

CAUGHT IN THE CROSSHAIRS

A dynamic action shot from a baseball game. A player in a white uniform with blue and red accents is sliding into a base, kicking up dirt. Another player in a white uniform with purple socks is reaching down with a glove to tag the sliding player. The background shows a green field and a yellow sign.

Middle Infield Positioning
to Prevent the Extra 90'

<https://github.com/TrevinC/SMT-Data-Challenge>

ABSTRACT

This paper investigates optimal defensive positioning as it relates to the placement of middle infielders about second base during an attempted steal of second base from first. The physical landscape of the game has been altered: an increase in base size has resulted in a 4.5" reduction in the length of the basepath, which limits the distance across which baserunners can be caught stealing, an interaction known to take place across remarkably small intervals. In addition to these new dimensions, the pitch timer rules implemented by MLB at the start of the 2023 season limits "pick-off" attempts by the pitcher, resulting in an increase in attempted steals. Given the increase in steals, we explored optimal defensive positioning as it relates to the placement of middle infielders around second base during an attempted steal of second base from first. "Plays of interest" (POI) were identified as those plays in which a player attempts to steal a base and was parsed from minor league games' data provided by Sports Media Technology (SMT). The defensive player positioning across these POIs in which the baserunner was successfully "caught stealing" was subjected to a spatial analysis used to define defensive "regions of success." This region for both the second baseman and the shortstop was found by plotting and finding the centroid of the initial second base and shortstop defensive positioning during successful caught stealing plays and is intended to illustrate the regions from which either position has had an improved success rate of catching a base runner in an attempt to steal second from first.

STEALS ARE IMPORTANT AGAIN

The pitch timer was implemented in Major League Baseball at the start of the 2023 season, after a successful period of experimentation in the minors. As part of the new rule, pitchers are relegated to two disengagements (pick-off attempts or step-offs) per plate appearance, a game change that has dramatically altered the context upon which one of baseball's more dramatic interactions unfolds: the first base baserunner's attempt to steal second, and the defensive team's efforts to thwart him. The physical landscape has been altered as well, with a 4.5" decrease in basepath distance attributed to increased base size.

According to MLB, minor league players attempted to steal bases more often in 2023 than they had prior to the rule changes: attempts increased from 2.2 to 2.8 per game, a 27% increase. Not only were players attempting to steal with increased frequency – but their rate of successful steals increased as well. The same data describing the 2022 season shows a 10% jump in success rate, from 68 to 78%. (1)

Increases in steal attempts should place more emphasis on player positioning of both the second base and shortstop defenders about second base. Increased steal attempts directly result in an increased number of corresponding opportunities defenders have to tag a base runner out during an attempted steal, a play in which the base runner is said to have been “caught stealing.” Defenders’ spatial orientation as it relates to second base should then be related to the success rate of defenders catching baserunners stealing. So what is the best positioning?



REGIONS OF SUCCESS

The investigation identifies two “regions of success:” one for each the second base and shortstop positions. These regions are circles containing and circumscribed by the initial player positioning during plays in which a baserunner is successfully caught stealing. These regions do not guarantee success – they also contain instances where the attempted “pick” (tagging a baserunner out) is unsuccessful. However, it remains that no successful pick falls outside of the region of success.

DATA

The plays in question involve players in the defensive pitcher, catcher, second base, and shortstop positions, while also involving offensive baserunners. Location data for each of the aforementioned positions was sourced in x,y coordinate format with the home plate specified as the coordinate plane’s origin. We decided to stick with data from the last two years that were recorded, the 1902 and 1903 years from the anonymized data we received from SMT. The early data was at times incomplete and broke the code because of inaccurate logs.

