Miles Kee

UCSAS 2022 Data Challenge

An Analysis of Player Movement by Position

1. Introduction

The analytics revolution that has changed baseball has had many profound and tangible impacts on the game that can be seen by even the most casual of fans. Compared to 10 or 20 years ago, today's batters focus more on power at the expense of contact than ever before. Teams value walks and extra base hits over singles from their hitters and strikeouts above all else from their pitchers. The analytics impact has even been seen on the basepaths, as stolen bases, bunts, and other "small ball" characteristics of baseball have been on the decline due to mathematical expectancies of those plays. One area that the analytics revolution has not made as much impact is on defense. Defensive metrics have long been an area that left more to be desired analytically, as the many aspects of defense have been hard to quantify. Before Statcast was implemented in 2015, it was quite hard to quantify things like arm strength, range, and routes to balls. The ability to now look at player tracking data has changed the way the analytics community views defense, as new metrics like OAA have replaced outdated ones and defensive wizards like Tommy Edman's value has begun to be appreciated. I chose to conduct a study on relative defensive positional difficulties based on movement during games. I looked at the distance moved by each player on the field during a game to attempt to quantify relative positional difficulty. I then applied these relative positional difficulties to defensive metrics at the major league value to see if certain players' defensive impact were being over or undervalued.

2. Methodology

All my analysis took place in the player_pos file. To start, I first filtered the data frame by position to create one each for first basemen, second basemen, third basemen, shortstops, left fielders, center fielders, and right fielders. I chose to not include pitchers or catchers as pitchers rarely move on defense, and the defensive difficulty of the catcher position is not due to movement. I conducted my analyses on each position independently, and then combined them at the end.

For each position, the first thing to do was find my sample of games to use. Since each position had millions of data points, it was unrealistic to write a loop for all those points. I instead used R to choose a random sample of 10 games for each position, which gave me over 10000 observations for each position. I then wrote a loop to mark the start and endpoints of each play in these games and sum the distance moved during the play using the coordinates and timestamps given. The next step was to compile the distances and times for each play, and I had my data to conduct my analysis.

3. Analysis

The first thing I wanted to look at was all the distances moved to get an idea of what my data looked like. Figure 1 is a histogram of the distances covered on each play in my study. As expected, the most common distances are smaller, and as distance increases, frequency decreases. Figure 2 reveals some interesting positional pieces of the data. The boxplots show that centerfielders have the highest range of movement, essentially meaning that they will have the highest maximum travel distance on a play. First basemen have the smallest range, which makes sense. Since they will have to be near the first base bag to cover it on any given play, they do not travel far distances. I was a bit surprised to see that second basemen have a wider range than shortstops. However, when you think about the modern game, it makes more sense. With the shift more prevalent than ever, second basemen are often positioned as the end point of the shift to right-handed batters. With the third basemen deep behind the bag, the shortstop farther in the hole, and the second basemen usually right behind the bag, that leaves the entire right side of the infield for them to cover. Any slow ground balls towards the normal second base position will cause the second basemen to travel far distances to attempt to make a play on them.

The plot of each position’s mean distance moved looks about how you would expect. Shortstops have the highest mean among infielders, followed by second base, then first base, then third base. Third basemen being as low as they are (especially relative to first basemen) can be explained by the amount of plays third basemen do not move. Since first basemen make some sort of movement to cover the bag on almost every play, they are moving on almost every play. Conversely, third basemen do not have a bag to cover on every play and therefore do not need to move on plays where the ball is not hit to them. In this study, third basemen were almost 10 times as likely to not move on a play than first basemen were, which explains why their average is much lower. In the outfield, centerfield has the highest mean, followed by right field, and then left field. This aligns well with intuitive defensive difficulties.

Order wise, the standard deviation plot is the same as the means plot. In the infield, shortstop has the highest standard deviation of movement, followed by second, first, and third basemen. In the outfield, center field is followed by right and left field. One interesting thing from this plot is how close second basemen and shortstop are. Although shortstops have more ground to cover than second basemen, the variance on a per play basis of each position’s distance moved is almost the same.

