Statcast scraper by Bill Petti. I filtered this data to those that resulted in in-plays since we are interested in data that contains exit velocity and launch angle. Then, in order to obtain peak launch angle for each player, I divided the launch angle of each player's data into 5-degree bins, and within each bin, I selected data that is in top 15 percent by exit velocity. Next, I made a quadratic fit of exit velocity versus launch angle for the selected data above. Finally, I found the maximum point of the fitted line, which gives the peak launch angle and its exit velocity. It turns out that adding foul balls to this data frame changes player's peak launch angle by few degrees. This occurs since when the launch angle is either very high or low, the outcome tends to be foul balls, so foul balls appears in the top 15 percent by exit velocity, which will contribute to calculation of the peak launch angle. This effect will minimize if I filter the range of launch angles to those that likely results in in-plays. But for the sake of this paper, we will focus on in-play data (no foul balls) since foul balls isn't the optimal outcome of an event.

3 Analysis

3.1 Exit Velocity versus Launch Angle

To begin with, I wrote a function that creates a scatter plot of exit velocity versus launch angle for a single player with quadratic fit for top 15 percent data (by exit velocity for each 5-degree bins), and included a vertical dashed line through the peak launch angle. Using this function, I created plots for the player with minimum peak launch angle (-2 degrees) in Figure 3 and the player with maximum peak launch angle (33 degrees) in Figure 4. Notice how the quadratic fit of Yuli Gurriel is shifted to horizontally towards the left compared to that of Joey Gallo.

Let's first take a look at Yuli Gurriel's swing:
https://www.mlb.com/video/yuli-gurriel-homers-1-on-a-fly-ball-to-left-field-svadgo.

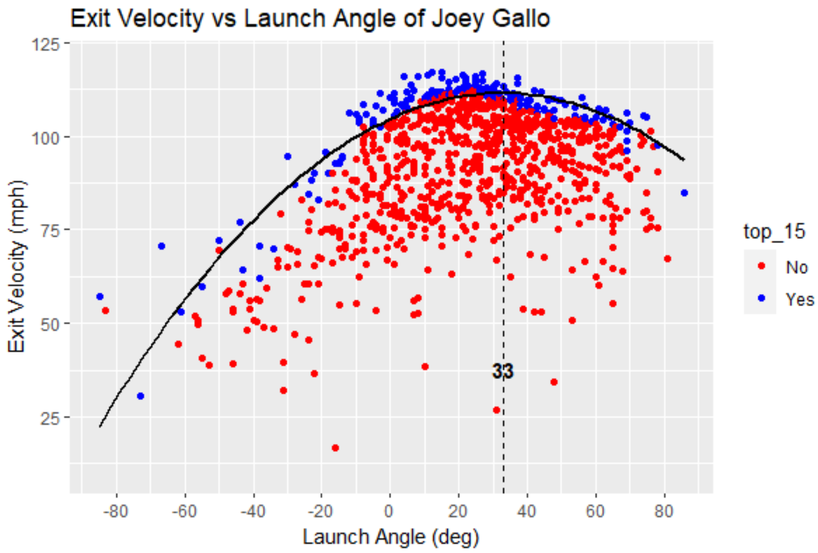

Figure 4: Comparison of Exit Velocity vs Launch Angle of player with highest peak launch angle.

His bat is leveled with the ball path as the ball approaches, with a low attack angle close to 0 degree. After his bat makes contact with the ball, his bat continues to follow the leveled path for a while. As shown by the video, his leveled swing path proves why his peak launch angle is very low.

Next, let's take a look at Joey Gallo's swing: `https://www.youtube.com/watch?v=-sBE5U6FAnA&t=4s`. His bat drops down fast as the ball approaches, with a high attack angle. After making a contact, he immediately brings up the bat and finishes his swing in a way similar to a golf swing where he brings up the grip above his shoulder and near his ear. This demonstrates how his uppercut swing relates to having such a high peak launch angle. Afterall, these examples indicate that given a launch angle, exit velocity varies a lot across player to player due to their variance in swing path.

3.2 Scaled Exit Velocity versus Launch Angle for Different Events

Next, let us define a phrase, scaled exit velocity. This is simply the exit velocity divided by the max exit velocity at the peak of the quadratic fit of each player. Using this, let's explore how each point of the graph represents different events. I selected Shohei Ohtani as the player, and color coded each point of scaled exit velocity vs launch angle graph by events shown in Figure 5. As expected, home runs tend to occur when the scaled exit velocity is near 1 and launch angle is in 20 - 30 degrees, which is a bit higher than the peak launch angle for Shohei Ohtani. In addition, most hits occur when the launch angle is close to the peak launch angle.

