MLB Pitcher Strategy in Stolen Base Opportunities

Jordan Wallach, Zach Esrig, Drew Mathieson

I. Introduction

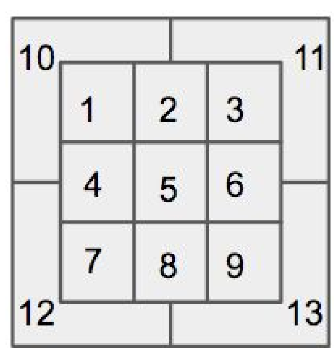
In baseball, it is easy to see how big of an impact having a fast player like Billy Hamilton on base can have. The game slows down, the pitcher changes his timing, pickoff attempts are made, and the crowd becomes restless. A less obvious effect of having runners on base is that a pitcher can change his strategy with regards to the batter – throwing balls that increase the likelihood of the catcher being able to throw the runner out on a potential stolen base attempt.

In this paper, we were interested in examining how pitchers respond in situations with a runner on first base. While formulating our research questions, we made a hypothesis about the propensity of pitchers to change their strategy, namely, that pitchers would select faster pitches with a runner on first, as it would intuitively minimize the amount of time between the split of the throw and the moment the catcher could pass the ball to the attempted stolen base. To either corroborate or disprove this hypothesis, we were first interested in asking how a pitcher's strategy changes when a runner is on first. Second, given a change in strategy, we were interested in whether this new strategy proved beneficial to the pitching team. Finally, with an optimal pitching strategy, our research aimed to make recommendations to pitchers on how to adjust their current strategy to maximize their team's success.

II. Methodology

Our research began by acquiring the entire 2015 MLB dataset from the PITCHf/x database using the R package, 'pitchRx'. The data was imported into a SQL database and was composed of five separate datatables: atbat, action, pitch, po, and runner. We then identified 78 qualifying pitchers from the FanGraphs database; these were pitchers who had pitched in at least one inning of every game throughout the season. We then queried every pitch for these qualifying pitchers, yielding a dataset of around 230,000 observations. Some important variables in the feature space were count, pitch x-axis position (px), pitch z-axis position (pz), play event, and type of pitch.

In order to investigate a pitcher's change in strategy, we first wanted to define *strategy*. In the context of pitching, we can identify pitch location and pitch type as two functions that drive a pitcher's decision making. Pitch location can be visualized in terms of the 14 regions in the strike zone. Our analysis uses these regions as a lens of classifying different pitches but relies on the (px, pz) pair, representing the exact location in the strike zone, to determine any changes in pitch location when a runner is on first. For pitch type, we decided to classify all pitches as either *fastball (FB)* or *off speed pitch (OS)*. This simplification of pitch type involved aggregating existing pitch types into these two categories. For example, FB is the combination of all fastballs, two-seam fastballs, four-seam fastballs, cutter fastballs, sinkers, and split-fingered balls, while OS is the combination of all sliders, curveballs, knuckle-curve balls, changeups, knuckleballs, and eephuses.



We then separated the data into two sets by event type. The first dataset queried all events involving a runner on first and an empty base on second; the second dataset compiled all other event types throughout the season. We were then able to calculate the number of fastball occurrences in each scenario and any changes in average pitch location. The results of these findings are discussed in the following section.

The next stage of our analysis attempted to answer the question: Should pitchers change their strategy when there is a runner on first base with second base empty? We grouped our data by three categories: count, pitch type (fastball vs. offspeed), and location (pitch zone, as defined earlier). We developed an objective function – the run expectancy of a given pitch – based on three components: run expectancy when the pitch was put in play (wOBA), run expectancy when the pitch was not put in play, and the change in run expectancy when there was a stolen base attempt. wOBA values were calculated using the "Guts!" data from Fangraphs, run expectancy changes for balls and strikes were adopted from Jon Walsh's 2008 research piece, "Searching for the game's best pitch," in The Hardball Times, and run expectancy changes on stolen base attempts were calculated based on the change in base-out states on the attempts. A distinction for the wOBA values and the run expectancy for non-batted balls is that each of those metrics were weighted by the number of pitches in the given category. Pitches with lower run expectancy values favored the pitcher, while pitches with high run expectancy values favored the hitter.

