

SMT Data Challenge

Jack Rogers, Jackson Balch, Isaac Blumhoefer, Kai Franke

University of Minnesota Twin Cities

## **Applications of Tracking Data in the Evaluation of Baserunner Performance**

Over the next few years, the advancements in baseball analytics will transform the game we know and love. With new technology like Hawk-Eye and the advancement of Stat-Cast, teams will have more data than ever before. Hawk-Eye opens up the vast new world of player tracking, which has the potential to revolutionize the way we analyze defense and baserunning. This new data could even be useful in determining optimal batter-pitcher matchups.

Stat-Cast has done this in the past; however, Hawk-Eye can help teams by having more accurate and granular data. In our project, we created new metrics to help evaluate both outfielders and baserunners when trying to score on a tag up or while rounding second base. We developed two primary metrics that find the momentum of the runner rounding the base and determine what a “close play at the plate” is. In addition, we had a running value representing the fielder’s distance from the ball, the fielder’s distance from the plate, and the baserunner’s distance from the next base. These metrics allow us to measure the performance of outfielders, baserunners, and even base coaches. We then applied these metrics to the situations of runners being sent home. In what situations do baserunners go from second to home? Should they be sent more? We hypothesized that runners are not being sent as often as they should, resulting in runs being left on the board.

Before we got into any analysis, however, we needed to clean and manipulate this massive dataset.

## **Data Cleaning and Manipulation**

The data we were given for this project was very raw. It had over 20,000,000 observations, so we really had to sift through the data to find anything meaningful. We attempted to add exit velocity and throwing velocity to our metrics, but found that the timestamps were not accurate enough to do so. We found outlier data points of exit velocities of over 240 mph! We had to abandon these metrics and all metrics that would utilize time in their calculations. Without reliable timestamps, we found that any metrics self-calculated using them would also be unreliable.

We decided that we needed to get the location of the catcher, the fielding of the baseball by the outfielder, and the end location of the throw and put them in the same observation so we could use those initial values to find relationships with the end values. We did this by filtering out to when the ball was thrown and when the ball was caught near the plate. This was used by a lead function in both R and Python. Through this, we were able to create ball start and end location variables. In order to make the data easier to work with, we made sure to filter out much of the unneeded information, mainly the locations of all non-involved players and player IDs, among many others.

## **New Variables**

Next, we decided how we wanted to define momentum and whether or not a play was close or not. Before that, though, we also added each of the fielders involved in the play and base runner locations for every play on single observations. This way, we could compare where each of these players were at all times and could see if the catcher with a ball was close to the base runner at the plate.

The definition of a close play at the plate, in our eyes, was a play where the catcher with the ball and runner were within 12 feet of each other. This may seem like a large number but we needed to add a cushion due to the common errors in the data we had already found. Out of 156 plays at the plate where the runner rounded third, we found 10 plays that were “close”.

That may not seem like much, but the dataset was relatively small compared to the length of an mlb season. With more data this same process can be applied to see better results. In addition, this small dataset confirmed some of our prior hypotheses: runners are not sent enough from second to home.

For momentum, we used data from just before the fielder throws the ball, from how much they move when they acquire the ball to when they throw it. Having more momentum going into the player's throw increases the velocity, which in turn makes it more likely that they will throw a player out at the plate, making it a close play, or having the runner hold up.