METHODOLOGY

SPATIAL DATA ANALYSIS

We were able to log 106 POIs from our chosen dataset. We parsed the data by singling out play IDs where a catcher threw the ball. It seemed the quickest way to remove the most amount of data that wasn't relevant to our investigation. After pulling these we eliminated back picks at first base and throwouts to 3rd by removing any plays where the first baseman or third baseman was the next to hold the ball after the catcher threw. We had to account for ball bounces and deflection off both defenders and baserunners. Next, we removed any plays that involved a hit, which dealt with bunts and dribble hits fielded by the catcher. We were able to pull all these in code and make a file of all our POIs. The data was then manually parsed, where we compared the game event data to the game info to piece together our final lists of caught stealing and successful steals. The nature of the game made parsing this computationally difficult, to account for every deflection, ball bounce, and missed throw. Originally when parsing our data we thought to just drop every play involving the first basemen, but with that, we would have lost out on a caught stealing that resulted in a pickle! We were able to find an exemption to most rules we could write. Once we had our manually parsed data, it was loaded back into the code to pull our timestamps and spatial data from the SMT datasets.



METHODOLOGY

Game String	At Bat	Play ID	Timestamp	Player Pos.	Event Code
xyz	8	25	824592	2	3
xyz	8	25	825792	6	2

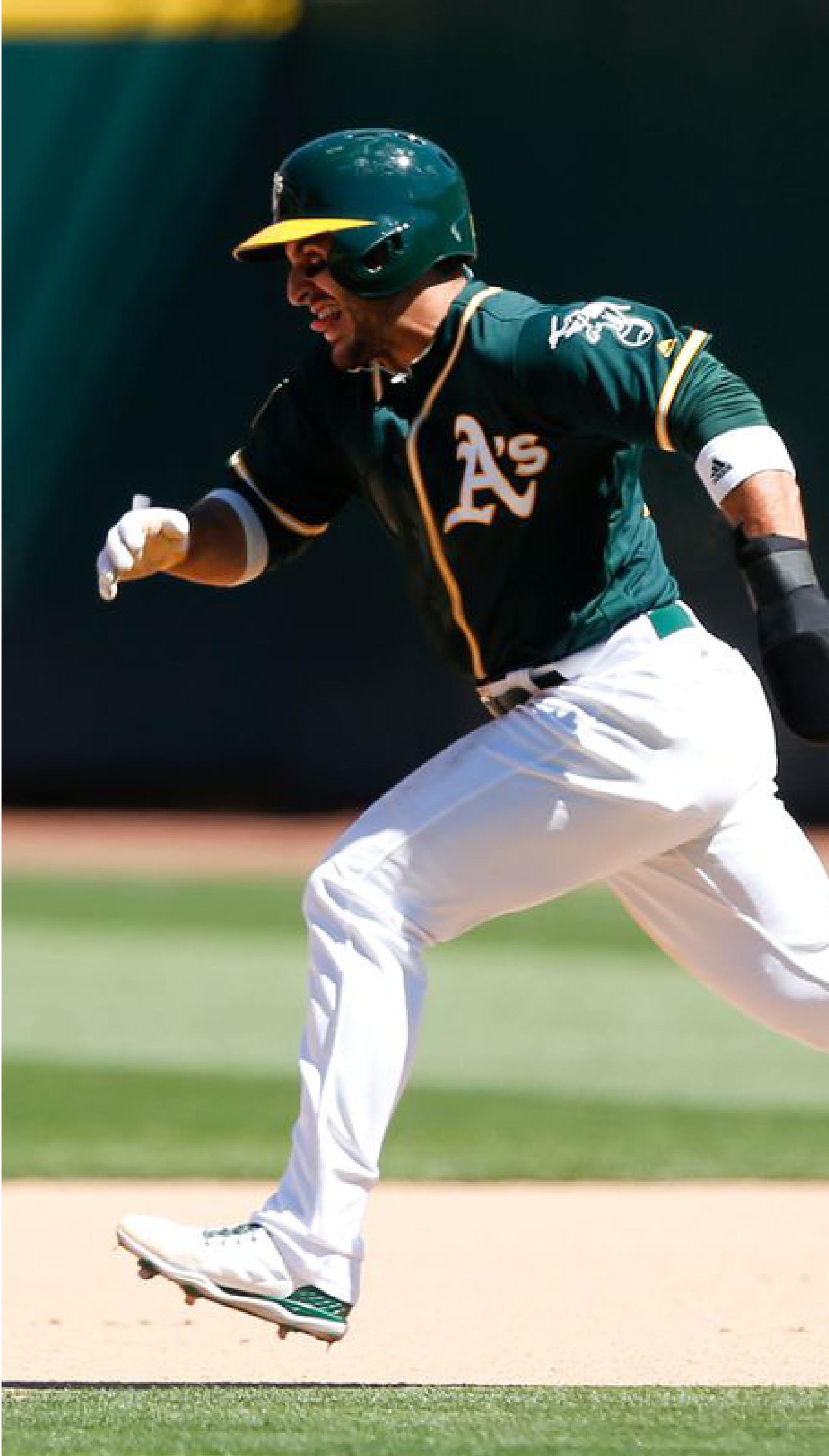
Table 1: In this table, the events of the 25th play of game xyz, which occurs during the 8th at bat of the game, are described by a timestamp, as well as event codes and player position. These events are Catcher throwing the ball (player position 2, event code 3) at timestamp 824592, with the Shortstop catching the ball (player position 6, event code 2) at timestamp 825792. Data resembling these events were parsed for further investigation into whether they described an attempt at catching a player stealing.

