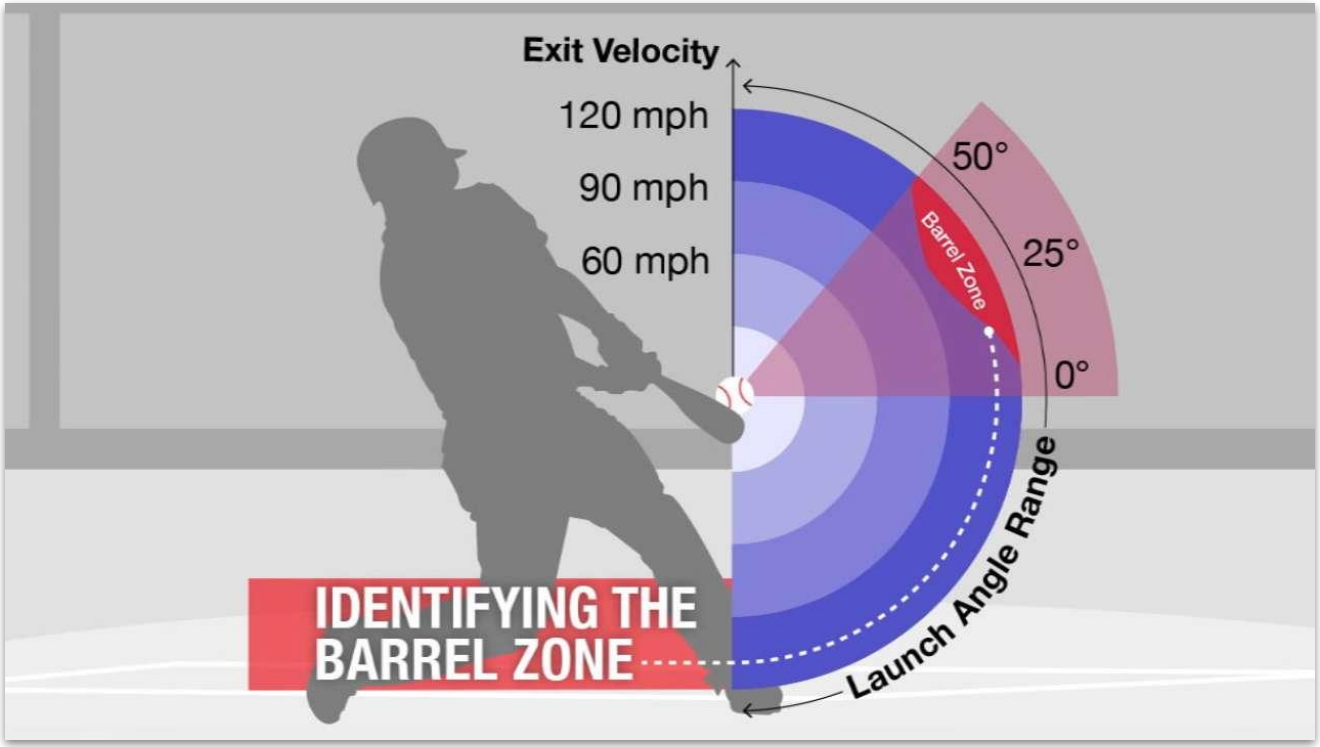
MLB Home Run Exit Velocities

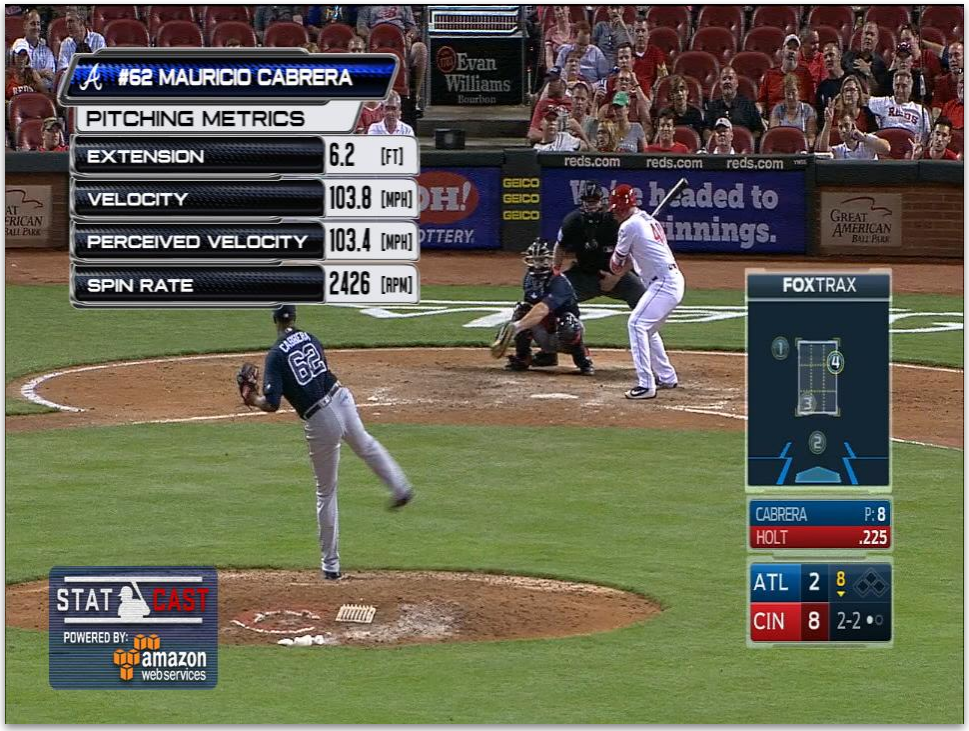


Jon Nelson

2017 Record Breaking Year for total Home Runs