Let's compare the plots color coded by events for Yuli Gurriel and Joey Gallo. Figure 6 depicts the plot for Yuli Gurriel. When he hits the ball hard at his peak launch angle, the result tends to be a single. In order for him to hit a double, triple, or home run, he needs to hit the ball with higher launch angle than his peak launch angle. Figure 7 depicts the plot for Joey Gallo. When he hits the ball hard at his peak launch angle, the result tends to be a home run. And when hits the ball hard at lower launch angle from his peak launch angle, the outcome tends to be double & triple or a single. It's interesting to see that when players hit the ball at their maximum velocity, the result of the play is different depending on their swing path. Thus player's outcome on their plate appearance is shaped by their peak launch angle.

3.3 Peak Launch Angle Trend Comparison between Good Hitters

To observe the trend of peak launch angle and batting stats, from baseball savant, I selected players who were placed in top 10 for one or multiple seasons in the last three years for following categories: most home runs, highest batting average, and highest slugging percentage. Then I obtained

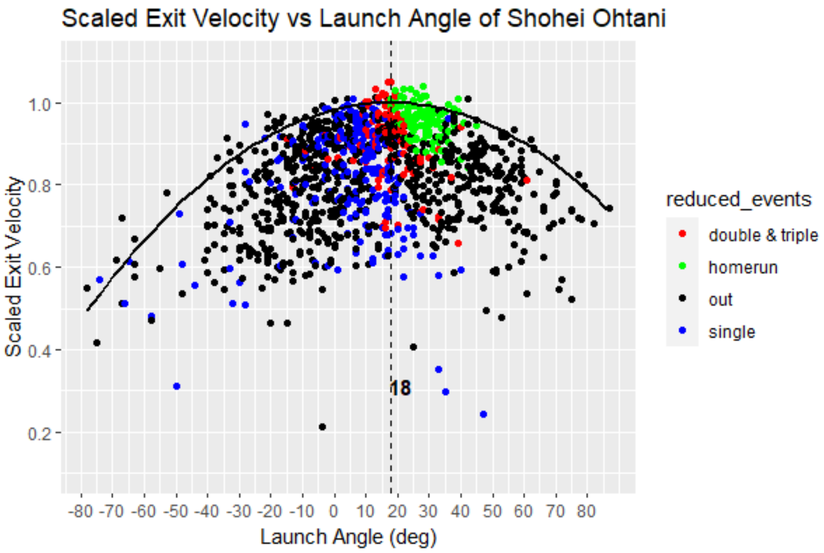

Figure 5: Scaled Exit Velocity vs Launch Angle of Shohei Ohtani, color coded by events.

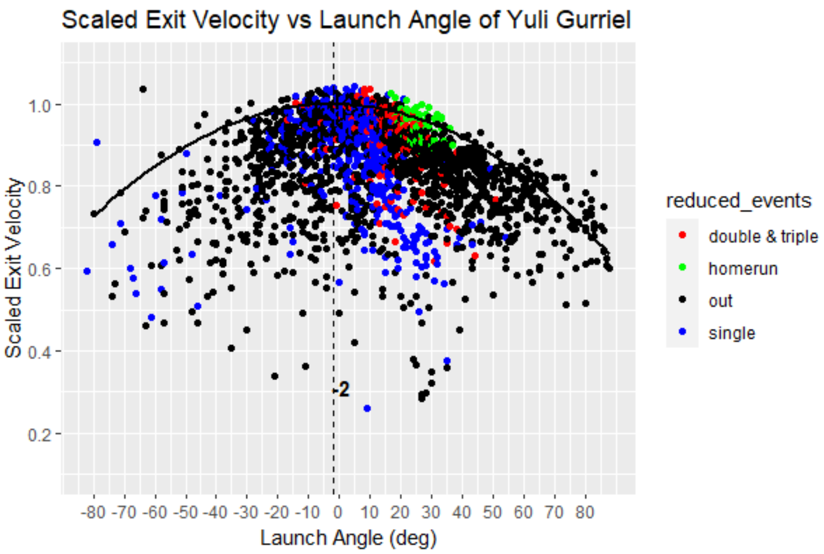

Figure 6: Scaled Exit Velocity vs Launch Angle of Yuli Gurriel, color coded by events.