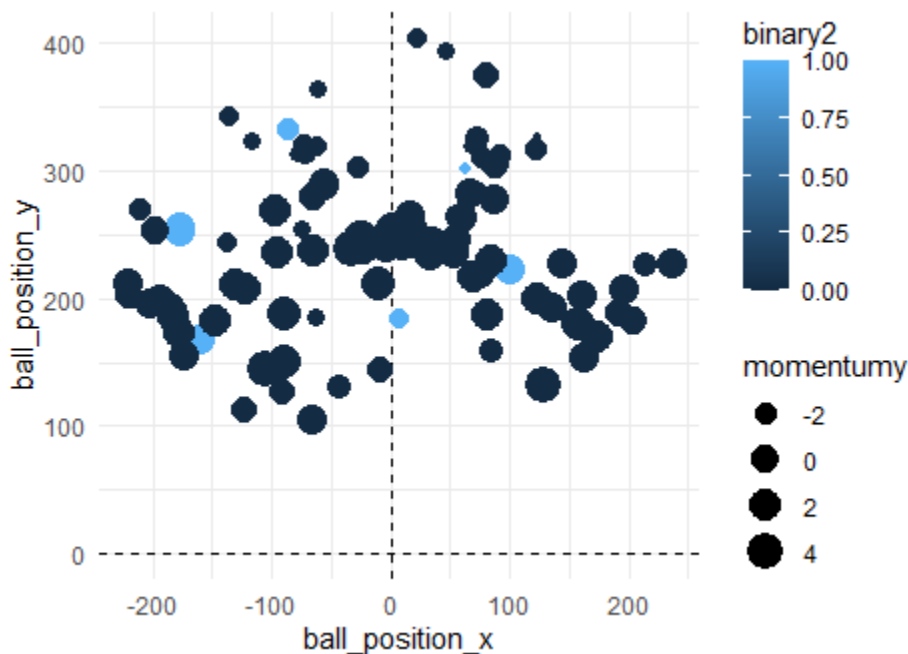
### **Significant Findings**

The next step after creating new variables is to find their applications. In the case of the “close plays” statistic, the most noteworthy finding was how few close plays there are within games. With runners on third and a pop fly to the outfield or runners on second and a ball hit to the outfield, only ~6% of the given plays resulted in a contested play at the plate. In other words, runners either easily crossed home plate or were held at third base in 94% of the cases, a shockingly high statistic. What this entails is that teams, or rather third base coaches, may be over-reluctant to send runners home, potentially to a degree that is costing batting teams multiple runs. What makes this even more surprising is the fact that batting averages have been decreasing across all baseball levels, meaning that trusting the next batter to drive the runner on second or third home is a much greater risk. Coaches are actively leaving runs on base, potentially costing their teams.

Another takeaway from the lack of close plays in this given selection is that some of the most exciting moments in baseball - the plays at the plate - are occurring a lot less frequently than they probably should, ultimately taking away from the overall product on the field. For if there are only 10 close plays at the plate in this entire dataset, that means that the on-field product is not as exciting as it should be for viewers of the game. As we are often seeing in the modern game, fans are the ones suffering the most.

We applied the momentum statistic and other fielding metrics we created to attempt to evaluate the performance of third base coaches. By comparing these metrics to make predictions on when a runner should score, we can go back and check for runners that were held at third in different situations. Due to the limited size of the dataset, we were not able to apply machine learning techniques, but we set a baseline of a few filters.

First, runners are almost never thrown out at the plate when a fielder's momentum tracker was taking them backwards. This is a hard and fast rule that applies to almost all of the data points (the only outliers being extremely short fly balls). In addition, the more a fielder is moving in when he makes the catch, the more likely he is to be thrown out at home. This might seem like common sense, but it is clear that third base coaches are not acting upon this as much as they should. We were also able to determine that the momentum of the fielder tends to be more predictive of the outcome at the plate than the distance of the throw. A visualization of these close plays is included below. The x and y axis represent the position of the ball on the field. The light blue dots are the plays at the plate and the size of the dot represents the forward momentum.



### Further Application

These metrics can be applied to many baserunning scenarios. With more data at an improved granularity, we can predict at any point in a play where the baserunners should end up. This would allow us to evaluate a baserunner on an individual level. How many extra bases over expected does a runner earn every year? And how many expected runs does that generate for the team? At the moment, there is not a great metric for evaluating baserunner performance, and this tracking data provides the potential for that. Similar to the case study above, this data can also be used to evaluate coaching. Which teams are improving their overall run productivity through the bases? How does a runner's performance change if they switch teams? With a larger sample size, the possibilities are endless.

## **Conclusion**

Our knowledge of the game of baseball grows exponentially every year, especially as new data becomes available. Most recently, player and ball tracking information are the latest revelations that are pushing us into a new frontier, one that allows us to dig even deeper into the nuances of the game. One such discovery from this newfound data is the significance of outfielder momentum on the number of close plays there are at home plate. What we found through a long process of data cleaning and statistical creation is that momentum when catching a flyball is absolutely critical to not only the runner safely crossing home plate, but also the decision by the third base coach to even send the runner in the first place. More specifically, that coaches and teams are not utilizing the knowledge of an outfielder's momentum enough in their decision making. Through our creation of these statistics, however, we strive to successfully bring to light this unknown advantage in hopes teams/coaches will make more informed decisions in the future. Simply reading an outfielder's position may be the difference between a run, a win, or even an entire playoff series. We believe that through deeper analysis of momentum and overall future advancements in player tracking data, however, those lost runs can eventually be scored.