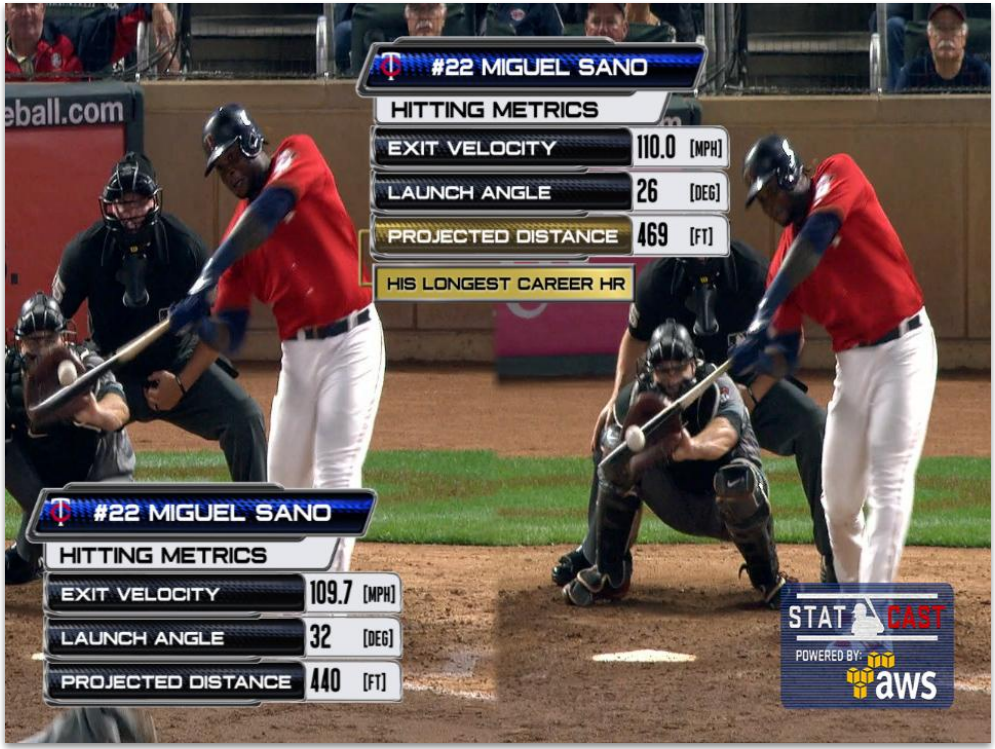
One of the most recent debates in Major League Baseball is focused on why more home runs were hit in the 2017 season than any other season in the leagues history.