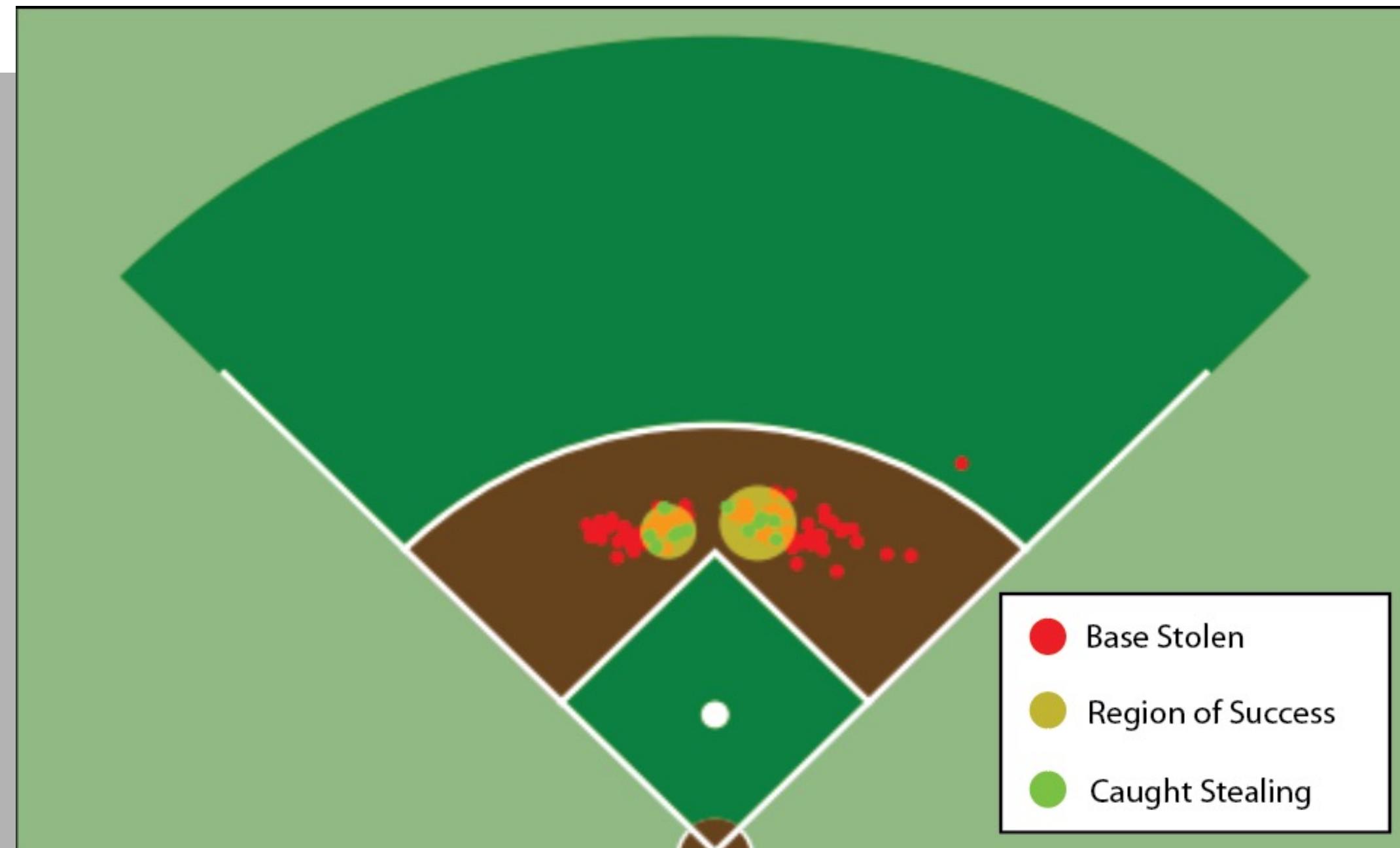
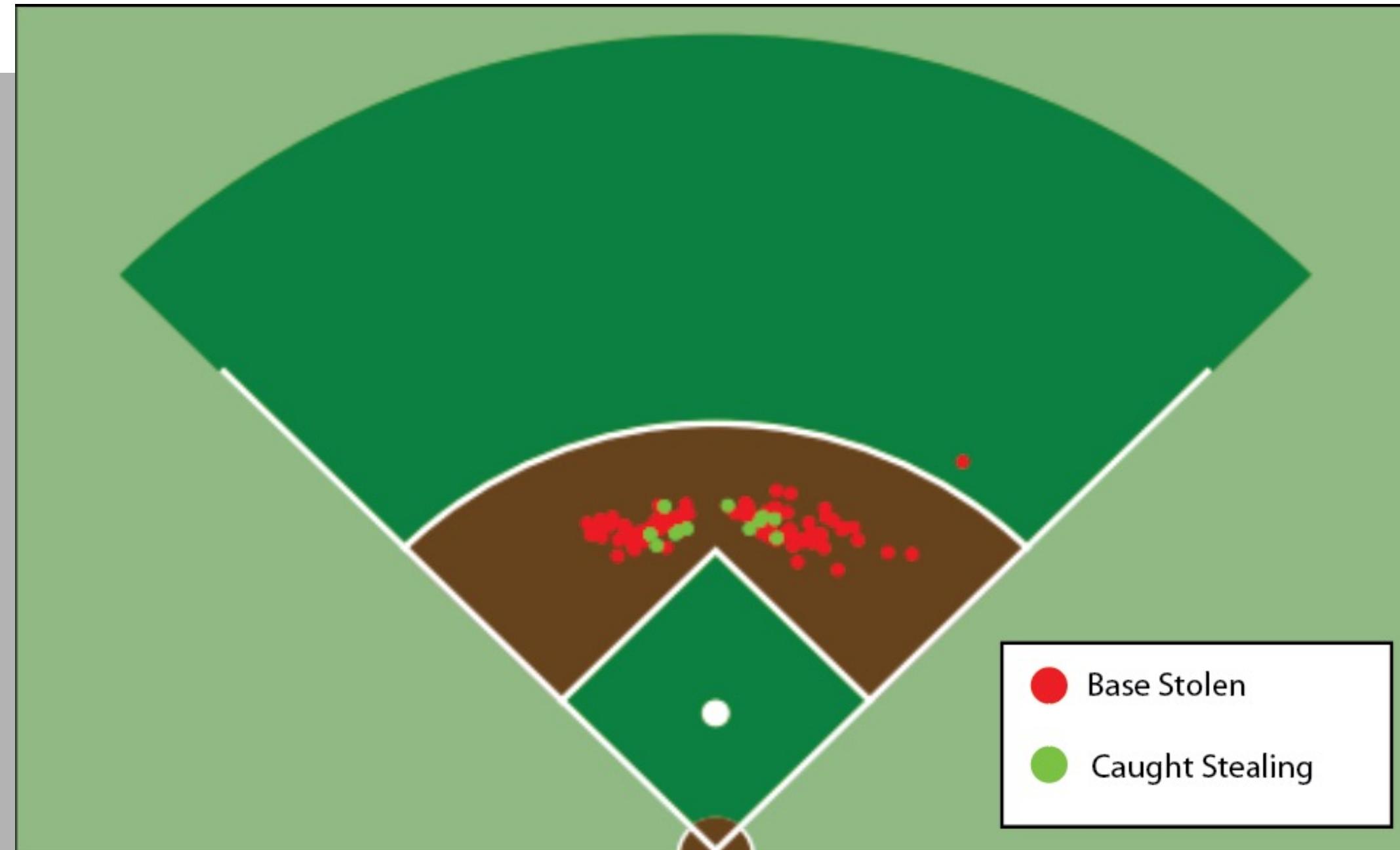
