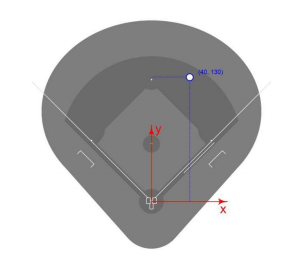
To be paired with .txt file “Hannon\_BlueJays” written in R Studio (Version 1.2.5042)

Task: Using provided *shortsopdefense.xlsx*, answer “which shortstop converted the most outs above average?”

To answer this question, I was first hoping to find a rigid, structured equation for calculating the Outs Above Average statistic. After perusing FanGraphs, mlb.com, Baseball Savant, and Tom Tango’s blogs on the subject, there was not really a formula to be found. The theme I got from the readings was that in order to calculate “Outs Above Average”, the important features are the distance a player needs to travel to get to “the intercept point”, the time to get there (how quickly the ball gets to the “intercept point”), distance from the base, and the speed of the runner.

With the given information, I determined the best way to determine the “distance” aspect of each play was to compare the players initial angle from home plate to the horizontal launch angle or “spray”. The players angle was determined using the arctan function as shown in Figure 1, utilizing the player\_x and player\_y variables as x and y. This was then converted from radians to degrees and the absolute value of the difference of the players angle and spray angle was stored as a new variable called Degrees Off. To account for how quickly the ball gets to the intercept point, a new variable was created called Degrees Off Adjusted which multiplied the Degrees Off variable by the launch\_speed. When filtering out routine plays, this multiplication also filtered out bunts due to their low velocity, which were initially problematic in thinking about if the spray angle was a fair way to determine a player’s ability to get to a ball. These also could have been filtered out using the is\_bunt variable.

Figure

Y

X

After the creation of these two variables, the entirety of the data set was filtered to just show plays in which a shortstop was given out credit. Then statistics were taken on the Degrees Off and Degrees Off Adjusted for this remaining data to determine a baseline of distance traveled and speed for an “average” out.

Chart, box and whisker chart

Description automatically generatedChart, scatter chart

Description automatically generated Figure 2 shows boxplots for the two baseline statistics created to quantify an “average” out. It was also calculated that the mean Degrees Off for a SS Out was 4.61 and the mean Degrees Off Adjusted was 393.82. To set a threshold for “outs above average”, it was decided to use the 3rd Quartile as a cutoff rather than the mean or median. This is from a “boxplot mentality” reasoning that the middle 50% of outs are contained between the 1st and 3rd quartiles and it is arguable that these plays are all roughly average, and the lowest quartile could be seen as “easy outs”.

Figure

So, the data was filtered to only plays in which the shortstop was 7 or more degrees off from the spray angle and the degrees off adjusted was above 572. This removed 4,550 observations from the dataset. These removed observations should have been fieldable by the average shortstop. The remaining data was then filtered to just the plays in which the shortstop recorded the out, meaning the shortstop needed to execute an above average play. There were 762 observations in this dataset. A table was created to tally how many of these above average outs were recorded by each player.

Lastly, errors on routine plays needed to be subtracted from the number of above average plays for each player. This was done by filtering the original shortstop defense dataset by event type being “field\_error” and position equal to 6. It was considered that the plays with Degrees Off above 7, etc. should be filtered out but due to the way the MLB only credits an error if the play was executable, these were not factored in. Finally, using a left join operation, the results were combined and the number of errors for each player was subtracted from the number of above average plays for each to give us an estimate of their OAA. See Figure 3 for the leaderboard.

A picture containing table

Description automatically generated

Figure -OAA Leaderboard

In addition to the given dataset, a few things would help answer this question more effectively. First, an agreed upon method of calculating OAA on all platforms would be ideal. Second, knowing information about the runner would help make these numbers more accurate. Even if a player does not have to move a full 7 degrees to get to a ball, having to move fast enough to throw out particularly fast runners may qualify as an above average play and thus change the final leaderboard.

Other thoughts and findings:

Table

Description automatically generated Having the OAA next to the number of opportunities definitely shows that the more opportunities a player has, the more likely they are to have high OAA. And while it is likely that many of the opportunities a player sees may not qualify for an OAA because it is a “routine ground ball” or something like that, players that have a high number of OAA in limited opportunity should definitely be valued for their impressive ability. For example, playerid 9074, ranked 3rd, had 19 OAA in just 126 opportunities, which means on 15% of his opportunities, he was able to make an above average play to get an out. Compare this to the top ranked player who only made above average plays for outs on 12.7% of his opportunities. To Account for this, a new table was made showing the percentage of OAA’s as well as number of errors. For this, any player with less than 10 opportunities was filtered out. Figure 4 shows a ranking by percentage of total opportunities. We see that player 167960 fell to number 9 when looking at average and we can also now see that he committed 15 errors which divides to be an error on 7.6% of balls hit to him.

As a final usage of this data, the player x and y coordinates were used to calculate what a generalized ideal location would be for a shortstop positioning to Text

Description automatically generatedcommit an out. The mean x and y were (-32.85, 141.09) as seen in the summary statistics in Figure 5. However it appears both the x and y are very spread. The accompanying boxplots also show that x (in blue) and y (in red) have a fairly large spread, even ignoring outlier that may Chart, box and whisker chart

Description automatically generatedbe due to defensive shifts. The spread of just the whiskers for x shows a spread of 82 feet while the spread of whiskers for y shows a spread of 40 feet, which combines to a large area for a shortstop to cover. Perhaps in a future project focusing on defensive shifts, one might take this further to factor in runner location and batter handedness and tendencies to calculate ideal starting position for shortstops.

Figure