You've Got to Know When to Hold 'em: Analyzing the Strategy of The Ultimate Gambler

Team 273

July 27, 2026

Abstract

When a runner is thrown out at home plate, they are often not the ones who can be blamed. In this data challenge, I have decided to take a deeper dive into the decision-making of the ultimate gambler: the third-base coach. My goal was to obtain a deeper understanding as to why a coach may hold a runner at third instead of sending one home. Utilizing player and ball tracking data, I created a model that reveals the precise thresholds that drive these high-stakes decisions. By analysing the marginal effects of the model's variables, I established that a runner's speed and the distance of the hit were the heaviest strategic levers in a coach's decision. More intriguing, though, as a third-base coach distances themselves from home plate, there is a considerable decrease in the probability of sending a runner home. Without focusing on the players themselves, I arrived at a powerful insight that has a large impact on winning a game.

Introduction

Background

In a stadium packed with spectators, there are plenty of people who are unfamiliar with the vast amount of rules and positions of baseball. When I read the prompt for this year's data challenge, my mind immediately focused on fans like those, ones like my mom and my sister, who were only there because of my love for the game.

If you look up at a stadium jumbotron, the screen is always filled with player metrics such as exit velocities, pitch velocities, and launch angles. Fans naturally learn to track the stars on the field. Yet, in the middle of a play at the plate, the most consequential figure is standing in a box in foul territory, all without a glove. We track every move a player makes, yet we rarely

give a window into the third-base coach's mind. By taking a deeper look into these high-stakes decisions, we can turn a confusing split-second play into an exciting, easy-to-understand moment of anticipation for every fan in the stands.

Context

A third-base coach may only have two options, but that doesn't make his ultimate decision any easier. With a wrong decision, the air can be taken out of the stadium in an instant. Just take a look at this Braves play from July 1 (Figure 1).



Figure 1: In this play, Dom Smith hits a single and Michael Harris II is sent home. Using the batted ball profile, the speed of the runner, and estimating the position of the coach, my model predicted about a 26% chance that the runner would be sent home.

Whether it was a heads-up call or due to defensive errors, a correct decision from a third-base coach will lead to them being praised for their confidence and guts. Take a look at this play from my dataset (Figure 2).

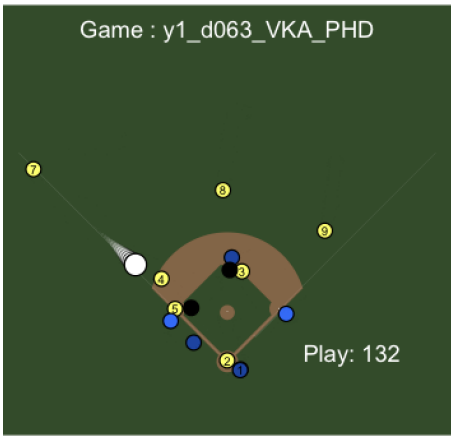


Figure 2: In this animation, the third-base coach sends the runners from 2nd and 1st to home. The ball is bobbled in the infield and both runs score. My model gave the runner on first a 51% chance to be sent home.

Data

SMT provided a year's worth of minor league ball tracking and player tracking data. Due to the anonymization of the data, game state info (runs, outs) was not included and was not necessary for this project. From the ball tracking data, I was able to pull out hits, exit velocity, and distances. Using the player tracking data, I could filter down to plays I needed with runners, the player's speed, and the coach position. More about the analysis techniques are in *Appendix I.*

Methods

A third-base coach's greatest effect is seen on plays where the ball is hit to the outfield and there is a runner on first or second base that has a chance to be sent home. To isolate plays like these from our large dataset, I filtered to plays where the ball was hit, immediately bounced, and after x amount bounces, was acquired by an outfielder. I then kept only the hits that occurred with a runner on first, second, or both. To find when at least one of the runners ran home, I used the Cartesian plane, filtering to plays where the runner was less than or equal to 5 ft away from home.

Integrating positional tracking for the third-base coach achieves my project's goal of shifting the analytical focus from the players to the coaches. To achieve my objective, I found the exact timestamp that the runner was closest to third base. I then pulled the tracking data for all third base coaches in the previously selected plays and found the coordinates of the coach as the runner was rounding third.

The last variables I needed for the batted ball portfolio were the hit distance and the exit velocity. The distance was found by grabbing the timestamp where the ball was acquired by one of the fielders and calculating the straight-line distance from home. I calculated the exit velocity using the first two frames after contact (100 ms). With the x,y, and z coordinates, I found the change in position over the change in time. More about the techniques are in *Appendix I.2*.