4. MLB Application

To apply my findings to the MLB level, the first thing I did was get the Fangraphs defensive leaderboards for each position, setting my qualification as a minimum of 100 innings played in the field. I then divided this data into two groups, infielders and outfielders. I did this because the distance needed to cover by infielders is much less than the distance outfielders need to cover, and therefore all outfielders have a higher mean distance traveled than infielders, so adjusting both groups together would be unfairly biased towards outfielders. I also used different metrics for infielders and outfielders, since I was already treating them as their own groups. I used OAA for outfielders since it is generally considered a more polished metric for outfielders rather than infielders, and DRS for infielders for the same reason.

The next thing I did was create an adjustment for each position. This adjustment was found by dividing each position's mean distance moved by the overall position group's mean. So, for example, the first basemen's adjustment was found by dividing the mean first base movement by the mean infield movement. This gave me a number for each position that was equivalent to the position's mean distance moved as a percentage of the average. I then found each player's adjusted metric by multiplying their DRS or OAA total by their adjustment, or dividing it by their adjustment if their metric value was less than 0. To create a value to compare across players, I subtracted their overall rank after the adjustment from their overall rank before the adjustment.

As expected, among infielders, first and third basemen were penalized the most and second basemen and shortstops were rewarded the most. The infielders who had the biggest drops were two third basemen, Max Muncy and Ryan McMahon. They were originally tied for 24th among infielders, but dropped to a tie for 58th after the adjustment. The infielders who saw the biggest reward were three shortstops: Tim Anderson, Didi Gregorius, and Javier Baez. The three were tied for 201st among infielders and moved up to a tie for 180th after the adjustment. At the top spot of the rankings, KeBryan Hayes was replaced by a tie of Jorge Mateo and Nico Hoerner. Among outfielders, left fielder were penalized most and center fielders were rewarded the most. The biggest droppers were a group of left fielders that includes Miguel Andujar, Adam Duvall, and Alex Verdugo. The group was tied for 71st originally, and after the adjustment moved to a tie for 84th. The biggest risers were a group of center fielders including Akil Badoo and Nick Senzel. They were originally tied for 139th and moved to 126th. Max Kepler, a right fielder, remained at the top of the rankings despite a neutral adjustment. Figures 5 and 6 show the average rank difference for each position, with a positive number meaning the position's rank moved up and a negative number meaning the position's rank lowered.

5. Conclusion

This study gave a helpful insight to positional movement. Popular defensive metrics like DRS do not include a positional movement adjustment, so the experiment above is an interesting preview of how defensive metrics could change as player coordinate data becomes more available and understood. The averages, standard deviations, and adjustments made that were mentioned can be found in figure 7. While these metrics are interesting and useful, they are not perfect. Obviously, position difficulty can not be equated to simply the distance traveled by each position. Distance traveled does not include how hard balls are hit at you, how often balls are hit to you, how strong your arm is, or any number of other key parts of a player's defensive prowess. This study is only a piece of the puzzle as we as an analytics community look to properly appreciate players' defensive value.