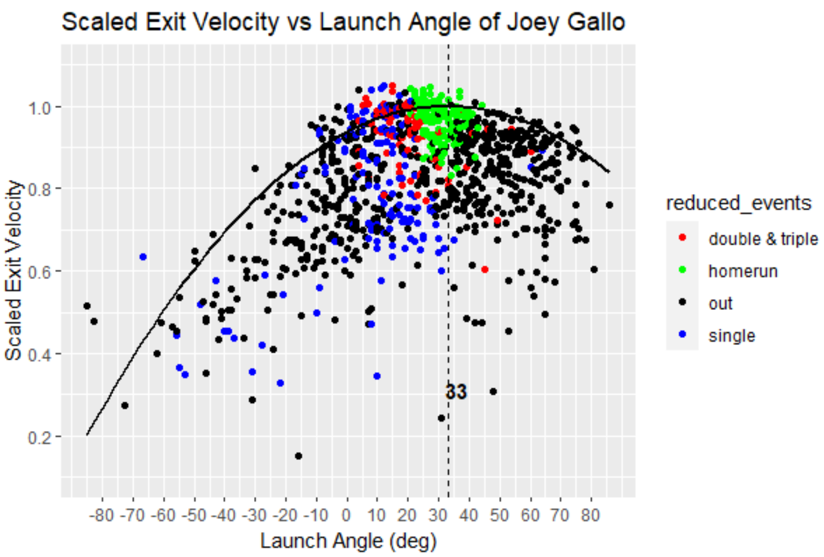

Figure 7: Scaled Exit Velocity vs Launch Angle of Joey Gallo, color coded by events.

peak launch angles for the same players in each category and computed the average. The results are as follows:

Mean peak launch angle for hitters who were place in top 10 over the last three years in:
- Most home runs ~ 15.36 $degrees$
- Highest batting average ~ 11.64 $degrees$
- Highest slugging percentage ~ 16.46 $degrees$

Although the peak launch angle varies a lot within each category, on average, hitters with top home runs and slugging percentage tend to have higher peak launch angle than hitters with top batting average. I also computed the peak exit velocity of the 10 players in each category, which is shown below:

Mean peak exit velocity for hitters who were place in top 10 over the last three years in:
- Most home runs ~ 111.30 mph
- Highest batting average ~ 109.45 mph
- Highest slugging percentage ~ 110.86 mph

This shows that on average, hitters with top home runs, batting average, and slugging percentage all have similar peak exit velocity. Although the data is not abundant, we could draw from this that hitters with high peak launch angle tends to hit home runs more frequently while the exit velocity of players aren't related to the peak launch angle.

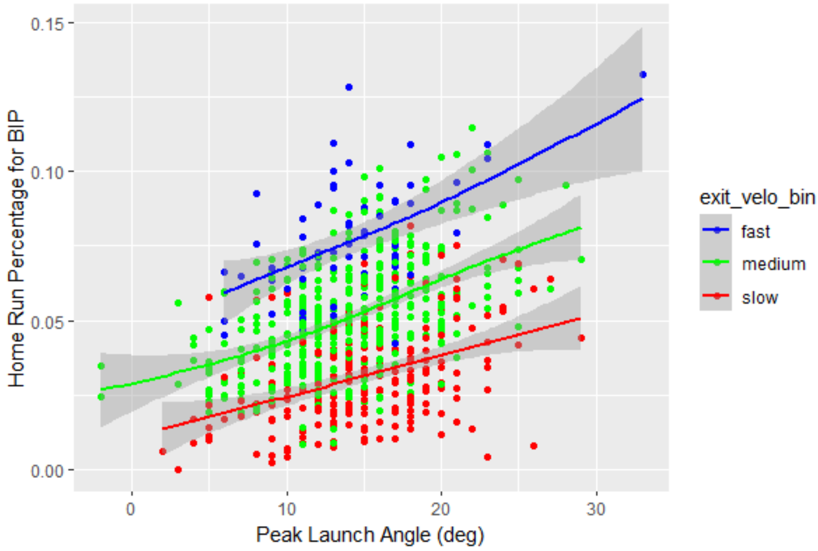

Figure 8: Scatter plot of home run percentage for balls in play versus peak launch angle, color coded into 3 exit velocity groups; blue for players with high exit velocity, green for players with medium exit velocity, and red for players with low exit velocity. Each group is fitted by GAM.

3.4 Correlation between Peak Launch Angle and Batting Statistics

To begin this section, I was interested in finding how batter's peak launch angle affects their performance. In order to observe the correlations, I plotted various batting statistics with respect to peak launch angle, and each point of the plots in this section represents a value for a single player.