1. Baseball Savant:
   a. Extracted each home run observation from the 2015, 2016 and 2017 seasons
      i. 16,626 observations with 89 features
2. Research Articles on physical baseball statistics
   a. Obtain the data from baseballs sampled from the 2015, 2016 and 2017 seasons
      i. 36 observations of baseballs, with 7 features
3. Baseball Savant Web Scrape
   a. Scraped each players personal stats:
      i. 750 players



Problem Statement

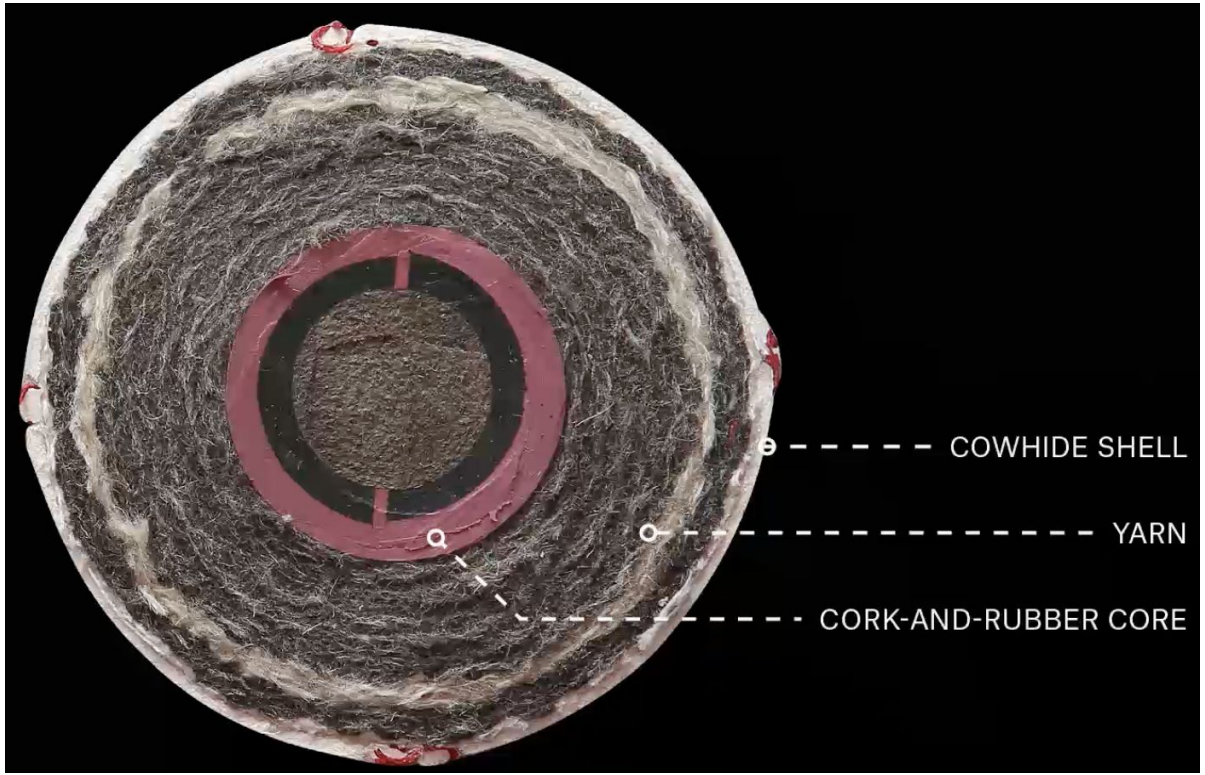
Using the physical data about baseballs, the personal stats from each player and the game statistics from each pitch that was hit for a home run I will draw a conclusion about the most influential features contributing to a batter's home run exit velocity.