Now in the closing stages of my data wrangling, my last step was to find the runner's speed. To understand a player's speed through a third-base coach's lens, we need to find the max sprint speed for each player; not just the speed they were running at during any exact play. I pulled out a list of every runner we needed to record the speeds for and found every play where they were on base in the dataset. After recording the coordinates from each play, I computed the speed of the runner and took the player's 98th percentile speed. More about the speed technique is in *Appendix I.2*.

Analysis

The first step of analysis for me was to take a look at the raw dataset itself and find the average value of each variable, isolated by if the runner was sent home or not (Figure 3).

Sent vs Held: Average Values

sent_home	hit_dist	exit_velo	sprint_speed	coach_dist	x_pos	y_pos
0.00	269.29	86.79	24.91	81.39	-64.75	49.08
1.00	288.31	90.38	25.57	76.99	-63.67	42.95

Figure 3: Sent vs Held Table

Clearly, there is a large discrepancy between the statistics when a runner is sent versus held. The most important variable to note is the coach distance from home, where there is almost a 4.5 ft difference from when the coach sends a runner versus holding one, an important difference that often goes unnoticed.

Before running the model, I decided to take a look at a graph of the runner's speed against the hit distance. In Figure 4, I can take away a self-explanatory result: slower runners can score on farther hit balls while faster runners are able to score on balls that are hit shallower. When connecting this graph to Figure 3, the runners sent home are in fact faster and are scoring on deeper hits than the ones held up.

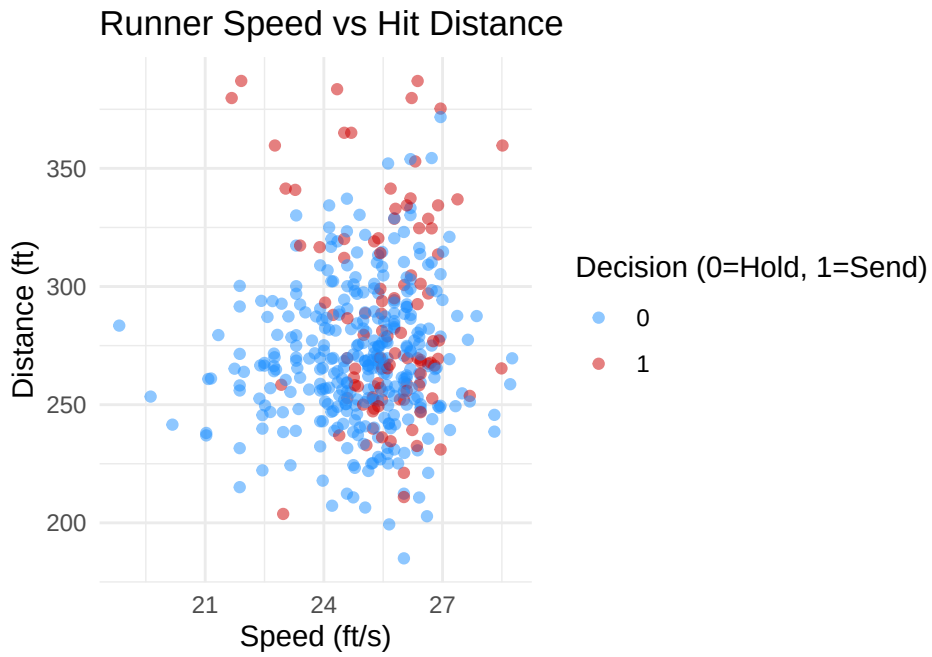


Figure 4: Speed vs Distance Graph

Model

To predict the probability of a runner being sent home on any given play, I ran a logistic regression using all of the aforementioned variables. There were 433 observations in the dataset, 334 runners held and 99 sent, and the AIC was 430.97. The model helped me decipher the effect of what a one unit increase on each variable would do to the probability of sending a runner home. The formula of the model is as shown:

$$\begin{aligned}\text{Runner Sent Home} = \beta_0 &+ \beta_1(\text{inning}) + \beta_2(\text{exit velo}) \\ &+ \beta_3(\text{hit distance}) + \beta_4(\text{sprint speed}) + \beta_4(\text{coach distance from home})\end{aligned}$$

More about the model, including the model summary, can be found in *Appendix II*. In a logistic regression model, the interpretation of the estimates can be the hardest part to pin down. In order to grasp a true understanding of each variable's effect, I transferred the log odds into direct probabilities (Figure 5).

