Analyzing Player Performance of MLB Silver Slugger Award Finalists

Samantha Nadler
Computer Science and Mathematics, Hofstra University

---

1. Introduction

The Major League Baseball (MLB) Silver Slugger is an annual award offered to the best offensive player at each position in both the National League and the American League, usually given out mid-November. The award, first awarded in 1980, acknowledges one catcher, four infielders (first baseman, second baseman, third baseman, shortstop), three outfielders (no distinction between outfield positions), one designated hitter, and one utility player (one who can play two or more of the aforementioned positions) in each league and are voted on by MLB coaching staff members unable to vote for representatives of their own team. While most voters consider batting average, slugging percentage, and on-base percentage as the most important metrics to determine each winner, the pitcher's behavior towards each batter's at-bats could play a part in determining newer metrics for a player's offensive performance.

1.1. Research Question

The main question posed by the provided data became: *How do Silver Slugger winners perform compared to their respective award category's runners-up?* This question considers details on a given pitch, the type of pitch thrown, the location of the pitch upon approaching a batter's strike zone, and over-time trends of estimated batting average and slugging percentage based on launch angle and speed to compare the batting performance of each position's top offensive players.

2. Methods

2.1. Data

The dataset *statcast_pitch_swing_data_20240402_20241030_with_arm_angle.csv* comprises data from over 750,000 pitches from the 2024 MLB season and includes over 100 Statcast variables that measure various aspects of pitching and batting characteristics, provided by Baseball Savant. The dataset was compiled by Brian Macdonald, senior lecturer and research scientist in Yale University's department of data science and statistics, and was created for the 2025 Connecticut Sports Analytics Symposium (CSAS) data challenge, encouraging participants to analyze some aspect of the pitcher-batter interaction.

2.2. Variables

Although the dataset contains 113 Statcast variables, only a few were selected to analyze various aspects of the at-bats of MLB's 2024 Silver Slugger Award finalists. Definitions of the following columns come from Baseball Savant:

player_name refers to the current batter in the at-bat.

description gives a brief summary of the outcome of the pitch.

pitch_type refers to the type of pitch thrown during the at-bat, derived from Statcast technology[1].

plate_x and *plate_z* refer to the horizontal and vertical position of the pitch as it crosses home plate as observed by the catcher, measured in feet[2].

launch_speed and *launch_angle* refer to the exit velocity and angle the moment the ball comes in contact with the bat. Launch speed is measured in miles per hour while launch angle is measured in degrees.

estimated_ba_using_speedangle and *estimated_slg_using_speedangle* refer to the batter's expected batting average and slugging percentage using the recorded launch speed and angle of a given at-bat[3].

2.3. Data Analysis

This research relies on Python's common data analysis and visualization packages to present trends in pitch type distributions, pitch locations, and time series analysis to study expected batting trends for Silver Slugger Award finalists. Matplotlib was mainly used as a visualization tool to create pie charts, scatter plots, and line graphs, while Pandas was used to determine measures of centrality and variability in given columns and NumPy was used for array construction and numerical analysis.

Due to the number of finalists and time constraints, only one group of finalists was analyzed: National League shortstop finalists. Three smaller datasets were constructed from the provided dataset, each where *player_name* was equal to its respective finalist: winner Francisco Lindor and runners-up Willy Adames and Elly De La Cruz. Each smaller dataset ranged around 2,500 entries. These datasets were then filtered to only include columns where the value of the *description* column was equal to *hit_into_play*, ensuring the data analysis was only performed on hits that ended each batter's at-bat. Pie charts were created to show the distribution of unique pitch types that occurred at least 5% of the time, to understand how diverse the pitches each batter faced were.

From there, a series of scatter plots were created using Matplotlib to visualize the locations of 100 sampled in-play balls for each batter. Since each batter had on average 450 in-play balls during the 2024 season, about a quarter were randomly sampled, using Pandas's `sample()` function, for a cleaner visualization.

The distance of each pitch from the center of the strike zone (0.0, 0.0), as represented in the columns *plate_x* and *plate_z*, respectively, was then calculated using the Euclidean distance formula, as pitches thrown closer to the center of the strike have a higher chance of being hit as an extra-base hit. New arrays were created with NumPy. The arrays were then combined into a Pandas DataFrame, and descriptive statistics, including measures of centrality and variability, were displayed to understand the spread of each batter's pitch locations.