$$RE_{pitch} = wOBA + RE_{Ball/Strike} + RE_{SB/CS}$$

$$wOBA = \frac{\sum RE_{BattedBall}}{count(BattedBalls)} * \frac{count(BattedBalls)}{count(Pitches)}$$

$$RE_{Ball/Strike} = \frac{\sum RE_{Ball} + \sum RE_{Strike}}{count(Ball + Strike)} * \frac{count(Ball + Strike)}{count(Pitches)}$$

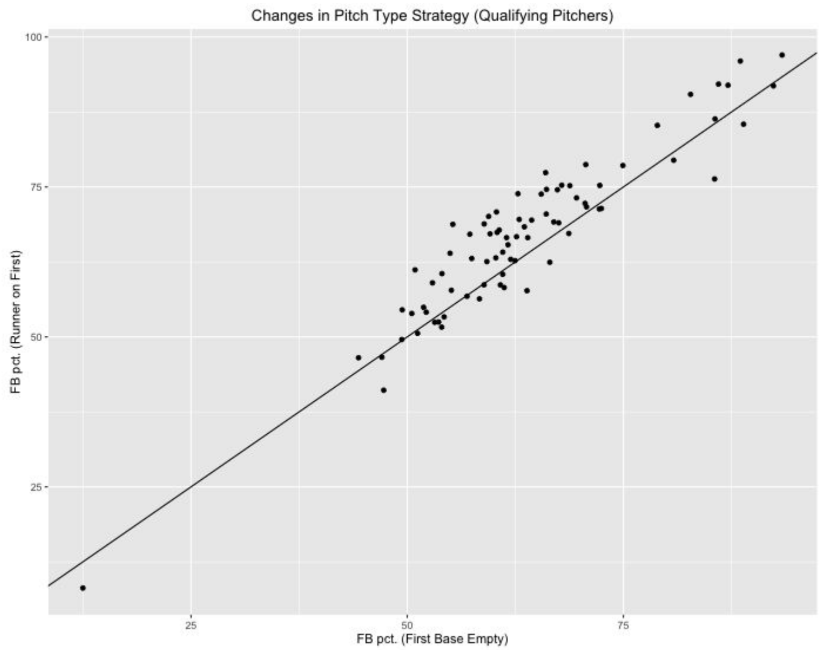
$$RE_{SB/CS} = \frac{\sum RE_{AfterAttempt} - RE_{BeforeAttempt}}{count(Attempts)}$$

Finally, we were interested in determining the optimal distribution of the recommended pitch strategy. Since we grouped by both location and type of pitch, each possible count had 28 pitch recommendations, all with a corresponding expected run value, as calculated above. To calculate the distribution of pitches, we calculated the percentages by assigning weights to the pitches relative to their magnitude. This ensured that, for example, an RE value of -.05 would be pitched 5 times more often than a pitch with a value of -.01. To do so, we took the top 6 pitches in for every count, negated the values and normalized the set. While this is an imperfect method, we believe it provides an approximation that is suitable to this analysis.

III. Analysis: Do pitchers change their strategy?

a. Pitch Selection/Pitch Speed

Our results favor the hypothesis that pitchers tend to pitch faster when there is a runner in stealing position. Using the data from the 78 qualifying pitchers, we found that there was a mean increase of fastball use of 4.7 percent, with 55 pitchers increasing their percentage of fastballs, whereas only 23 of them decreasing in percentage. Examining the graph below, we see that the majority of pitchers fall above the $y = x$ line, indicating that these pitchers use fastballs more frequently when there is a runner in base-stealing position. The further from the $y = x$ line the pitchers dot is, the more he changes his fastball percentage given a base stealing situation.



Further examining the data on pitchers who increased their percentage vs. decreased their percentage, we found that the pitchers who increased tended to do so more drastically and consistently. The top nine pitchers who increased fastball rate increased by a mean of 18.3, with a variance of 6.2, suggesting that these pitchers actively were changing their pitch selection to account for the base stealing situation. The bottom 9 pitchers, excluding R.A. Dickey who was our only knuckleballer and a clear outlier in the data(with a -34.9 percent decrease in fastball usage, a 21.9 percent difference from the next nearest pitcher), decreased their fastball percentage by a mean of 7.0 perecent, with a variance of 13.0. This small mean decrease and high variance suggests that these pitchers most likely kept their pitching strategies the same, and their decrease in fastball usage was simply random noise. It is most likely that these players fall into the category of players who do not actively change their strategy, similarly to the players who saw increased fastball usage of between 0 and 13 percent.