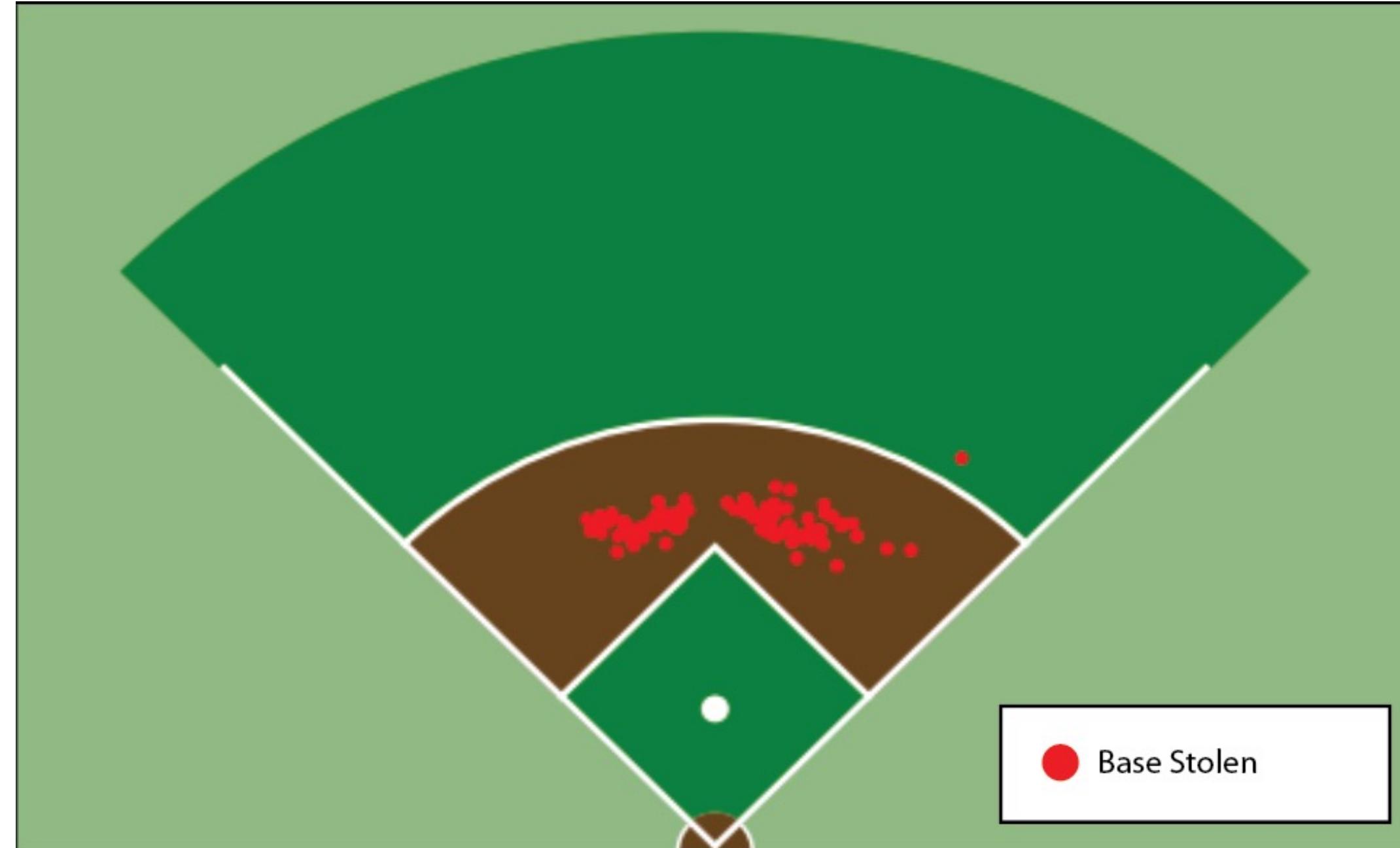
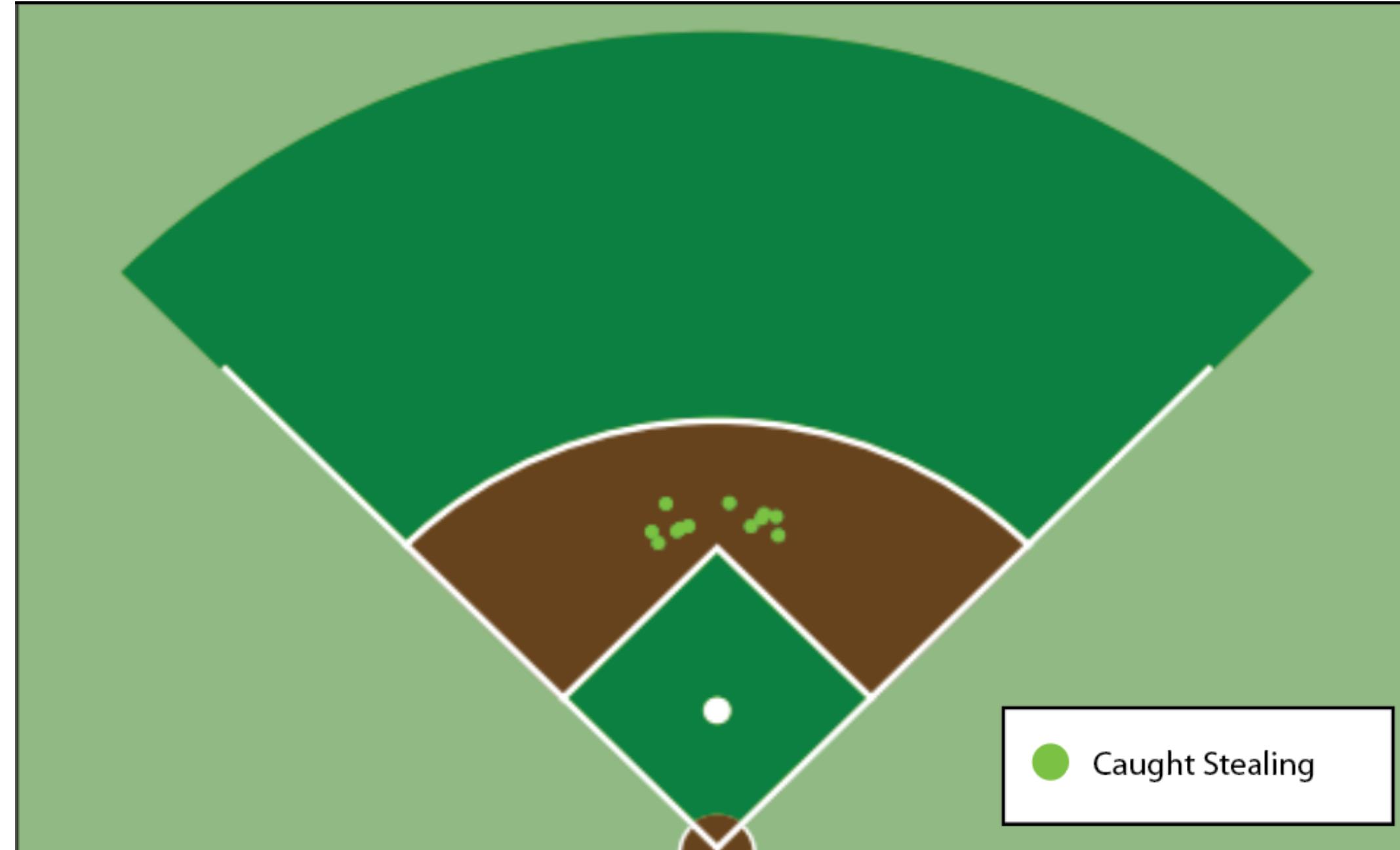
Game String	At Bat	Play ID	Catcher	SS	Batter	1st Base	2nd Base	3rd Base
xyz	8	25	9185	8787	2386	2766	0	0
xyz	8	26	9185	8787	2386	0	0	0