---

[1] The following Statcast pitch abbreviations are used in the dataset: 'CH' - Changeup, 'CU' - Curveball, 'FC' - Cutter, 'EP' - Eephus, 'FA' - Fastball (general), 'FO' - Forkball, 'FF' - Four-seam fastball, 'KN' - Knuckleball, 'KC' - Knuckle-curve, 'PO' - Putout, 'SC' - Screwball, 'SI' - Sinker, 'SL' - Slider, 'SV' - Slurve, 'FS' - Splitter, 'ST' - Sweeper.

[2] Negative values for *plate_x* signify the pitch crossed home plate to the left of the catcher's center.

[3] The variables *estimated_ba_using_speedangle* and *estimated_slg_using_speedangle* are not accurate readings of the batter's batting average and slugging percentage at the time and are just approximations using other data.

Time-series analysis was performed on the columns *estimated_ba_using_speedangle* and *estimated_slg_using_speed_angle* to analyze the over-time changes in each batter's estimated batting average and slugging percentage. While the given values may not reflect the batters' actual statistics, they can give a good idea of a player's batting performance and any possible archetypes. Similar to the previous step, each player's values for *estimated_ba_using_speedangle*, *estimated_slg_using_speedangle*, *launch_speed*, and *launch_angle* were each combined into a separate Pandas DataFrame, and descriptive statistics, including measures of centrality and variability, were displayed to understand if their contact rate, launch speed, and exit velocities, indicate their batting archetype.

3. Results

Figs. 1 shows a trio of pie charts displaying the distribution of each unique pitch seen in at least 5% (about at least 100 pitches) of all pitches thrown to each Silver Slugger finalist. An abbreviation key is provided in Footnote 2.

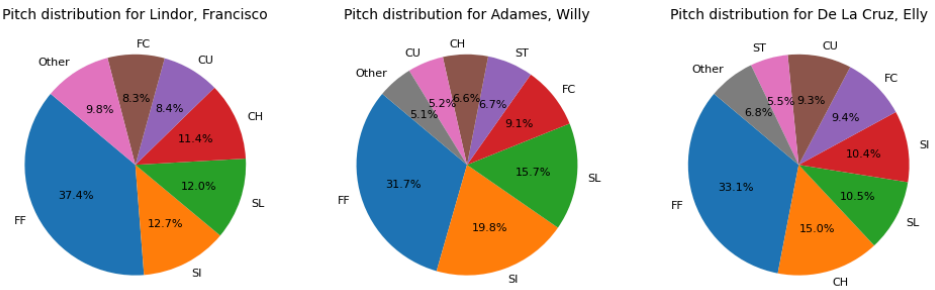

Fig. 1. Pitch distributions for Francisco Lindor, Willy Adames, and Elly De La Cruz.

Lindor, the winner for the National League shortstop category, faced the fewest unique pitches thrown at least 5% of the time, while his fastball percentage (the most common pitch thrown by MLB pitchers) was just over 4% higher than the next-highest fastball percentage from De La Cruz. As Lindor faced the least amount of unique pitches thrown at the 5% threshold, he was being thrown more consistent pitches, which over time became easier to identify and hit off.

Fig. 2 shows the final locations of 100 random pitches hit into play by Lindor, Adames, and De La Cruz, measured in feet from the center of the strike zone as noted in Footnote 3.

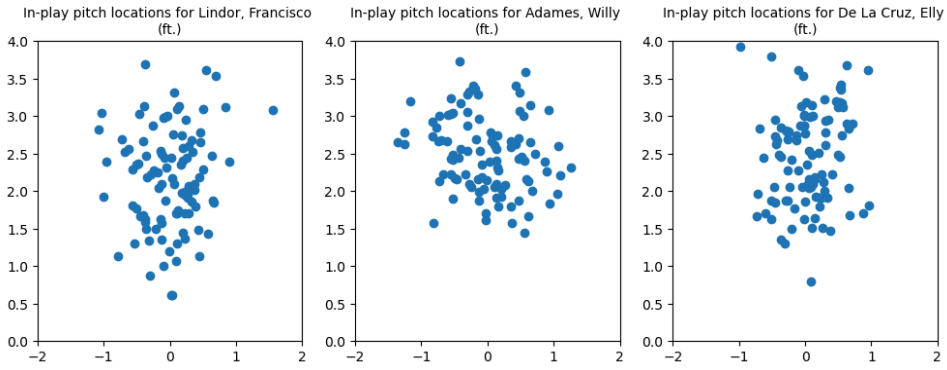

Fig. 2. In-play final pitch locations for Francisco Lindor, Willy Adames, and Elly De La Cruz.