Logistic Model Term Effects

Interpretation Helper

Variables	Probability	p-value
coach_distance_from_home	-0.35%	0.02
exit_velo_mph	0.1%	0.26
hit_distance	0.24%	0.00
inning	-0.77%	0.31
max_sprint_speed	5.32%	0.00

For every one unit increase in ***variable****, the probability of the coach sending the runner home changes by* ***probability****.*

Figure 5: Model Term Effects Table

The sprint speed of the runner looms large in the mind of the third-base coach as for every 1 ft/s increase, the probability of sending the runner increases by 5.32% (average marginal effect). Another detail to take away is that the inning variable was not statistically significant, contrary to my original belief. Late in games, a third-base coach may become aggressive, but they may also become extremely conservative depending on the score (a factor we must omit).

For every 10 ft increase in the coaches distance away from home, they are 3.5% less likely to send a runner, granting us a quantitative value for our variable of focus. Figure 6 shows the third-base coach position variable throughout the whole dataset, providing a clear visual picture of what their position would look like on the diamond. There is a distinct cluster of points where runners were sent home instead of being held up. As the plot moves to the left of that cluster and up the third-base line, the red points start to fade out, contrary to as when we move down the line, some red dots are still appearing.

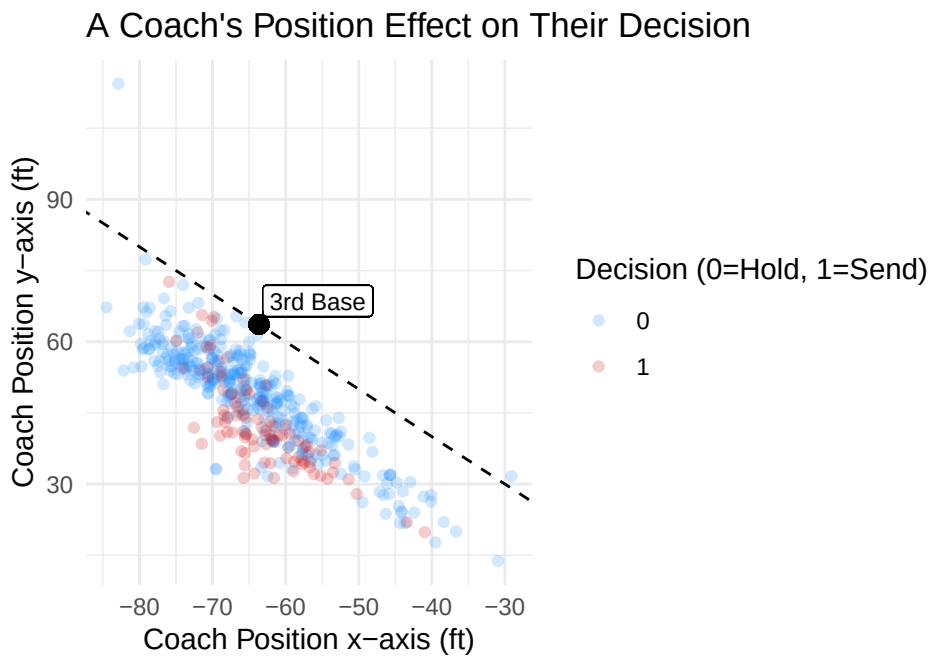


Figure 6: Coach Position Graph (The back tip of home plate is at (0,0))

Limitations

From my results, we can draw the conclusion that a third-base coach has a very tough decision to make, but there are certain variables they rely on to help them make that split-second judgment, most importantly where they position themselves and the speed of the runner.

There were a couple key variables that were intentionally left out from my model that could have had a huge effect on the decision to send a runner home. As previously mentioned, the anonymization of the data made it so we could not decipher the score or outs of the games, two components that most definitely alter the coach's decision in a large way. A coach's aggressiveness is completely erased from the project, therefore we can only apply the findings to the extent of variables covered.

In addition to the anonymization limitation, I decided to exclude outfielder arm strength. Due to the time restraints and the immense amount of data wrangling that goes into finding each outfielder's arm strength, I decided to bolster the rest of the project instead. Not every play is a clean, outfielder to catcher laser. Including raw arm strength without accounting for the infielder's relay speed and accuracy would create noise in the model. A high exit velocity from the outfielder's hand means very little if the cutoff man bobbles the transfer or makes a poor throw home. Future iterations of this model could incorporate an outfield velocity variable that, I predict, would be significant at the 0.05 level.

