It's Not My Fault: a Balanced Perspective on Stolen Bases

Andy Hutchison, Max Greaves, Ryan Dajani, Shamar Phillips

Friday August 9, 2024

Abstract
In baseball, stolen bases present a unique situation where a player can be charged with an error or a caught stealing without ever touching or throwing the ball. Currently, there are few metrics available to measure the impact of a caught stealing on the pitcher. While a catcher's performance can be evaluated using caught stealing above average, which compares a catcher's effectiveness against the league average, similar metrics for pitchers are lacking. Utilizing player and ball tracking data, we analyzed pitch time, pop time, and ball location in the strike zone to determine how pitchers and catchers influence a steal. This analysis enabled us to develop a caught stealing above average metric for pitchers, providing a new way to assess their impact on the play. Contrary to common belief, our findings suggest that catchers may often be blamed for stolen bases that are not entirely their fault. Our research indicates that pitchers' roles in controlling the running game are under-analyzed and should be given more attention when defending against stolen bases. Further data and analysis are necessary to fully understand both the catcher's and pitcher's impact on the play and to refine the metrics used to evaluate stolen bases.

Introduction

At the end of September, the Atlanta Braves were in first place in the NL West, and clinched their division and a position in the 2023 Playoffs. Ronald Acuna Jr. was fighting for something else that day. A player has hit 40 home runs in a season 364 times, and 132 players have stolen 70 bases in a season (MLB Stats, Scores, History, & Records | Baseball-Reference.com, n.d.). As Acuna Jr. took off for second base, he was chasing something that no player in the history of professional baseball had: the 40/70 season.

Baseball, like a symphony, requires perfect harmony among all its musicians. The art of preventing stolen bases is a team effort, involving much more than the catcher's reflexes and strong arm. By focusing solely on the catcher in stolen base attempts, we may miss the music that our sport's musicians create: a complex interplay of timing, strategy, and coordination that determines success or failure in the split second of a steal.

In 2023 Major League Baseball (MLB) debuted a set of new rule changes to decrease the overall time of games. Some key changes are limited pickoff attempts allowed, wider bases, and the introduction of pitch clock. MLB accomplished their goal as game times dropped from 3 hours and 6 minutes in 2022 to 2 hours and 42 minutes in 2023 (2023 MLB Rule Changes | MLB.com, n.d.). Another consequence of these rule changes was a drastic increase in stolen bases. The league leader –Ronald Acuña Jr.– had 73 stolen bases, the highest total since 2007, and the total stolen bases was 3500 compared to the 2487 in 2022 (Yearly League Leaders & Records for Stolen Bases | Baseball-Reference.com, n.d.) (Baseball Almanac, Inc., n.d.). With this influx of stolen bases a question arises: does the defensive blame lie in the hands of the pitcher or glove of the catcher?

Data

The data used in this analysis came from over 330 Minor League Baseball (MiLB) games across two seasons and four levels of the minor league system. The data was compared using game events information coded for each play. Information was provided with respect to the 2D (x,y) coordinates for player positions, 3D (x,y,z) ball position, with timestamps every 50 milliseconds distinguishing differences in location over time.

Each type of ball movement was uniquely numbered (eg: 1 for pitch, 2 for catch), and included in the data to understand the progression of the play over time. Each ball movement had a timestamp every 50 milliseconds allowing us to analyze stolen base plays. A filtering process was done to filter down to 236 stolen base attempts exclusively to second and third base. These attempts are what the analysis will be based upon.

Methodology

The analysis aims to understand the probability a player (and by that, catcher and pitcher) has of being caught stealing. An understanding of pop time will be used to better understand when a catcher should have a chance to throw a runner out. Pop time is the time between when a catcher catches a pitch and when their throw is caught at the base. Further, an understanding of where a pitch was caught by the catcher will help to understand why a catcher might struggle to throw out a runner attempting to steal.

When looking at stolen bases extracted from the data that has been provided by SMT, there are a few things we need to do to ensure that a player has attempted to steal a base, and they are as follows:

1. A player's ID has been allocated to a starting base
2. A throw has been attempted to the advancing base
3. The advancing base is unoccupied, meaning that no baserunner is at the advancing base

For a stolen base attempt to count as successful, these must happen:

1. A player's ID has been allocated to two consecutive bases within back-to-back plays
2. A throw has been attempted to the advancing base

A player has been caught stealing if:

1. Within back-to-back plays, a player's ID has been shown only on their starting base and the ID does not appear on the advancing base
2. A throw has been made to the advancing base

Figure 1: The player sliding into second base has their center of mass much further behind their hand that touches the base first. This is the main reasoning behind our buffer zone around the bases.