Visually, the pitches thrown to Lindor appear more symmetric and closer to the center of the strike zone, predetermined to be (0.0, 2.0). Lindor was thrown more balls closer to the ground but also had some thrown higher up. However, Adames and De La Cruz saw more pitches higher in the strike zone. Pitches that are closer to the center of the strike zone are deemed “easier” and have a higher chance of being hit as an extra-base hit.

Figs. 3a, 3b, and 3c are taken from the lists of final pitch locations in relation to the center of the strike zone (0.0, 2.0), using the Euclidean distance formula to calculate each pitch location’s distance from the center of the strike zone. Fig. 3a contains data of the distances of pitches thrown to Lindor, while Fig. 3b contains data of the distances of pitches thrown to Adames, and Fig. 3c contains data of the distances of pitches thrown to De La Cruz.

```
# Sort Lindor's distances for ease of viewing differences in distances
lindor_distance = np.sort(lindor_distance)
print('Sorted distances from center of strike zone from Lindor\n', lindor_distance)
```

```
Sorted distances from center of strike zone from Lindor
 [0.12206556 0.13892444 0.16401219 0.17464249 0.17464249 0.1811077
 0.23021729 0.23537205 0.27166155 0.27730849 0.28460499 0.29832868
 0.31016125 0.31400637 0.31622777 0.32557641 0.35846897 0.36235342
 0.36796739 0.37013511 0.37054015 0.38078866 0.38483763 0.39560081
 0.39824616 0.40311289 0.43680659 0.43829214 0.4472136  0.4472136
 0.45044423 0.47539457 0.51419841 0.51971146 0.52009614 0.52325902
 0.54626001 0.55785303 0.55803226 0.55901699 0.55946403 0.58668561
 0.59135438 0.5920304  0.60083276 0.60083276 0.6060528  0.61611687
 0.62128898 0.62425956 0.63953108 0.65306967 0.66098411 0.66211781
 0.6670832  0.67067131 0.67683085 0.70859015 0.73348483 0.7558439
 0.76164296 0.77129761 0.78032045 0.7940403  0.79630396 0.8000625
 0.81320354 0.84291162 0.85603738 0.87800911 0.90553851 0.91760558
 0.93434469 0.9749359  0.97984693 0.98086696 0.98412398 0.99247166
 0.99724621 1.00498756 1.01079177 1.01242284 1.03619496 1.10548632
 1.12805142 1.1374533  1.15948264 1.16846053 1.19540788 1.20341182
 1.3118689  1.3480727  1.38057959 1.39014388 1.4        1.46372812
 1.69162644 1.71081852 1.73219514 1.8973666 ]
```

Fig. 3a. Code snippet and sorted list of pitch distances from the center of the strike zone, using Euclidean distance formula, for Francisco Lindor.

```
# Sort Adames's distances for ease of viewing differences in distances
adames_distance = np.sort(adames_distance)
print('Sorted distances from center of strike zone from Adames\n', adames_distance)
```

```
Sorted distances from center of strike zone from Adames
 [0.05830952 0.12529964 0.13038405 0.15033296 0.16643317 0.17204651
 0.17691806 0.21023796 0.21213203 0.21260292 0.21540659 0.24738634
 0.26400758 0.28160256 0.29732137 0.30066593 0.31906112 0.34438351
 0.37107951 0.39051248 0.39115214 0.40816663 0.41048752 0.42059482
 0.48507731 0.49648766 0.5186521  0.52952809 0.54341513 0.56727418
 0.57723479 0.59236813 0.60166436 0.61814238 0.62225397 0.6226556
 0.62369865 0.63702433 0.668431   0.6726812  0.68007353 0.68680419
 0.68818602 0.69115845 0.69404611 0.70235319 0.70604532 0.70837843
 0.71309186 0.71505245 0.73437048 0.741485   0.74706091 0.75504967
 0.75716577 0.78160092 0.78492038 0.85146932 0.87658428 0.9060905
 0.91923882 0.92590496 0.92973114 0.93680307 0.95341491 0.9535198
 0.9687621  0.97590983 0.99824847 1.05042848 1.09480592 1.10535062
 1.11986606 1.14127122 1.14690889 1.16275535 1.17239072 1.17686023
 1.18338498 1.22674366 1.23648696 1.24619421 1.24651514 1.29869165
 1.3        1.3160927  1.3244244  1.35915415 1.36058811 1.38311243
 1.3986422  1.40872283 1.41566239 1.41873183 1.46454771 1.48189068
 1.50734867 1.67597733 1.68908259 1.78025279]
```