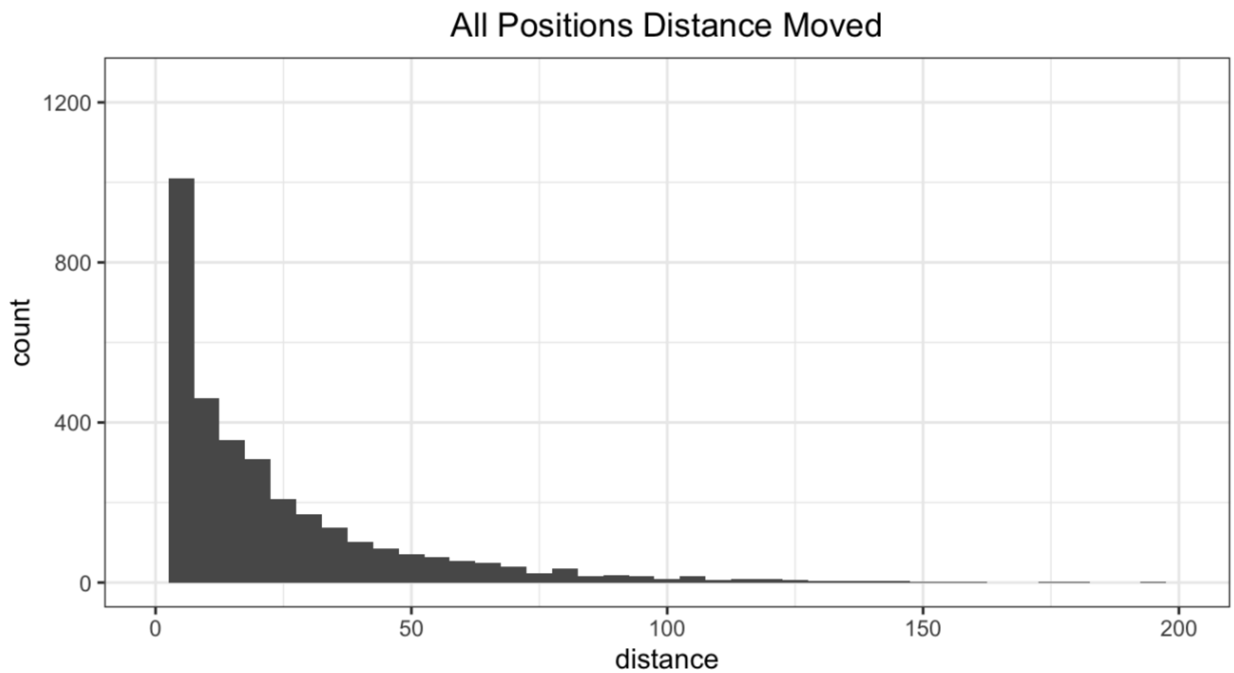

Figure 1: All distances moved.

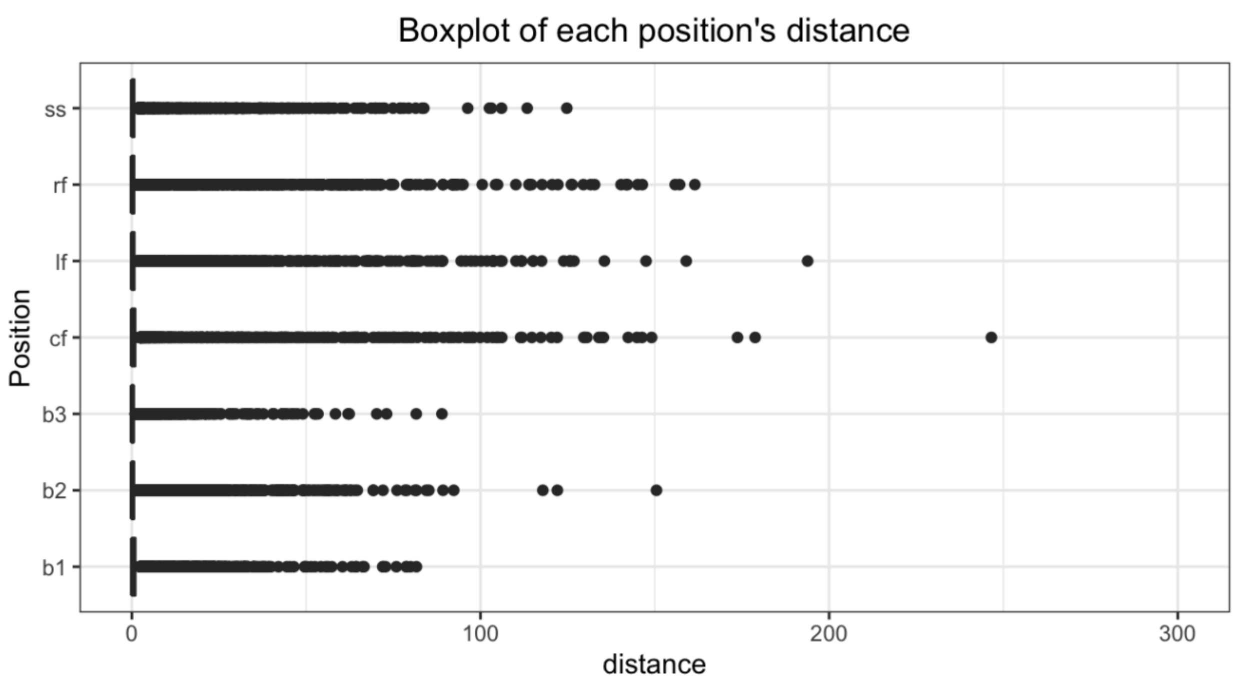

Figure 2: Each position's distance moved.