Table 2: This table shows an excerpt of game info data, which describes which players are occupying defensive and offensive positions across plays and at bats. Players are demarcated with four-digit identifying codes, with defensive players highlighted in blue and offensive players highlighted in yellow. Omitted from this excerpt are players that did not necessarily participate in the determining events of the 25th play of game xyz, e.g. 1st and 3rd base defenders. Notice how between plays 25 and 26, the 1st base baserunner is ruled out (highlighted in red) before the same at bat resumes. We assume this to be the result of the events described in Table 1, or the Shortstop tagging the runner out at 2nd base after catching the ball thrown by the Catcher.

Game String	Play ID	Timestamp	Player Pos.	X Coordinate	Y Coordinate
xyz	25	820292	6	-15.1044	137.6448
xyz	25	824592	6	-3.9309	129.8019
xyz	25	826642	6	2.2239	126.7959

Table 1: This table displays the position of the Shortstop (6) from the start of the 25th play of game xyz (timestamp 820292), to the moment the Catcher throws the ball in an attempt to catch the 1st base baserunner stealing (timestamp 824592), to the moment the play concludes (timestamp 826642). Position is represented by xy coordinates across an xy plane with the origin at home plate, towards 1st and 2nd base indicating the positive x- and y-directions, respectively. Notice how the Shortstop moves from the negative to positive x, indicating that they have traveled into the baserunning path between 1st and 2nd base during this successful “caught stealing.” Their starting coordinates (-15.1044, 137.6448) were incorporated into the region of success for Shortstops.





CONCLUSION AND FUTURE OPPORTUNITIES

These findings highlight the importance of precise defensive positioning in response to the increased steal attempts. By aligning themselves within these defined regions of success, teams can maximize their chances of catching baserunners stealing second base, effectively countering the resurgence of stealing as a pivotal aspect of the game in the 2023 season. These regions are based on successful caught stealing plays and provide valuable insights for defensive strategies.

Most current player positioning aligns exclusively with the hitters' spray pattern, but if we were able to take our player positioning theory and weigh it with the hitter's spray pattern, the runner at first' sprint speed, and the catcher pop time, we believe we could make a multi-objective optimization program to align the factors for total run reduction. On the other side as well, once runners could be educated about ideal player positioning, they could take advantage of this. It could give runners an edge if they notice the shortstop or second baseman has strayed from the "radius of success."