Fig. 3b. Code snippet and sorted list of pitch distances from the center of the strike zone, using Euclidean distance formula, for Willy Adames.

```
# Sort De La Cruz's distances for ease of viewing differences in distances
de_la_cruz_distance = np.sort(de_la_cruz_distance)
print('Sorted distances from center of strike zone from De La Cruz\n', de_la_cruz_distance)
```

```
Sorted distances from center of strike zone from De La Cruz
 [0.07211103 0.07211103 0.09848858 0.10816654 0.14       0.15620499
0.16763055 0.17888544 0.21260292 0.22803509 0.23769729 0.24738634
0.26570661 0.28017851 0.29017236 0.29068884 0.29546573 0.29732137
0.30083218 0.31780497 0.33970576 0.35128336 0.35171011 0.37054015
0.38209946 0.39       0.46957428 0.47010637 0.47381431 0.48383882
0.49030603 0.50009999 0.53235327 0.53366656 0.53851648 0.54781384
0.55470713 0.59033889 0.62481997 0.63007936 0.65215029 0.66121101
0.66640828 0.67977938 0.69231496 0.69296465 0.70178344 0.72111026
0.73925638 0.74094534 0.74249579 0.75133215 0.76557168 0.76843998
0.76843998 0.77       0.79924965 0.80529498 0.82540899 0.8357033
0.83815273 0.85755466 0.8637708  0.87005747 0.87367042 0.88022724
0.9202717  0.92703829 0.93600214 0.97862148 0.98671171 0.99247166
1.00603181 1.0100495  1.01242284 1.02019606 1.0511898  1.05422958
1.08706026 1.09859001 1.1315918  1.14634201 1.15433964 1.18016948
1.20266371 1.22249744 1.25399362 1.25670203 1.28875909 1.29356098
1.29711218 1.4539945  1.47827602 1.51921032 1.53029409 1.61375339
1.79424079 1.86123615 1.86938493 2.15564376]
```

Fig. 3c. Code snippet and sorted list of pitch distances from the center of the strike zone, using Euclidean distance formula, for Elly De La Cruz.

Table 1 contains descriptive statistics for measures of centrality and variability for the pitch distances thrown to Lindor, Adames, and De La Cruz.

	Lindor, Francisco	Adames, Willy	De La Cruz, Elly
count	100.000000	100.000000	100.000000
mean	0.716812	0.800285	0.766308
std	0.391037	0.422588	0.444619
min	0.122066	0.058310	0.072111
25%	0.401896	0.493635	0.388025
50%	0.631895	0.724711	0.741721
75%	0.980102	1.150871	1.014366
max	1.897367	1.780253	2.155644

Table 1. Descriptive statistics for measures of centrality and variability for the pitch distances thrown to Francisco Lindor, Willy Adames, and Elly De La Cruz.

While Adames has the smallest distance from the center of the strike zone, Lindor has the smallest mean, median, standard deviation, and interquartile range. A smaller spread of Lindor's pitch data signifies the pitches thrown to him were more consistent then those thrown to Adames or De La Cruz. Additionally, a lower mean distance from the center of the strike zone implies pitches thrown to Lindor were more central to the strike zone, making hitting, especially hitting extra-base hits, easier compared to Adames or De La Cruz.

Figs. 4a, 4b, and 4c show a time-series analysis of each batter's estimated batting average and slugging percentage over time, using launch angle and exit velocity (launch speed) for each at-bat.