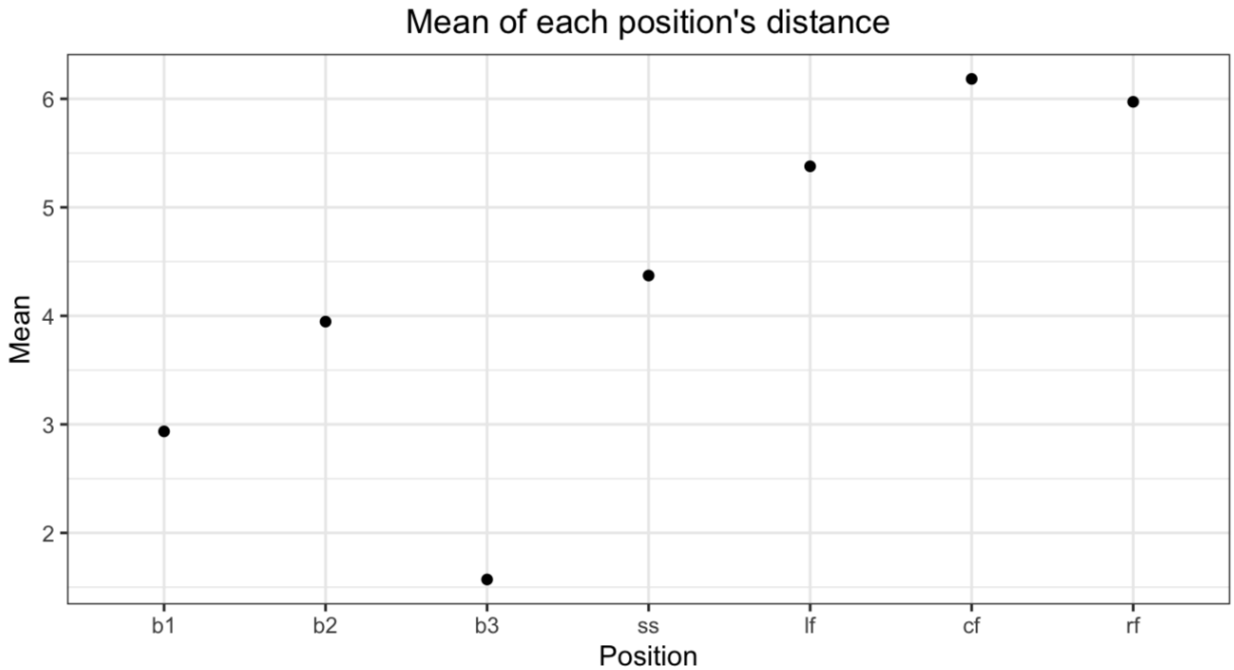

Figure 3: Mean of each position's distance moved.