Players with Highest and Lowest Percent Change of Fastball Usage:

Top 9

Name	%FB Change
Kyle Gibson	24.4
Chris Sale	20.2

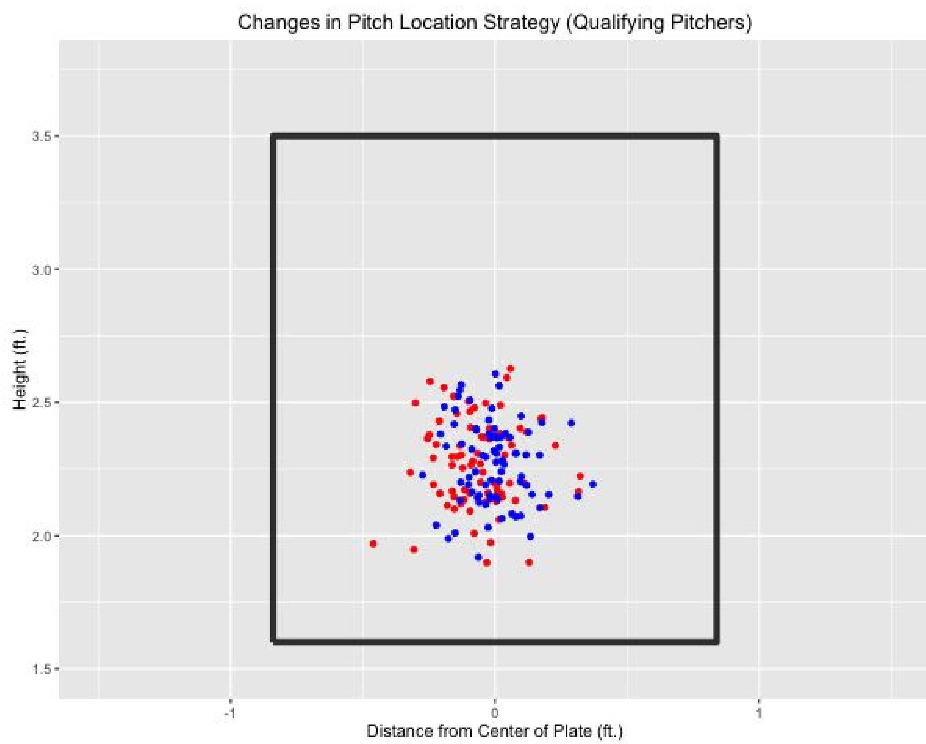
Bottom 9

Name	%FB Change
Rubby de la Rosa	-3.5
John Lackey	-3.9

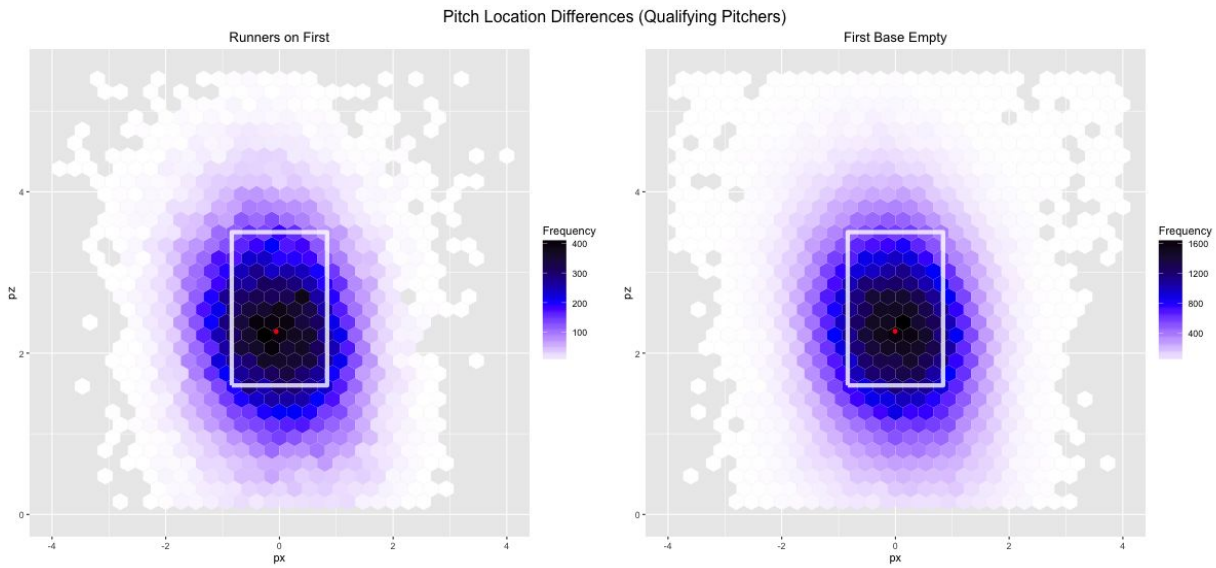
Matt Harvey	18	Tyson Ross	-4.4
Rick Porcello	17.6	Jimmy Nelson	-4.9
Anthony DeSclafani	17.4	Dallas Keuchel	-6.1
Jose Quintana	17.3	Alex Wood	-9.6
Danny Salazar	17.2	Bartolo Colon	-10.8
Chris Heston	16.9	Colby Lewis	-13
Carlos Martinez	16.4	R.A. Dickey	-34.9