3.4.1 Home Run Percentage for Balls in Play

First, I analyzed the correlation between peak launch angle and home run percentage for balls in play. To do this, I computed the home run percentage for balls in play of each batter by dividing their total home runs in a season by their total in play events per season. Then, I made a scatter plot of home run percentage for balls in play vs peak launch angle for each player. In addition, I color coded players into 3 exit velocity groups; blue for players with exit velocity greater than 110 mph, green for players with exit velocity less than 110 mph and greater than 105 mph, and red for players with exit velocity less than 105 mph. For each groups, I fitted a generalized additive model (GAM) on it, color coded by its group color. The result displayed in Figure 8 shows that for all 3 groups, home run percentage for balls in play tends to increase as the peak launch angle increases. Also, it is obvious that having a high exit velocity will result in higher home run percentage for balls in play. Note that exit velocity here refers to the maximum exit velocity of each players based on their quadratic fit.

3.4.2 Slugging Percentage for Balls in Play

Similarly to what I did for Figure 8, I created correlation plots for slugging percentage for balls in play with respect to peak launch angle; this time color coded by exit velocity without grouping. The slugging percentage shown in Figure 9 displays similar relationship with home run percentage, where the value increases as the peak launch angle increases. Again, players with high exit velocity tends to have higher slugging percentage for balls in play.

3.5 Batting Average for Balls in Play

Next, I plotted batting average for balls in play with respect to peak launch angle, color coded by exit velocity shown in Figure 10. In contrast to previous plots, the fitted line for this plot remains constant with respect to peak launch angle. Therefore, batting average isn't much determined by the peak launch angle. Again, players with high exit velocity tends to have higher batting average for balls in play.

Figure 9: Scatter plot of slugging percentage for balls in play versus peak launch angle with GAM, color coded by exit velocity.

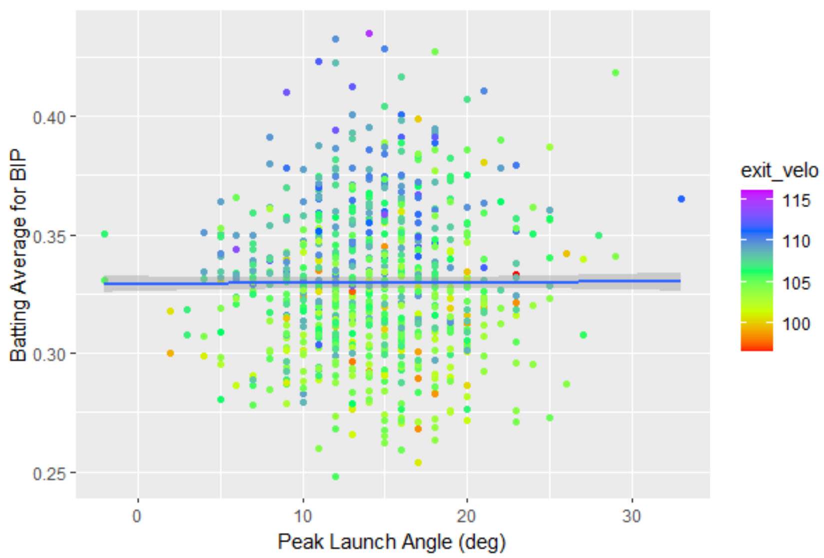

Figure 10: Scatter plot of batting average for balls in play versus peak launch angle with GAM, color coded by exit velocity.

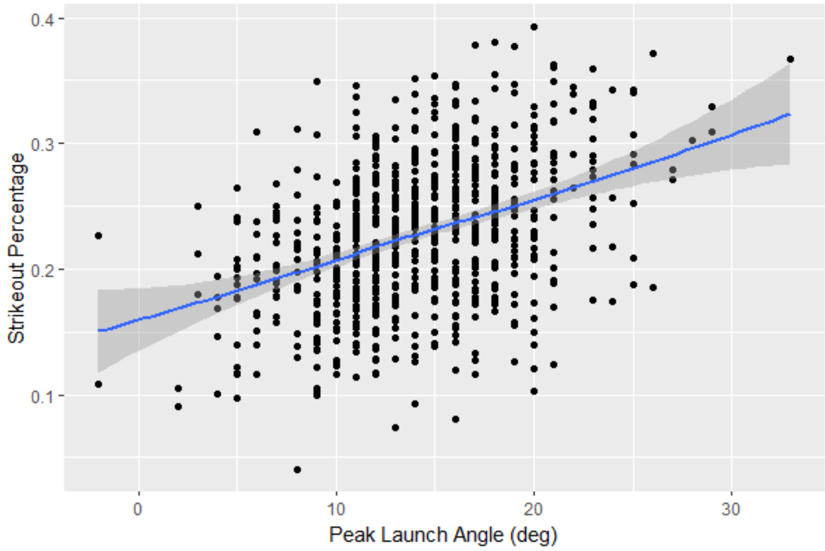

Figure 11: Scatter plot of strikeout percentage versus peak launch angle with GAM, color coded by exit velocity..

