**Introduction**:

This document is detailed with the proposal for answering the questions asked by an MLB infield coach:

“One of our infielders, Player X, seems to be struggling in the field. He’s got a great arm, but he’s made a few errors this season and is failing to get to some balls. Could you look into this and identify any problem areas that we can target with drills?

Measuring fielding ability quantitatively has been an ongoing challenge for baseball statisticians. However, baseball has recently been equipped with higher statistics, such as Statcast, which provides more detailed information and playing tracking in the field. This new form of data allows statisticians to build more sophisticated models modeling player fielding performances. This proposal aims to show how this new data can pick out certain players’ weaknesses in the field. They can then take those weaknesses to build drills and practices around bettering their mistakes.

**Potential Approaches:**

Investigating issues with ball tracking and ground ball errors can take several forms, including:

* Examining fielder placement before the ball is hit depends on the batter and pitcher matchup. Depending on Player X's position, where they position themselves in the field based on the hitter at bat can shed light on whether they have the chance to get to a ball or not. For example, a more than often lefty pull hitter (teams shade, but do not shift for him) is at bat, and Player X is a second baseman. If he plays straight up or double play depth, then he is more than likely not going to get to balls pulled between the first basemen and him.
* Examining the ball in play characteristics. Balls put in play at a certain velocity as well as angle are sometimes harder to field than normal. Balls batted at above 95+ MPH are considered hard hit and have an increased chance of performing an error on said ball. As a result, if Player X is receiving a plethora of hard-hit balls in his area and he shows difficulty fielding those, that is an area to be worked on.
* A third thing to be examined could be fielding statistics based on position in the field. Since Player X has a great arm, he would most likely be slotted at third or shortstop, given their position's longer throws. But if an analysis of his errors (specifically throwing errors) per position shows that he makes more at second base, maybe he doesn't have a problem with power. Instead, accuracy and drills surrounding increasing a thrower's accuracy can be targeted.
* Included together with what’s above, but maybe separately, is an examination of Player X's misplays. What kind of errors and mistakes is he committing often? Poor throws (whether that's an overthrow or under throw, wide left or right), simple errors (ball rolls under glove), two fielders go after same ball, and any other infielder error. These will tell us his most common error, highlighting what he needs to work on.

Hopefully, data available through any of the baseball websites (Savant, baseball-reference, FanGraphs, or retrosheet) would be able to provide for analysis on each of these primary questions.

**Data**:

Like I mentioned in the introduction, fielding statistics have been questionable for a long time in the baseball world. However, now that we have statcast data that provides more detailed data and position tracking data for everything happening on the field, this kind of data will be better in fielding analysis.

Play-by-play data, as well as basic baseball statistics data, would be primary to use.

Variable:

locationX - player location on an x-axis grid of a baseball field (included in statcast)

locationY - player location on a y-axis grid of a baseball field

Shift - a variable detailing a shift or not (0 = no shift, 1 = shift)

Zone Rating (ZR) - plays a player makes in his given zones divided by opportunities. Numerically tracks the percentage of fielding.

Direction - direction of batted ball (indicated by a vector)

Speed - MPH of batted ball (exit velocity, categorized in three categories of soft, medium, hard)

Type - classified as groundball, liner, fly ball, “fliner” (balls considered between fly ball and line drive)

Distance - how far the batted ball went (in feet)

Ultimate Zone Rating (UZR) - similar to ZR but measures the actual number of runs a player saves due to his defense

Defensive Runs Saved (DRS) - metric for tracking runs prevented based on location of the ball in play

Probabilistic Model of Range (PMR) - where a player fields balls relative to the average player at his position.

**Limitations**:

There’s a multitude of limitations on defensive metrics even today. I think the study has greatly improved over the last ten years, but we also have more room to grow. For this project specifically, I see a problem with the shifting metric of a team. Each team does the shift differently, sometimes adding an extra fielding over there or just shifting the players over but staying on their side of second base. There's also the current debate about some of the variables put forth for analysis. ZR is measured subjectively, with the scorekeeper dividing the field and charting the path of the hit himself. This can cause a human error in charting this statistic.

**Methodology:**

As described above, there is a multitude of directions a project like this could take. I think the best plan would be to incorporate the most common data applicable to us. The infield coach mentions that Player X has a hard time getting to balls. More often than not, this is not due to athletic ability but player placement, depending on the batter. Dansby Swanson is not known to have a great arm for a shortstop, but what he does exceptionally well is place himself in the best position prior to the pitch to get to the ball as soon as possible to beat out a quick runner. He is knowledgeable about where to be at the right time.

An analysis of player placement is difficult, but with “direction,” we can track whether the side of the body that Player X fields the ball on has an effect on his chance of making an error. If we can visually show or numerically show that one side of his body is worse than the other, the coach can target the weaker side in practice. We can also use the location data in Statcast compared with the ZR as well as the PMR for further position analysis. ZR will show us where the balls are hit in Player X’s zone that he either made a play on or did not. If he is constantly not making plays on balls hit in the hole, and we compare where his starting location is, and he is shaded towards second, then we can obviously see a problem. An important disclaimer is needed, however. This is all dependent on the hitter, so it would be useful to filter through the most common hitters that Player X faces in a season. PMR actually plays the biggest part in positioning analysis. The graph shows how good Player X is at fielding balls relative to the second-base bag (if he's playing short or second, third if he’s playing third) and also relative to the average major-league infielder.

This can be combined with the batted ball characteristics. We would need statistics showing how often Player X made an error on soft, medium, and hard hit balls and the type of hit (groundball, liner, fly ball, and fliner). This can give us a quick answer to the balls that Player X has the most trouble with. Produce these two reports, which can be handed to the coach for direction on what he wants to specify during practice.

**Resources:**

[Sabr](https://sabr.org/journal/article/measuring-defense-entering-the-zones-of-fielding-statistics/) - measuring defense: entering the zones of fielding statistics (helped inform me about tracking a fielders effects for a team)

[JQAS (Journal of Quantitative Analysis in Sports)](https://repository.upenn.edu/cgi/viewcontent.cgi?article=1154&context=statistics_papers) - sparked some ideas on models that could be used in the analysis of fielding statistics