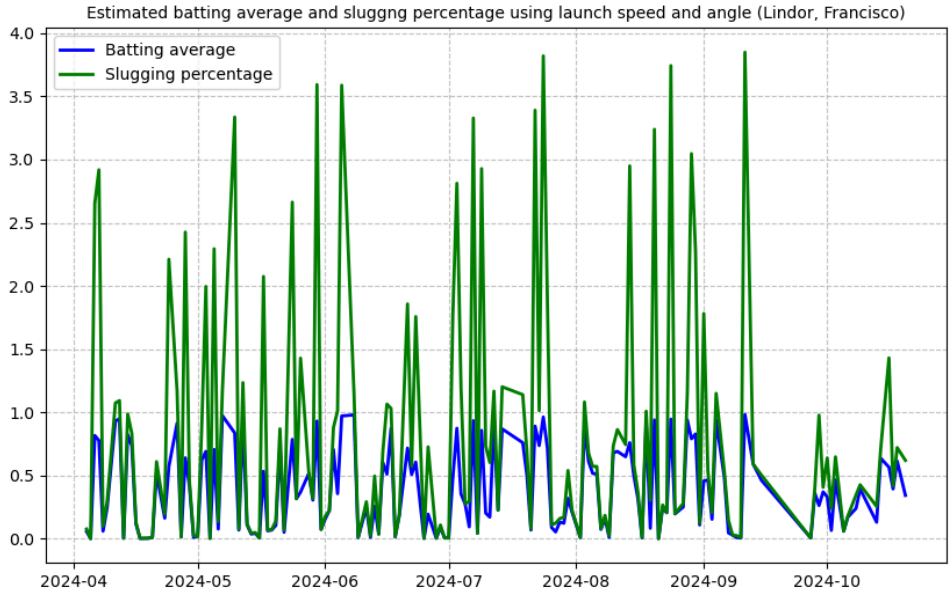

Fig. 4a. Estimated batting average and slugging percentage over time, using launch speed and angle, for Francisco Lindor.

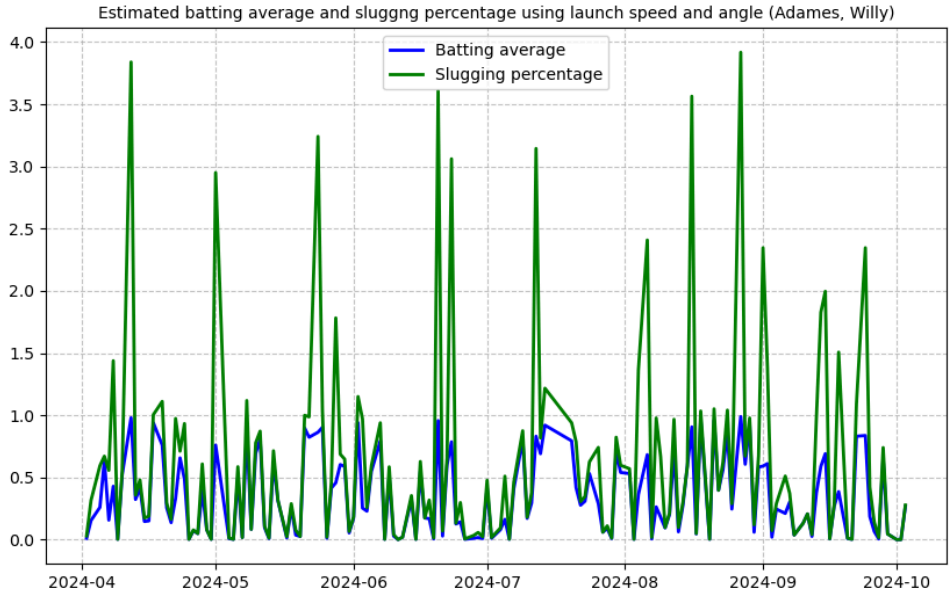

Fig. 4b. Estimated batting average and slugging percentage over time, using launch speed and angle, for Willy Adames.