3.5.1 Strikeout Percentage & Whiff Rate

Then, I computed the correlation between peak launch angle and strikeout percentage. The strikeout percentage was calculated by summing the total strike outs for each player and dividing that number by total at bats for each player excluding events such as caught stealing, pick offs, catcher interference, stolen bases, challenges (other out), wild pitches, and game advisory (rain delay). After that, I graphed a scatter plot of strike out percentage vs peak launch angle and fitted a GAM over it. The result depicted in Figure 11 shows that strikeout percentage increases gradually as the peak launch angle increases.

To explore deeper into this, I created a graph of whiff rate (swing and miss percentage) versus peak launch angle for all pitches that were swung at. The whiff rate was calculated by dividing total swing and misses by total swings. The result shown in Figure 12 is similar to strikeout percentage, and whiff rate increases as the peak launch angle increases. Since the peak launch angle is directly related to attack angle of the bat, the above two graphs depicts that players with leveled swing tend to have lower swing and misses, and lower strikeout percentage than players with non-leveled swing. This result is reasonable since batters with leveled swing have a higher duration of time where they could contact the ball, where as batters with non-leveled swing can only contact the ball at the proper timing. For these plots, exit velocity doesn't play an important role in determining the percentage of outcomes.

3.5.2 Foul Percentage

Furthermore, I explored how foul percentages relates to peak launch angle. Foul percentages was calculated by dividing the amount of fouls by total swings. The result shown in Figure 13 has a fitted curve that peaks around 13 degrees. When compared to fitted curve for home run and slugging percentage, this curve is nearly flat over the range where the most players' peak launch angles are. This demonstrates that the foul percentage isn't much determined by player's peak launch angle. Likewise, exit velocity doesn't effect foul percentage.

3.5.3 wOBA for BIP and wOBA for ALL Events

Finally, I computed the relationship between wOBA and peak launch angle. I plotted two types of wOBA: wOBA for balls in play shown in Figure 14 and wOBA for all events shown in Figure 15. In order to compute these wOBAs, I used the "wOBA values" Statcast generates, and computed the mean

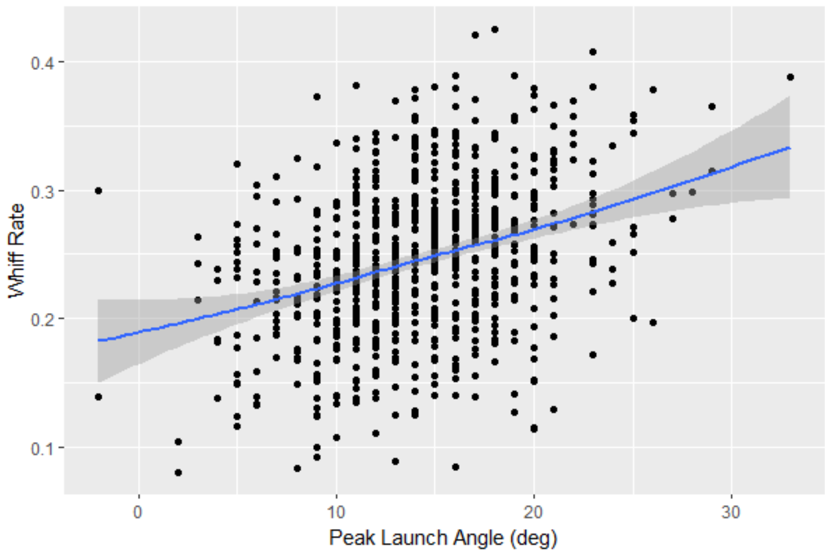

Figure 12: Scatter plot of swing and miss percentage versus peak launch angle with GAM, color coded by exit velocity.