We create a zone stretching from the base to ten feet before the base, as well as four feet wide. We created 4 feet of width because the runner might not run head on to the base, and maybe slide around the base to avoid the infielder's tag.

We compared two timings: the time a baserunner first enters the zone, and the time at which the fielder acquires the ball. We concluded that a runner was expected to be caught stealing if the fielder acquired the ball prior to the runner entering the zone. We expected a stolen base if the runner entered the zone prior to the ball entering the zone. The runner entering the zone is a filtering step, as there are situations where a throw can occur to a base with the runner assigned to it, while there is no "steal attempt".

We defined Expected Caught Stealing and Expected Caught Stealing percentage value (xCS and xCS%) to quantify the value of pitchers and catcher stolen bases against the expected marks. Based on the criteria defined for a stolen base, for xCS we gave the values of a stolen base 1, and a caught stealing 0. We then calculated the expected stolen base numbers. xCS is taken by adding up a catcher or pitcher's number of baserunners caught stealing, and xCS% is found by dividing xCS by the number of attempts. We can compare the data between CS and xCS for each pitcher and catcher, and subtract the values to find a new value, Caught Stealing Above Expected (CSAx). This number will show how good a catcher or pitcher is at preventing steals, a positive number being good, a negative number being bad, and 0 implies the player has the expected number of steals caught.

Our goal is to understand the pitcher's impact on the stolen base. The use of pitch time, the time it takes to get the ball from the hand of the pitcher to the glove of the catcher, is one point of analysis. If pitch time is higher, the catcher has less time to get the throw to the base to catch the runner stealing. When a catcher is set up behind the plate, square to the pitcher, we see that their throwing hand is on the first base side of the plate. We wanted to investigate whether pitch location is another important factor in pop time.

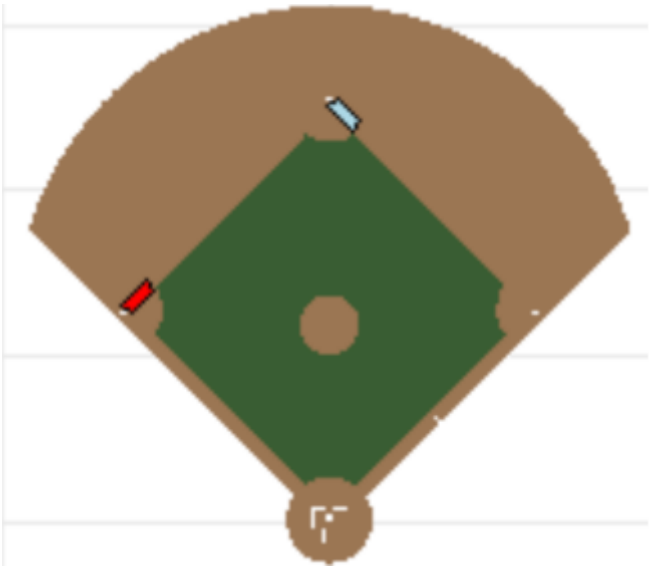

Figure 2: Buffer zone

Pop time is one of the most widely-used statistics to measure a catcher's defensive effectiveness. A faster pop time is often correlated with a higher likelihood of CS% for catchers. The league average caught stealing rate over the last ten years is about 30%, with an average pop time of 1.98 seconds. As the pop time of catchers decreases, we will see that the caught stealing rate increases considerably.