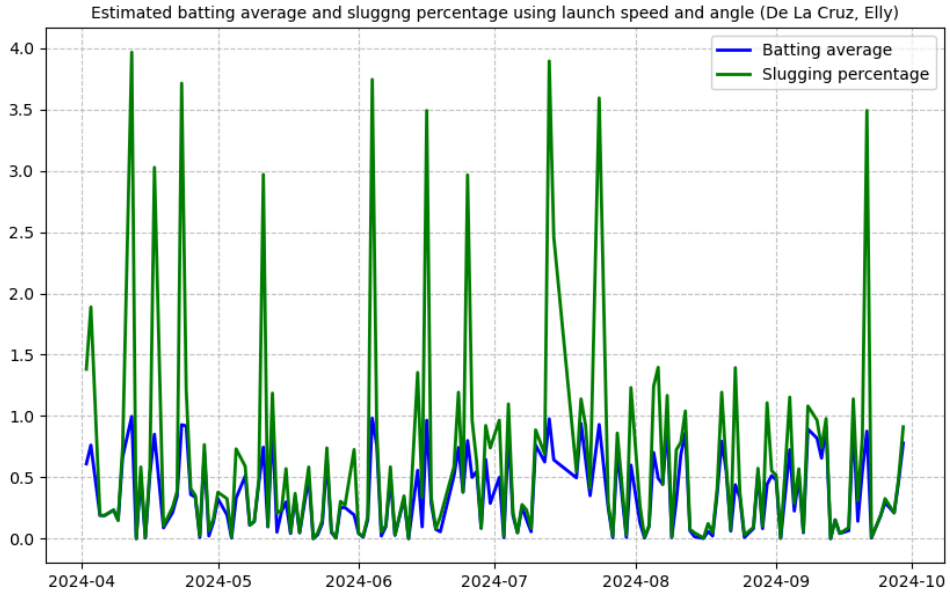

Fig. 4c. Estimated batting average and slugging percentage over time, using launch speed and angle, for Elly De La Cruz.

From these estimations, it can be concluded that Lindor, Adames, and De La Cruz may all fit the "power hitter" archetype; that is, these players all tend to have high slugging percentages that suggest that they often hit extra-base hits, including many home runs to put their slugging percentage close to a perfect but very rare 4.000.

Fig. 5 shows the relation between launch angle (in degrees) and launch speed (in miles per hour) of 360 randomly sampled in-play balls hit by Lindor, Adames, and De La Cruz.

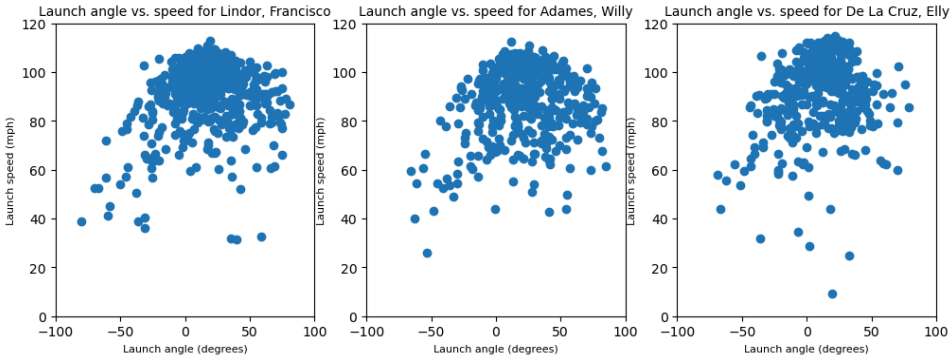

Fig. 5. Relations between launch angle (degrees, on the x-axis) and launch speed (miles per hour, on the y-axis) for in-play balls hit by Francisco Lindor, Willy Adames, and Elly De La Cruz.

Just by visualizing, it is clear that each batter consistently hits in a given cluster where their launch angle and speed are similar. However, Adames appears to hit the ball at a larger angle compared to the other two batters, while De La Cruz appears to hit the ball at a faster speed.

Table 2 contains descriptive statistics for measures of centrality and variability for launch angles of in-play balls hit by Lindor, Adames, and De La Cruz.

	Lindor, Francisco	Adames, Willy	De La Cruz, Elly
count	360.000000	360.000000	360.000000
mean	17.863889	21.513889	9.111111
std	28.221583	26.869432	25.088454
min	-80.000000	-66.000000	-69.000000
25%	3.000000	3.750000	-7.000000
50%	18.000000	24.000000	10.000000
75%	37.000000	39.000000	26.000000
max	81.000000	85.000000	79.000000

Table 2. Descriptive statistics for measures of centrality and variability for launch angles of in-play balls hit by Francisco Lindor, Willy Adames, and Elly De La Cruz.

From the provided data, Adames has the highest launch angle and highest average launch angle, meaning pitches thrown to him tend to fall shorter in the field, many which could be pop-ups or easier outs. Lindor, on the other hand, has the widest spread in launch angles, which implies he hits the most diverse types of hits (from largest to smallest average launch angle: pop-ups, line drives, and fly balls). While Lindor is facing less diverse pitch types, as seen in Fig. 1, he is able to bat various types of hits, making it more difficult for fielders to determine the final location of a hit in the field.