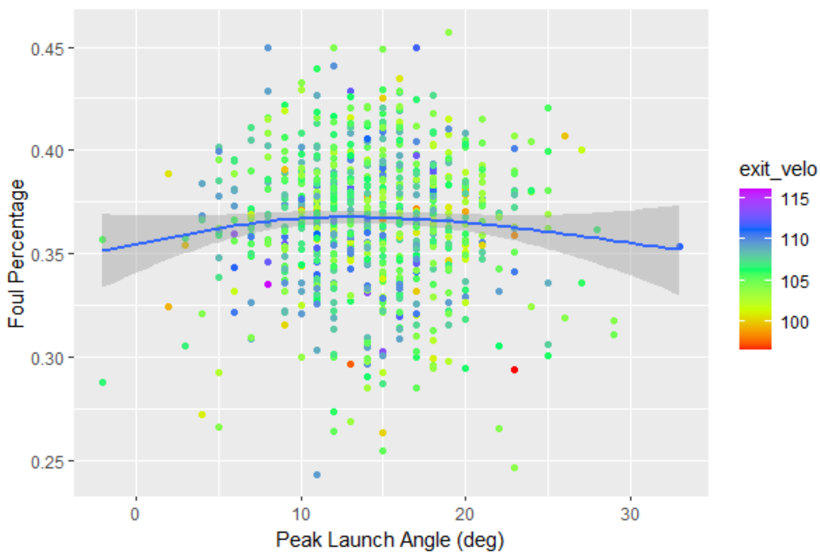

Figure 13: Scatter plot of foul percentage versus peak launch angle with GAM, color coded by exit velocity.

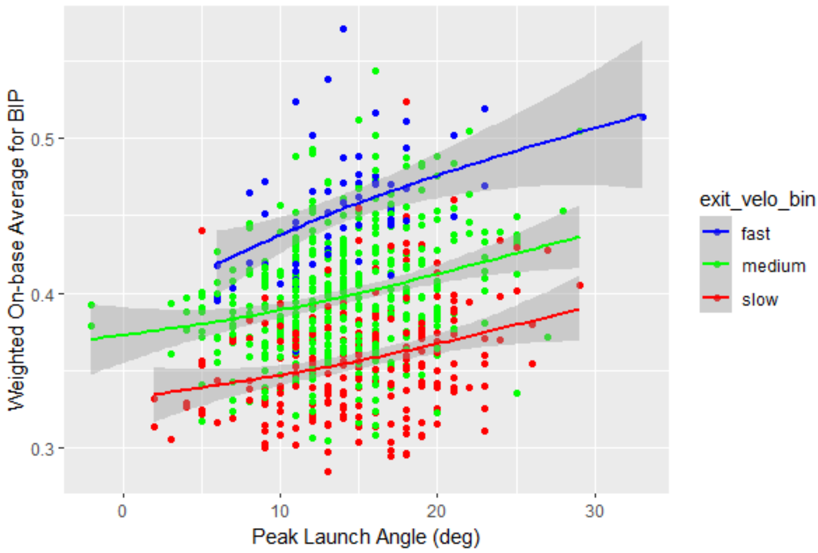

Figure 14: Scatter plot of weighted on base average for balls in play versus peak launch angle, color coded into 3 exit velocity groups; blue for players with high exit velocity, green for players with medium exit velocity, and red for players with low exit velocity. Each group is fitted by GAM

value for balls in play and for all events. Similar to plot for home run percentage, I color coded players into 3 exit velocity groups. From Figure 14, it is clear that wOBA for balls in play increases with higher peak launch angle. This was expected since this plot is technically an identical version of slugging percentage, but with different weights on hits. Next let's look at Figure 15 for wOBA including all events. This wOBA measures overall hitter's performance, including all at bats with non-NA wOBA value (including walks, hit by pitches, etc), with home run being weighed the highest. So this technically represents the combination of all previous plots. From Figure 11 and 12, I concluded that hitters with high peak launch angle tends to strikeout more often than hitters with low peak launch angle. However, Figure 8, 9, and 14 depicts that hitters with higher peak launch angle also has higher home run, slugging percentage, and wOBA for balls in play. From looking at Figure 15, the smooth fitted line indicates no change in wOBA as peak launch angle increases, thus the two effects seems to cancel out. Interestingly, this plot is closest to the batting average for balls in play plot in Figure 5. Thus,the amount of hits a batter obtain for balls in play regardless of its type could represent his overall batting status. Nevertheless, having a higher exit velocity will result in higher wOBA.