It is also worth noting that the coach position finding may run both ways. A coach might position himself closer to home because he has already decided to send the runner and wants to wave him in, meaning the position partly reflects the decision rather than only driving it. My model treats position as a predictor, so reverse causality is something to keep in mind when interpreting that result.

Beyond defensive metrics and game state, additional physical and contextual constraints exist within the data. First, the model does not account for runners who are stealing or who have abnormally large leads. Second, the data does not account for stadium dimensions or weather conditions, both of which can help make a send or a hold.

Conclusion

Ultimately, I wanted to create a project that reels in casual fans who might not otherwise appreciate the art of the third-base coach, while offering an analytical perspective on one of baseball's most frequent plays. On the field, coaches can learn from these split-second decisions, utilizing the live 'Send Probability' displayed on the jumbotron to sharpen their own judgment in future games (detailed in *Appendix III)*. Plays at the plate are easily the most thrilling moments in baseball. By giving third-base coaches a window into how their own intuition operates under pressure, we can encourage bolder decisions on the diamond, sending more runners home, and elevating the excitement of the game for everyone in the stands.

Appendix

I: Code Descriptions

1 Datasets and Variables Used

ball-events:game_string, play_per_game, timestamp, player_id, ball_event_code

ball-positions:game_string, play_per_game, timestamp, ball_position_x, ball_position_y, ball_position_z

Lineups.csv: game_string, inning, play_per_game, batter, on_1b, on_2b, on_3b

Player-positions: game_string, play_per_game, timestamp, player_id, field_x, field_y

2 Algorithm Overview

Hit definition:

Code: filter((player_id == batter & ball_eventcode == ball hit) &

(lead(player_id) == ball & lead(ball_eventcode) == bounces))

The ball is hit and bounces right after.

Code: filter(((player_id == left-fielder & ball_eventcode == ball acquired) |

(player_id == center-fielder & ball_eventcode == ball acquired) |

(player_id == right-fielder & ball_eventcode == ball acquired)) &

(lag(player_id) == ball & lag(ball_eventcode == bounced)))

The ball is picked up by an outfielder directly after bouncing.

Exit Velo:

Code: mutate(dist = sqrt(change_x^2 + change_y^2 + change_z^2),

exit_velo_ft_ms = dist / change_time,

exit_velo_mph = exit_velo_ft_ms / 5280 * (1000*60*60))

Calculate the distance changed between the two frames (1 frame = 50 ms), use the distance formula, divide it by the change in time and convert to mph.

Bases:

Code: mutate(distance_to_third = sqrt((field_y - 45 * sqrt(2)) ^ 2 + (field_x + 45 * sqrt(2)) ^ 2))

A cartesian plane on a baseball field with the back tip of home plate at (0,0). Third-base is at (-45√2, 45√2), so this will find the runners distance to third.

Code: mutate(coach_distance_from_home = sqrt(as.numeric(coach_x)^2 + as.numeric(coach_y)^2))

Finding the third-base coach's distance from home plate using the (0,0) point.

Time Difference Variable:

Code: mutate(time_difference = abs(as.numeric(timestamp) - as.numeric(third_timestamp)))

time_difference: When the frame of the coach's current location was taken minus the timestamp the runner touched third. Used to find where the coach was standing as they were making their decision. Needed due to the player_positions data not having every frame of every play.

Sprint Speed Percentile:

Code: summarize(max_sprint_speed = quantile(speed_ft_s, 0.98, na.rm = TRUE))

Due to the possibility of data issues, I decided to take the 98th percentile of the player's max sprint speed.

II: Model

Table 1: Coach Model Coefficients

	Coach Model
(Intercept)	−12.577***
	(2.740)
inning	−0.049
	(0.048)
exit_velo_mph	0.006
	(0.006)
hit_distance	0.015***
	(0.004)
max_sprint_speed	0.339***
	(0.095)
as.numeric(coach_distance_from_home)	−0.022*
	(0.009)

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

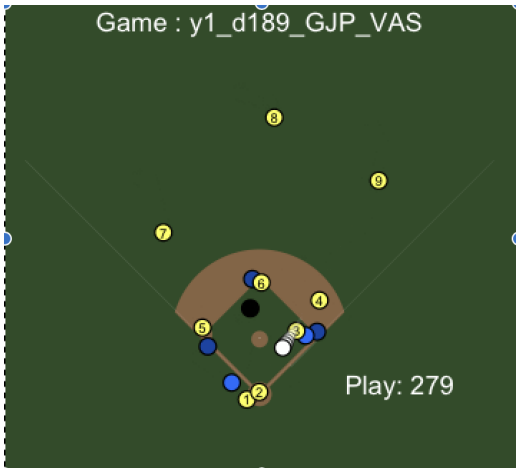
Code: coach_model <- glm(data = master_coach, sent_home ~ inning + exit_velo_mph + hit_distance + max_sprint_speed + as.numeric(coach_distance_from_home), family = binomial)

The model summary table shows the coefficients of each of the variables in the model. The numbers are log odds.

