In this project, I was to project the “true” average speed-off-bat (on the scale of system A, system B, or some combination of the two systems) for each batter in the following season. I did so using linear regression and other methods in R to come up with a complete projection using the averages of each hit type for each batter recorded through each system.

Through looking at the linear models that I would possibly use to predict the launch angle and exit velocity, I found that the best place to start was finding the launch angles. Finding the launch angle of system A using the data from system B could be done with an accuracy of 94.9%, and finding B using system A could be done with an accuracy of 93.8%. However, predicting the exit velocities of both A and B could only be done with 75.6% accuracy and 82.5% accuracy respectively. When adding the launch angle of system A into the model to find exit velocity A the accuracy improved to 77.8% and adding the launch angle of system B into the model to find exit velocity B improved the accuracy of the model to 84.1%, so I decided to predict the launch angle of each first.

I provided code in R to predict the launch angle of system B using the data in system A if provided. I then iterated through a new copy of the battedBallData dataset and added the predicted values for each NA in the vangle\_B column, checking to make sure there were no missing values left. I repeated this for speed\_B, vangle\_A, and speed\_A to have a complete dataset to make projections of each batters exit velocity with. After, I made sure to check the frequency of each batter ID to see if some may be underrepresented, and there were many cases of few hits recorded.

I then created matrices of the average exit velocity of each batter in system A by hit type and then system B by hit type to provide better projections of true exit velocity. The creation of a second matrix allowed me to combine the two prior and find the average exit velocity by hit type for each batter using both systems A and B together. All that was left to do was to find the total average of each batter’s exit velocity by taking the exit velocities of their hit types. In doing so, I created a matrix that provides each batter and their projected exit velocity for the next season at hand.

If I had more time, I would have liked to have done a cluster analysis of each hit type and the launch angles recorded to better group the exit velocities recorded. I would then have used the averages of each of these clusters to project more accurately a batters exit velocity.