Discussion

We analyzed the pitches for each stolen base attempt to be able to draw conclusions and paint general trends that are able to help explore the preventative defense of pitchers. A well-placed pitch to defend against a steal is one that is on the catcher's right side by their right hand, low and away for a right handed hitter. Receiving the pitch by their throwing hand allows catchers to minimize transfer time resulting in a faster throw. Figure 3 provides a visual on this idea contrary to what we expected to see, highlighting that pitches thrown on the catcher left and high in the zone have a better caught stealing percentage. An explanation for this subversion of our expectations could lie in the idea that catchers are able to generate more accurate and powerful throws to the base when throwing from their feet compared to on their knees. Our graph may show inflated numbers on higher areas in the zone due to lack of data points from the fact runners ideally attempt stolen bases on pitches they know are breaking balls, and breaking balls generally break into the lower portion of the zone. Pitchouts are another factor that must be spoken about when it comes to caught steals. According to Sam Miller of Baseball Prospectus, it was found that pitchouts can cost a defense around 18 runs per season, which isn't significant, but with the rise of the running game we might see a rise in the pitchout again.

Pop time and pitch time are crucial metrics in regards to defending stolen bases. A faster pop time and a faster pitch time give the runner less time to make it to the next base, meaning we expect to see higher caught stealing rates with lower times for both.

Figure 4 shows the trends from our data on pitch time, pop time, and the associated caught stealing percentages. It is important to note there are empty rows since the majority of the data is taken every 50 milliseconds and due to an absence of data for more extreme pitch times. Our data confirmed our initial thoughts that lower pitch time and pop time result in a higher caught stealing rate.

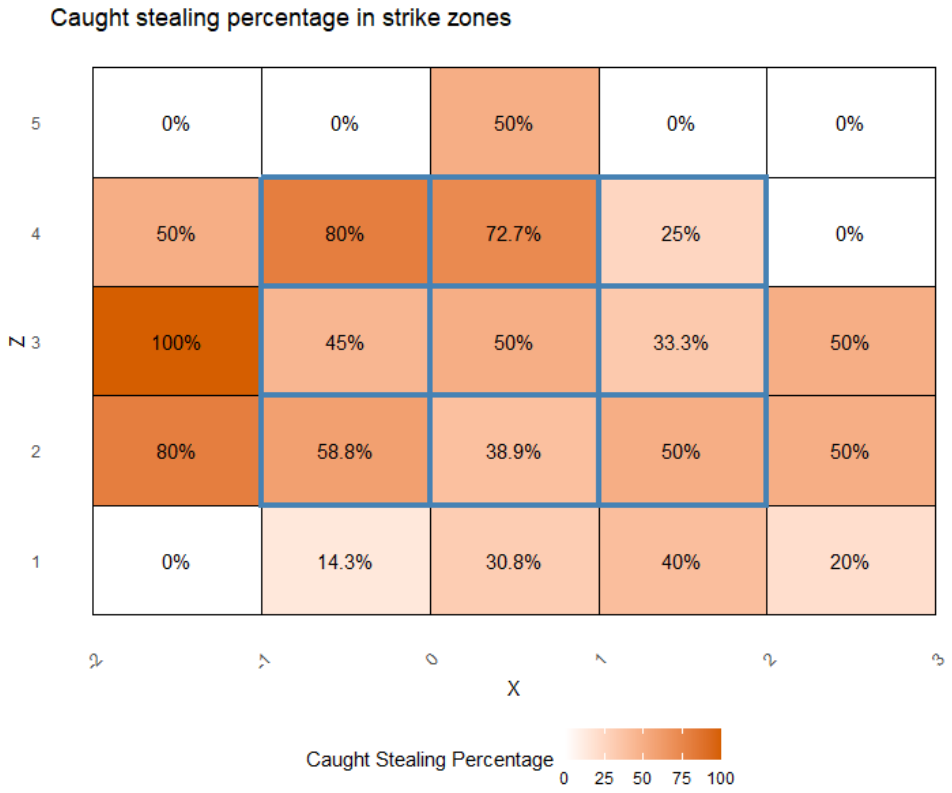

Figure 3: Caught stealing percentages by pitch placement from catcher's perspective