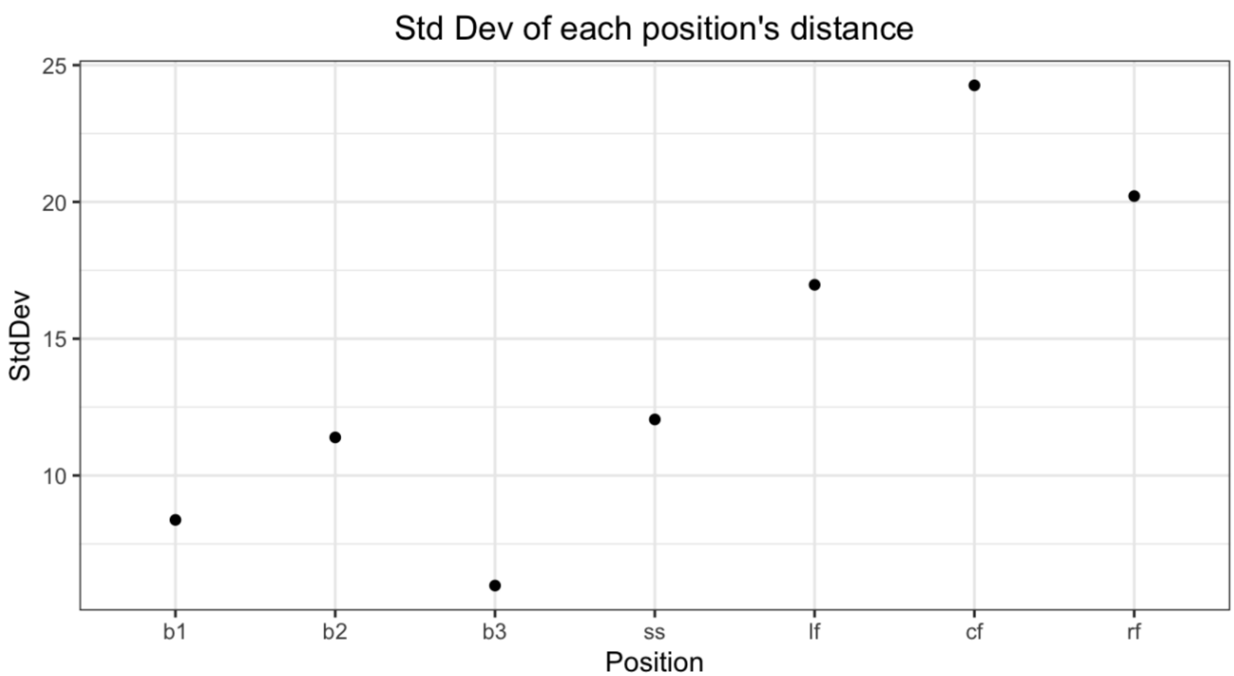

Figure 4: Standard deviation of each position's distance moved.