b. Pitch Location

When analyzing the data on pitch location, we found that there was no difference in pitch location selection due to a base stealing situation. Examining the below graph, we found that the mean difference between situational centers for all pitchers was only 1.2 inches, an insignificant amount. The red points correspond to the average position for each pitcher in a base stealing position and the blue points represent the average pitch in all other events.



Further, by combining all the data for pitchers, we were able to create a heat map of pitch location during base stealing situations vs. non base stealing situations, and it is clear that the pitch location distribution remains fairly consistent in both situations. Note that the red dot represents the overall mean position of the ball in the strike zone during these two scenarios.



After examining all of the data, we concluded that some pitchers, though not all, alter their pitching strategy when faced with a base stealing situation. These pitchers alter their strategy by throwing more fastballs, as would be expected, as a fastball gives the catcher more time to throw out the base runner at second. We concluded that pitchers do not in general change their pitch location based on the situation, as the pitch location distribution for all of our pitchers remained fairly constant in both base stealing and non base stealing scenarios.

IV. Analysis: Should Pitchers Change their Strategy, and How?

The short answer is, yes, pitchers should change their strategy when a runner is in base stealing position. For our analysis, we chose to focus on the 0-0 count, as it had the most data points and therefore the most reliable results. Our analysis can be applied to the other pitch counts, but lack of data points made the results for uncommon counts less reliable. Observing at the pitch values for 0-0 counts, it is fairly clear that fastballs are more beneficial than off speed pitches. The mean run expectancy of fastballs in any zone with a runner in base stealing position is .046, as compared to .087 for an offspeed pitch. Looking at the chart for most beneficial pitches below, we can see that the pitch with the lowest run expectancy is a Zone 10 fastball, followed by Zone 3 and Zone 1 fastballs. This is unsurprising, as all three of these pitches are great for catchers throwing out potential base stealers. Being a fastball means that the ball reaches the catcher more quickly, giving him extra time to pick off the base runner. Additionally, Zones 10, 3, and 1 are all high in the strike zone, making it faster for the catcher to catch and throw, giving him additional extra time to pick off the base runner. It is curious, however, that the most optimal pitch is a ball, as one would expect a strike to decrease run expectancy more than a ball would. The reason a ball could be the optimal pitch is that, on a 0-0 count, whether the pitch is a ball or a strike is far less important than on a higher pitch count, and a thrown ball is harder to put into play, thus potentially decreasing the wOBA on those specific pitches. As we get into higher pitch counts, especially ones with 3 ball or 2 strikes, the value of throwing a strike skyrockets, and we see much lower instances of thrown balls being included in the most optimal pitches.

The chart at the right is potentially very useful to pitchers attempting to optimize their strategies. Simply, a pitcher should attempt to throw more of the pitches higher up on the chart, and avoid pitches lower down. It's important to note, however, that it is not as simple as only throwing zone 10 fastballs any time the count is 0-0. The pitcher must weigh the cost of throwing a less optimal pitch against the cost of throwing a more predictable pitch. If a pitcher threw only zone 10 fastballs at 0-0 counts, the batters would quickly catch on and be able to take advantage. Finding the optimal way to weigh these two decisions would certainly be a very involved and interesting area of further study.