Table 3 contains descriptive statistics for measures of centrality and variability for launch speeds (exit velocities) of in-play balls hit by Lindor, Adames, and De La Cruz.

	Lindor, Francisco	Adames, Willy	De La Cruz, Elly
count	360.000000	360.000000	360.000000
mean	90.481111	88.502778	91.125000
std	14.575174	14.715410	15.543216
min	31.600000	26.100000	9.400000
25%	83.800000	80.275000	82.650000
50%	94.450000	92.500000	93.100000
75%	100.900000	99.575000	102.250000
max	109.600000	112.300000	114.700000

Table 3. Descriptive statistics for measures of centrality and variability for launch speeds of in-play balls hit by Francisco Lindor, Willy Adames, and Elly De La Cruz.

While De La Cruz has the highest launch speed for any in-play ball for each player, Lindor has the highest minimum launch speed and highest median launch speed, despite some of the three batter's statistics appearing similar. However, in the case for De La Cruz's high bat speed, his original dataset

had over 100 less samples compared to either other batter, so the random selection of his at-bats is more representative of his full-season performance and could include outliers.

Table 4 contains descriptive statistics for measures of centrality and variability for the estimated batting average, determined using launch speed and launch angle, of in-play balls hit by Lindor, Adames, and De La Cruz.

	Lindor, Francisco	Adames, Willy	De La Cruz, Elly
count	360.000000	360.000000	360.000000
mean	0.348219	0.337683	0.361786
std	0.313597	0.311496	0.307201
min	0.001000	0.001000	0.001000
25%	0.056250	0.060000	0.077000
50%	0.251000	0.234000	0.275500
75%	0.611000	0.589500	0.577000
max	0.994000	0.989000	0.994000

Table 4. Descriptive statistics for measures of centrality and variability for the estimated batting average, determined using launch speed and launch angle, of in-play balls hit by Francisco Lindor, Willy Adames, and Elly De La Cruz.

Table 5 contains descriptive statistics for measures of centrality and variability for launch speeds (exit velocities) of in-play balls hit by Lindor, Adames, and De La Cruz.

	Lindor, Francisco	Adames, Willy	De La Cruz, Elly
count	360.000000	360.000000	360.000000
mean	0.632822	0.667969	0.642792
std	0.779781	0.869204	0.840980
min	0.001000	0.001000	0.001000
25%	0.076250	0.072500	0.102000
50%	0.340000	0.338000	0.335500
75%	0.905500	0.934500	0.863250
max	3.820000	3.917000	3.953000

Table 5. Descriptive statistics for measures of centrality and variability for the estimated batting average, determined using launch speed and launch angle, of in-play balls hit by Francisco Lindor, Willy Adames, and Elly De La Cruz.

4. Discussion

As mentioned in Footnote 5, the *estimated_ba_using_speedangle* and *estimated_slg_using_speedangle* columns are not completely representative of a player's batting average and slugging percentage.

MLB.com was consulted to determine the batter's true batting average and slugging percentage during the 2024 season. For the National League shortstop finalists, Lindor finished with a .273 batting average and a .500 slugging percentage, Adames finished with a .251 batting average and a .462 batting average, and De La Cruz finished with a .259 batting average and a .471 slugging percentage. As hypothesized before, the player's real batting average and slugging percentage do suggest these players fall under the "power hitter" archetype, as MLB's average slugging percentage is about .414.

The bottom line discovered from the research was that exit velocities and launch angles might not be that indicative of a player's performance. The readings found in the *estimated_ba_using_speedangle* and *estimated_slg_using_speedangle* columns were also inaccurate and negatively skewed. The average estimated batting averages were around .345, which is relatively high in the modern era of baseball, and the highest estimated slugging percentages were just under 4.000, highly impossible slugging percentage numbers for even the greatest power hitters in the history of baseball.

For future research, one could analyze another set of Silver Slugger finalists using the following links provided by MLB.com. Using the same dataset and methods explained in **Section 3. Methods**, future research could analyze and visualize the performance of any group of 3-5 batters selected as possible winners for their respective league and position. However, comparing batting behaviors for the newly-added "Team" category may be difficult, as this is a fan-voted category and further dataset manipulations would need to be performed, including filtering the dataset by all players of a team. (There are no features in the original dataset that indicate a given batter's team.)

- Award winners: https://www.mlb.com/awards/silver-slugger
- Finalists: https://www.mlb.com/news/silver-slugger-award-finalists-2024