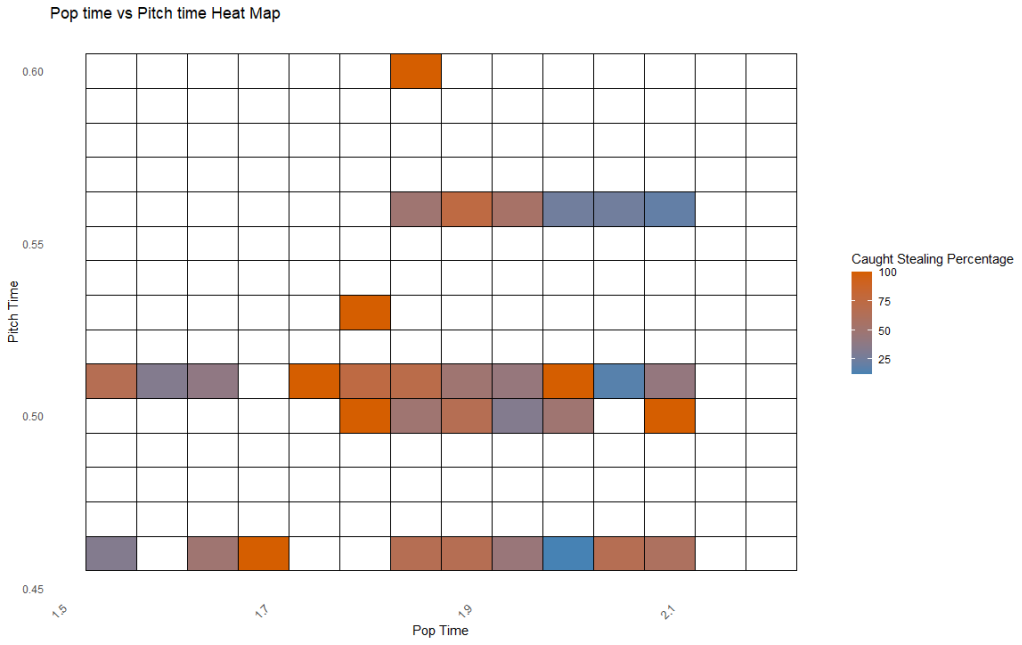

Figure 4: Heat map showing the relationship between pitch time and pop time with caught stealing percentage

Top Ten Pitchers In Caught Stealing Above Average					
Steals Attempted and Allowed					
Pitcher ID	Total Steals Allowed	Total Attempts	Steals Allowed With Average Catcher	Average Pitch Time	Caught Stealing Above Average
6685	0	3	3	0.50	3
523	1	5	3	0.51	2
549	1	3	3	0.53	2
602	0	2	2	0.50	2
674	1	4	3	0.49	2
722	3	5	5	0.50	2
733	0	2	2	0.50	2
811	3	5	5	0.47	2
814	1	3	3	0.50	2
966	1	6	3	0.50	2
SMT Data					

Figure 5: Top 10 pitchers in CSAx

Top Ten Catchers In Caught Stealing Above Average					
Steals Attempted and Allowed					
Catcher ID	Total Steals Allowed	Total Attempts	Steals Allowed With Average Pitcher	Average Pop Time	Caught Stealing Above Average
598	6	16	13	1.88	7
920	3	13	10	1.85	7
953	1	9	6	1.94	5
786	11	17	15	1.84	4
520	6	10	9	1.94	3
530	2	5	5	1.73	3
2323	0	4	3	1.98	3
8815	0	3	3	1.87	3
737	1	3	3	1.95	2
2252	0	2	2	2.00	2
SMT Data					

Figure 6: Top 10 catchers in CSAx

Bottom Ten Pitchers In Caught Stealing Above Average

Steals Attempted and Allowed

Pitcher ID	Total Steals Allowed	Total Attempts	Steals Allowed With Average Catcher	Average Pitch Time	Caught Stealing Above Average
3268	1	1	0	0.50	−1
6451	1	1	0	0.45	−1
6499	1	3	0	0.50	−1
8549	1	1	0	0.50	−1
367	2	2	2	0.45	0
380	2	2	2	0.45	0
456	2	2	2	0.50	0
473	3	3	3	0.53	0
511	0	1	0	0.60	0
551	1	1	1	0.55	0

SMT Data

Figure 7: Bottom 10 pitchers in CSAx

Bottom Ten Catchers In Caught Stealing Above Average

Steals Attempted and Allowed

Catcher ID	Total Steals Allowed	Total Attempts	Steals Allowed With Average Pitcher	Average Pop Time	Caught Stealing Above Average
2771	1	1	0	1.80	−1
3042	2	2	1	2.08	−1
3305	1	1	0	1.95	−1
5205	1	3	0	2.07	−1
5444	1	1	0	1.55	−1
8771	1	1	0	1.95	−1
348	3	4	3	1.88	0
418	2	2	2	1.82	0
647	1	1	1	1.95	0
719	2	2	2	1.75	0