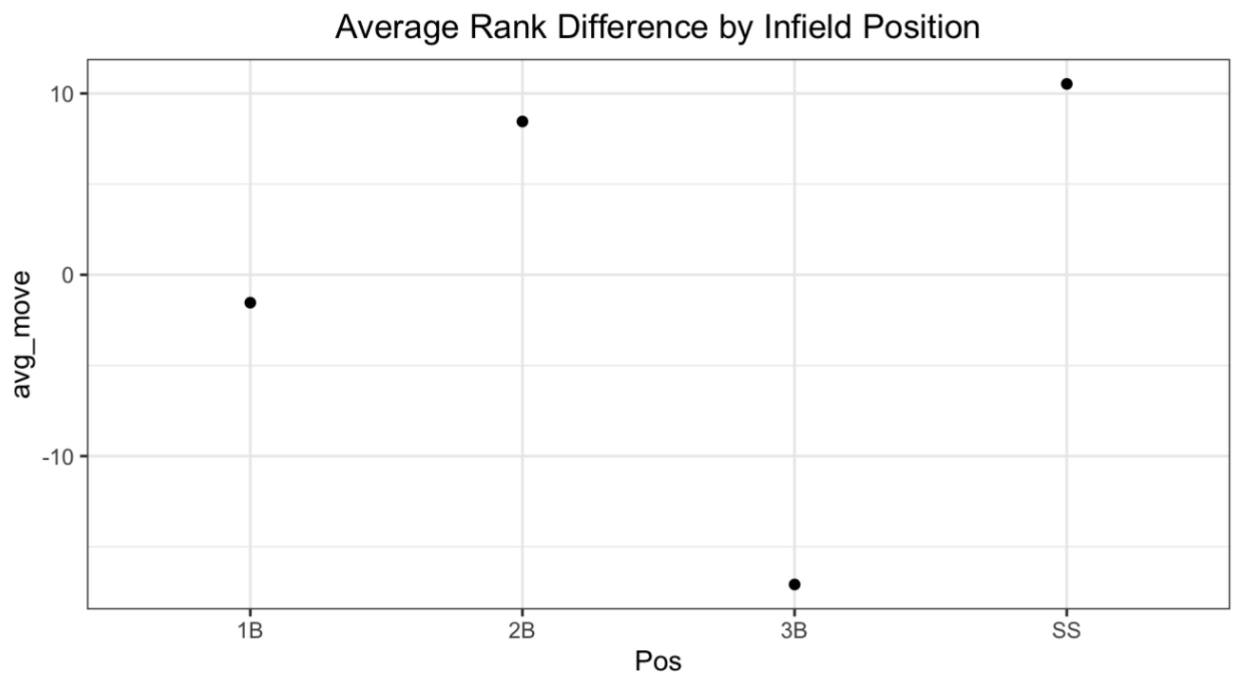

Figure 5: Average Rank Difference (before-after adjustment) of each infield position.

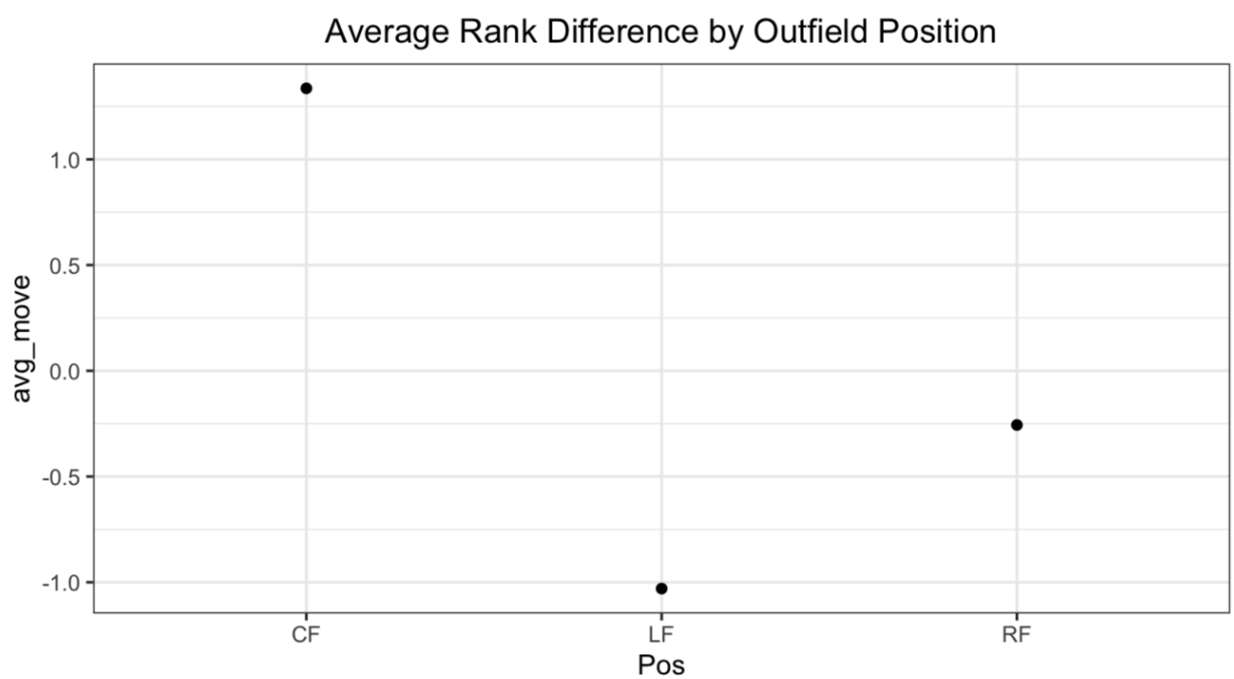

Figure 6: Average Rank Difference (before-after adjustment) of each outfield position.

Pos	Mean	Standard Deviation	Adjustment
1B	2.935567917596254	8.376442368775367	0.9156950777411831
2B	3.946800921	11.391713207335133	1.2311301518594713
3B	1.5706242303484688	5.975937207065677	0.489926623087224
SS	4.371418881601272	12.048276575904909	1.36358172080885
LF	5.378053423741786	16.97039592533667	0.9156840815205175
CF	6.1835984258350525	24.261451903090425	1.0528386758034454
RF	5.972398926027788	20.213693620228753	1.016879192927194

Figure 7: Table of Means, Standard Deviations, and Adjustments.