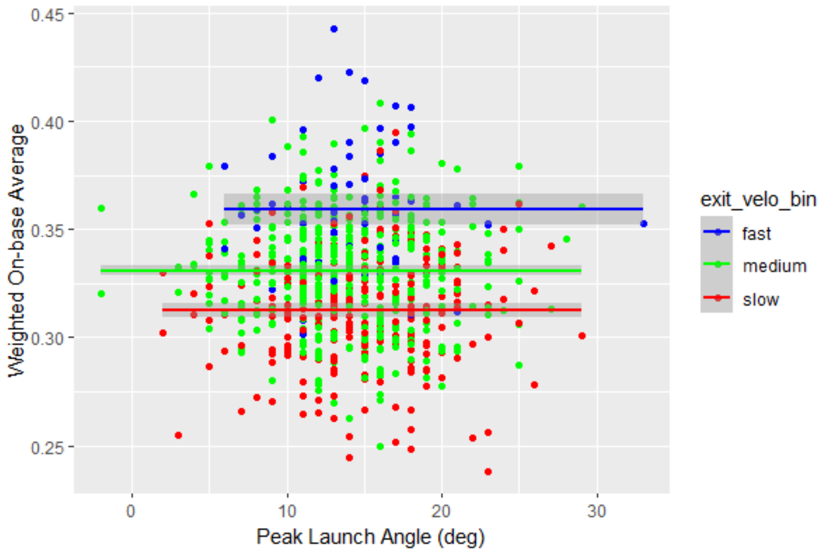

Figure 15: Scatter plot of home run percentage for all events versus peak launch angle, color coded into 3 exit velocity groups; blue for players with high exit velocity, green for players with medium exit velocity, and red for players with low exit velocity. Each group is fitted by GAM

4 Traverse Ball Bat Collision

In this section, I analyzed how attack angle and center line angle are related to launch angle and exit velocity. Figure 13 depicts what attack angle and center line angle means in ball-bat collision.

First, using Alan Nathan's calculations on Traverse-Ball-Bat-Collision, I obtained a function that returns a data frame of batted ball parameters from two inputs: attack angle and center line angle. Then I made a contour plot of scaled exit velocity from 0 (lowest) to 1 (highest) with attack angle in the y-axis and center line angle in the x-axis, and launch angle included as a dashed line, as shown in Figure 16. Notice how the scaled exit velocity is at maximum when attack angle is equal to centerline angle, depicted by black solid line on the figure.

Using the data from Traverse-Ball-Bat-Collision, I picked a certain attack angle for Yuli Gurriel and Joey Gallo, and see how that fits on the scaled exit velocity vs launch angle plot. Figure 17 shows the scaled exit velocity versus launch angle for Joey Gallo. For his quadratic fit curve with peak launch angle of -2 degrees, an attack angle of 6 degrees (as depicted with green points) seems to match closely. Additionally, I created a similar plot for Joey Gallo in Figure 18, who has a peak launch angle of 33 degrees. In order to match his quadratic fit curve, we highered the attack angle to 25 degrees, which overlaps it well. We can take away from these plots that there is a clear association between peak launch angle and attack angle. The green points moves depending on the initial parameters of the Traverse-Ball-Bat-Collision calculation, so our next step is to check whether the attack angle fitted for each player matches the actual attack angle for them, using the bat-tracking data.

4.1 Comparing contour plot to quadratic fit

In order to verify the relationship between peak launch angle and attack angle, we used bat-tracking data provided by the MLB, and plotted peak launch angle with respect to median attack angle - shown in Figure 19. The Simple Ball-Bat Collision Model which was derived from Traverse-Ball-Bat-Collision data, is shown in blue line. Looking at the plot, the relationship is approximately linear as predicted by the model. Now the reason why there are only few points on this plot is because this data was taken from 2022 where only two MLB stadiums captured bat-tracking data. Thus, as more data comes in, this relationship can be explored further.

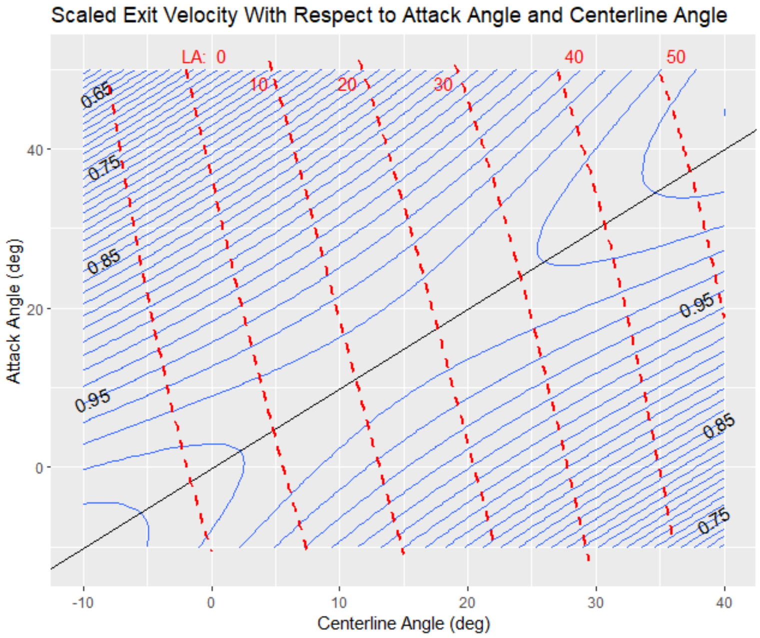

Figure 16: Contour plot of scaled exit velocity with respect to attack angle and centerline angle. Launch angle is depicted by the red dashed lines. Black solid line represents when attack angle is equal to centerline angle.