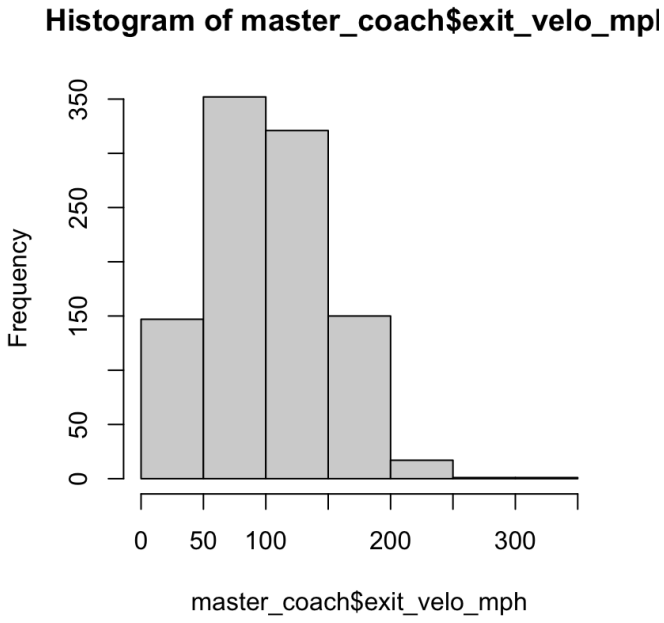
III: Other

1 Other Figures

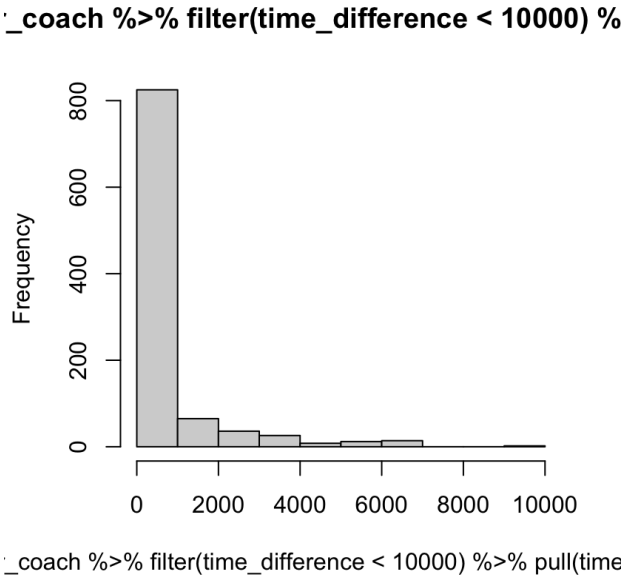
Link



An extra animation depicting a third-base coach hold up their runner at third base



A histogram of the exit velocities in the final dataset, showing the reason we limited the range of exit velocities.



A histogram of the time differences against their frequency in the final dataset. We eliminated any that were over 1000 milliseconds.

Live Tracking

Decision Analytics Display

Stat	Value
Live Runner Speed (Max)	17.4 MPH
Batted Ball Distance	315 Ft
Coach Setup Distance	78 Ft from Home
LIVE SEND PROBABILITY	**51%**

2 Scoreboard Bug

The scoreboard bug is meant to flash up on the board right after a play has concluded, specifically one where the third-base coach had to assist in the runner's decision. The "Live send probability" will flash and then fans, players, and coaches can examine what the odds were of the runner being sent and compare to the result of the play.

Acknowledgements

I'd like to acknowledge Billy Fryer, Meredith Wills, Cameron Adams and the rest of the SMT team for their help throughout the summer. My project would not be the same without the tutorials and the coding help I received. I truly appreciate all their hard work.

References

Newest Questions - Stack Overflow, stackoverflow.com/questions. Accessed 25 July 2026.

Major League Baseball Video Search | MLB Film Room | MLB.Com, www.mlb.com/video. Accessed 25 July 2026.