SMT Data

Figure 8: Bottom 10 catchers in CSAx

Evaluating CSAx gives us an idea of how much pitchers and catchers individually impact the outcome of a stolen base attempt. Figures 5-8 give the extremes of a player's impact on stolen bases. The catchers with the faster average pop times were the ones at the top of CSAx list, this cements the idea that pop time is critical when defending a steal attempt. Pitchers don't play nearly as frequently as catchers, so data is limited with them. So far in the 2024 MLB season, there have been 116 games, and the top catcher (Salvador Perez) has caught 113 games. The top pitcher (Seth Lugo) has 155 innings pitched, which is just over 17 total games (2024 MLB Player Hitting Stat Leaders, n.d.). The floor and ceiling for catchers varies much more than pitchers for this reason, they're simply more involved in the game.

Conclusion

This paper has outlined the traditionally overlooked importance of a pitcher in defending against stolen bases. Conventional statistics such as pop time, pitch time, and pitch location were evaluated to gain an understanding of what factors of stolen bases are. xCS and CSAx were introduced to help emphasize that a pitcher and catcher both must be quick to prevent a stolen base. Using the ideas and statistics introduced in this paper, teams could be better able to assign blame from a steal to a player and develop their skills to prepare for the future. To conclude, pitchers and catchers must both bear the responsibility of defending stolen bases.

Future Work

While the analysis takes an encouraging look at player speed, pitch time, and pitch velocity there is still room for improvement with additional data. For example, the time it takes for a pitcher to release a pitch, the windup of a pitcher, is crucial information in determining critical steps for defense against stealing. Seeking new data could increase the accuracy of our measurements and validity of our conclusions. Many of our calculations use a buffer zone to calculate when players arrive at the base. We use a buffer zone because there is no data that indicates contact with the base since we are given only the center of mass of the player. The buffer zone gives us an educated guess on when a runner gets to a base, meaning expected caught stealing values we calculated differ from true values. Using data that includes the outcomes of plays will reduce the ambiguity on some of the data we generated surrounding close plays at a base. In addition to reducing the ambiguity of close plays, outcome-specific data better enables us to characterize stolen bases compared with using play sequences.

The greatest thing we can do moving forward with this project is increasing the amount of data. Having more data allows us to increase the accuracy of our existing data, as well as give us the opportunity to branch off into new ideas focusing on other aspects around defending stolen bases.

Acknowledgments

This paper would not have been possible without the contributions of our brilliant collaborators Dr. Meredith Wells, David Awosoga, Dr. Jack Davis, and UWAGGS.

References

2023 MLB Rule Changes. (n.d.). MLB.com. https://www.mlb.com/rule-changes-2023

Baseball Reference. (n.d.). MLB Stats, Scores, History, & Records | Baseball-Reference.com. Baseball-Reference.com. https://www.baseball-reference.com/

Baseball Reference. (2024). Yearly League Leaders & Records for Stolen Bases | Baseball-Reference.com. Baseball-Reference.com. https://www.baseball-reference.com/leaders/SB_leagues.shtml

Major League Miscellaneous Year-by-Year Averages and Totals. (n.d.). Baseball-Reference.com. https://www.baseball-reference.com/leagues/majors/misc.shtml

Miller, S. (2013, January 14). Pebble Hunting: The Probably Pointless Pitchout | Baseball Prospectus. Baseball Prospectus. https://www.baseballprospectus.com/news/article/19373/pebble-hunting-the-probably-pointless-pitchout/

Miller, S. (2024, July 30). Elly De La Cruz Jumps But They Get Increasingly Preposterous. Substack.com; Pebble Hunting. https://pebblehunting.substack.com/p/elly-de-la-cruz-jumps-but-they-get

MLB Stats | Baseball Stats. (n.d.). MLB.com. https://www.mlb.com/stats/

MLB Stolen Base League Totals | Baseball Almanac. (n.d.). Www.baseball-Almanac.com. https://www.baseball-almanac.com/hitting/hisb3.shtml

Statcast Catcher Throwing Leaderboard. (n.d.). Baseballsavant.com. https://baseballsavant.mlb.com/leaderboard/catcher-throwing