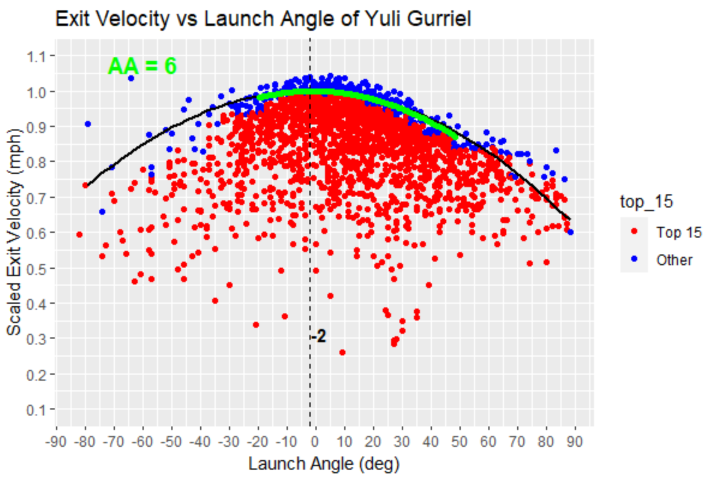

Figure 17: Scaled Exit Velocity with respect to Launch Angle for Yuli Gurriel, with green points depicting an Attack Angle of 6 degrees from Traverse-Ball-Bat-Collision data.

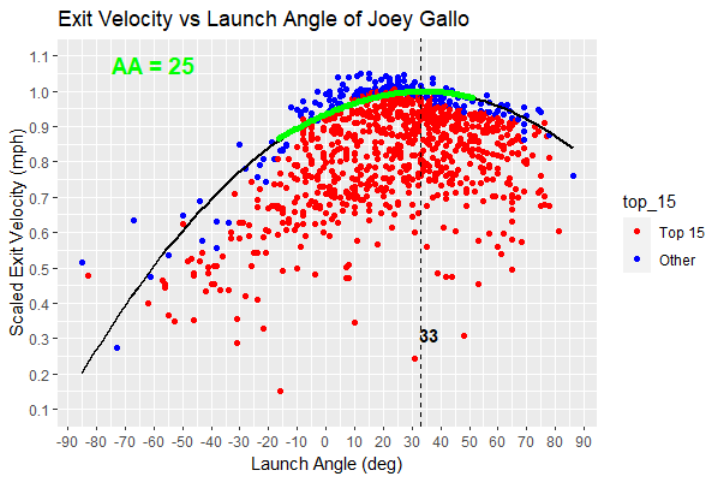

Figure 18: Scaled Exit Velocity with respect to Launch Angle for Joey Gallo, with green points depicting an Attack Angle of 25 degrees from Traverse-Ball-Bat-Collision data.

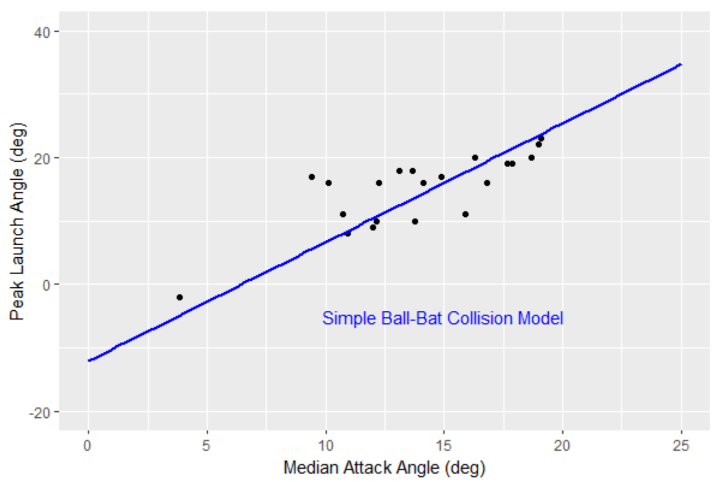

Figure 19: Peak Launch Angle with respect to Median Attack Angle of players. The Attack Angles are taken from bat-tracking data provided by the MLB. The blue line represents Simple Ball-Bat Collision Model, which is a model driven from Traverse-Ball-Bat-Collision data.