Count	Zone	Pitch Type	Pitch Value
0-0	10	FB	-.1257412
0-0	3	FB	-.0742766
0-0	1	FB	-.0371552
0-0	7	OS	-.0263108
0-0	9	FB	-.0149205
0-0	3	OS	-.0100417
0-0	13	OS	-.0008116
0-0	11	OS	.0166525
0-0	12	FB	.0197576
0-0	2	FB	.0208645
0-0	7	FB	.0301951
0-0	9	OS	.034186
0-0	2	OS	.0519219
0-0	13	FB	.0531164
0-0	6	OS	.0594854
0-0	0	FB	.0657125
0-0	11	FB	.0743714
0-0	0	OS	.0792293
0-0	12	OS	.1017401
0-0	4	FB	.1179624
0-0	8	OS	121214
0-0	5	FB	.1295268
0-0	1	OS	.1316602
0-0	10	OS	.1459091
0-0	6	FB	.1558297
0-0	5	OS	.2030969
0-0	8	FB	.2288444
0-0	4	OS	.2973895

When determining how a pitcher can change his strategy to a more optimal one, we decided to only look at the top six most optimal pitches for a given count, as it is infeasible for a pitcher to hold 28 different types of pitches in his repertoire and in his mind at any one time. In the table below, we have those six pitches, with the percentage of the time they are being thrown (considering only the times those six are thrown) and the percentage of the time our model suggests they should be thrown, calculated using the technique described in our methodology section. Our model encourages pitchers to throw more Zone 10 and Zone 3 fastballs, but far less Zone 9 fastballs. These results suggest that throwing high fastballs becomes more advantages in base stealing situations, and should be thrown more frequently than they are being thrown currently.

In our previous results, we found that pitchers tend not to change up their pitch location during base stealing situations, so one way for pitchers to use these results to create a more optimal pitching strategy would be to throw more high fastballs in those situations. One curious result from this data set is that our model does not encourage to throw pitchers to throw more fastballs. The rate of fastballs went from 87.3 percent to 87.4 percent, a negligible change. However, it is important to note that 87.3 percent is already an uncharacteristically high rate to be throwing fastballs, and this surprising result is most likely caused by the 6 optimal pitches having an uncharacteristic distribution, rather than discrediting the notion that pitchers should throw more fastballs in base stealing situations.

Count	Zone	Type	Frequency	Percentage	Suggested Percentage
0-0	10	FB	872	34.9%	43.6%
0-0	3	FB	370	14.8%	25.8%
0-0	1	FB	435	17.4%	12.9%
0-0	7	OS	222	8.9%	9.1%
0-0	9	FB	506	20.2%	5.2%
0-0	3	OS	96	3.8%	3.5%

Ultimately, our suggestion to pitchers in a base stealing situation would be in general to increase the number of fastballs pitched, as well as try to pitch more balls in zones 10, 11, 1, 2, and 3, in order to give the catcher a better chance at throwing out the base stealer. For a more nuanced suggestion, the pitcher should examine our chart for pitch value at each count, and try to increase the number of pitches higher up on the chart, and decrease the number of pitches lower down.

V. Limitations and Possible Extensions

The main limitation for our project was a lack of data. Despite having over 200,000 different pitches to analyze, due to how rare an event a stolen base is, we were unable to have a large enough dataset to analyze some more nuanced situations. One extension we tried to implement was us k-mean clustering to provide more targeted recommendations to different types of pitchers. We were able to successfully cluster our 78 qualifying pitchers into 6 different groups, but ultimately were unable to find anything significant differences between the success of different strategies between different groups of pitchers due to the low occurrences of stolen base attempts in our data set.

We also tried to further granularize our pitch selection options, as we had to essential combine all fastballs and all offspeed pitches into two categories, even though there are many different types of pitches that classify as an offspeed or a fastball, each of which most likely has varying success in base stealing situations. Again, we were unable to do this, as we were lacking in the data to subdivide more than the binary of fastball vs off speed. If we had kept the original pitch classifications without combination, we would have had 12 different pitches instead of 2, resulting in approximately one sixth of that data per pitch as we had originally. As it is, we already had to subdivide the data into 12 pitch counts * 13 zones * 2 pitches = 312 different groups, which resulted in many of the more rare pitch count/zone combinations having scarce data. Dividing it into 1872 different classifications would have rendered our results essentially meaningless and inconsequential.

Finally, an interesting extension would be to see how pitcher strategy changes in regards to the potential base stealer. Namely, how does having a known base stealer on first affect strategy as opposed to a generic player. Obviously, pitchers are not going to change their strategies if Prince Fielder of Pablo Sandoval is on first base, but when are in a base stealing position, pitchers are obviously thinking about it. Again, this would require more data then we have, as parsing our already sparse data by player on first would lead to too few data points for any significant results.