Background on the baseball

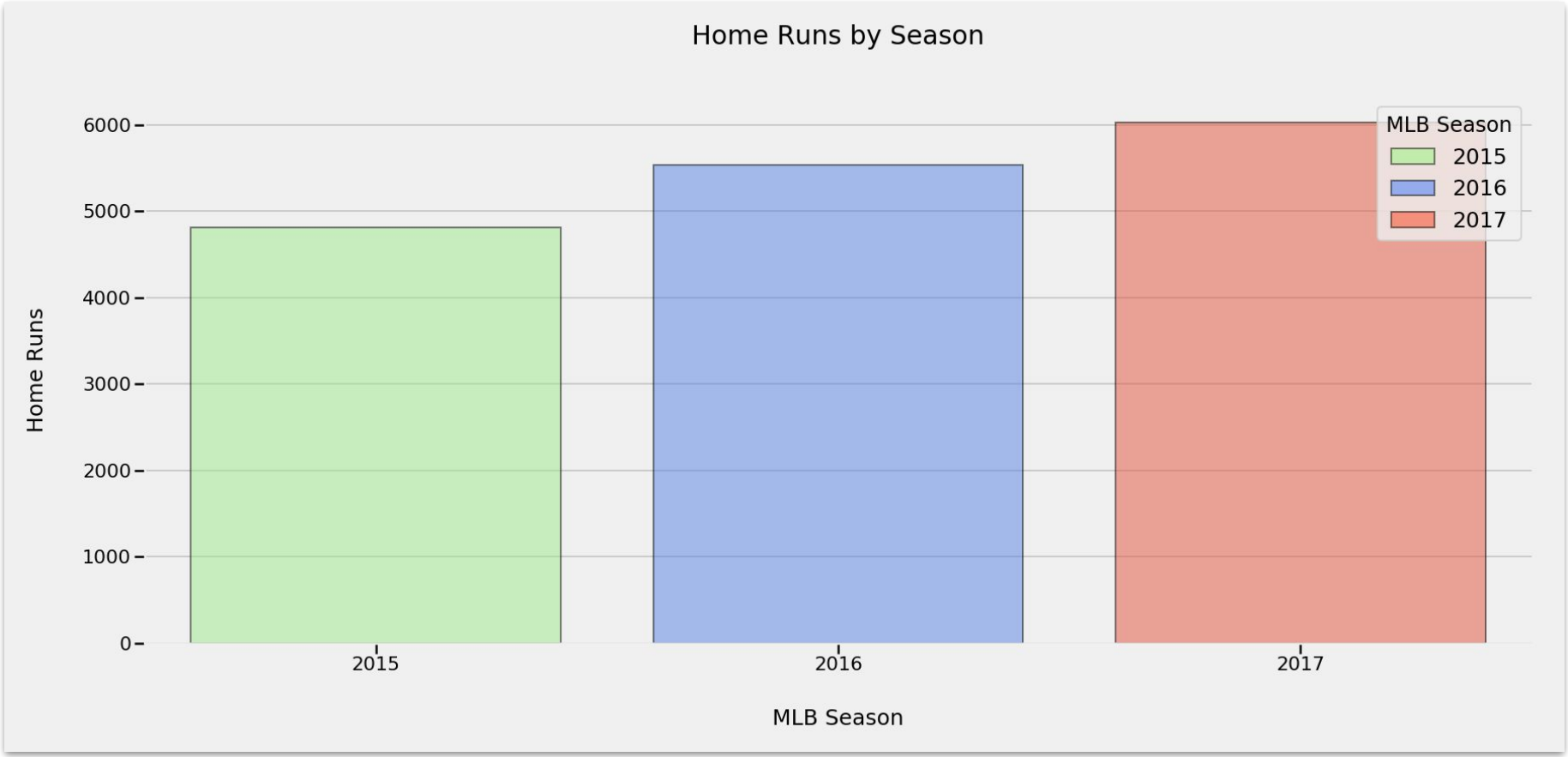
Important Features:

- **Weight**: the weight of the baseballs in ounces.
- **Seam Height**`: seam height was defined as the average radial distance from the seam to the ear.
- **CCOR**: cylindrical coefficient of restitution (ccor) is the measurement of the "bounciness" of the baseball and is the core ingredient of "the pill" the middle rubber of the baseball.
- **DS**: Dynamic Stiffness is a measure of a ball's hardness. Its measurement is conducted to represent bat-ball impact forces.

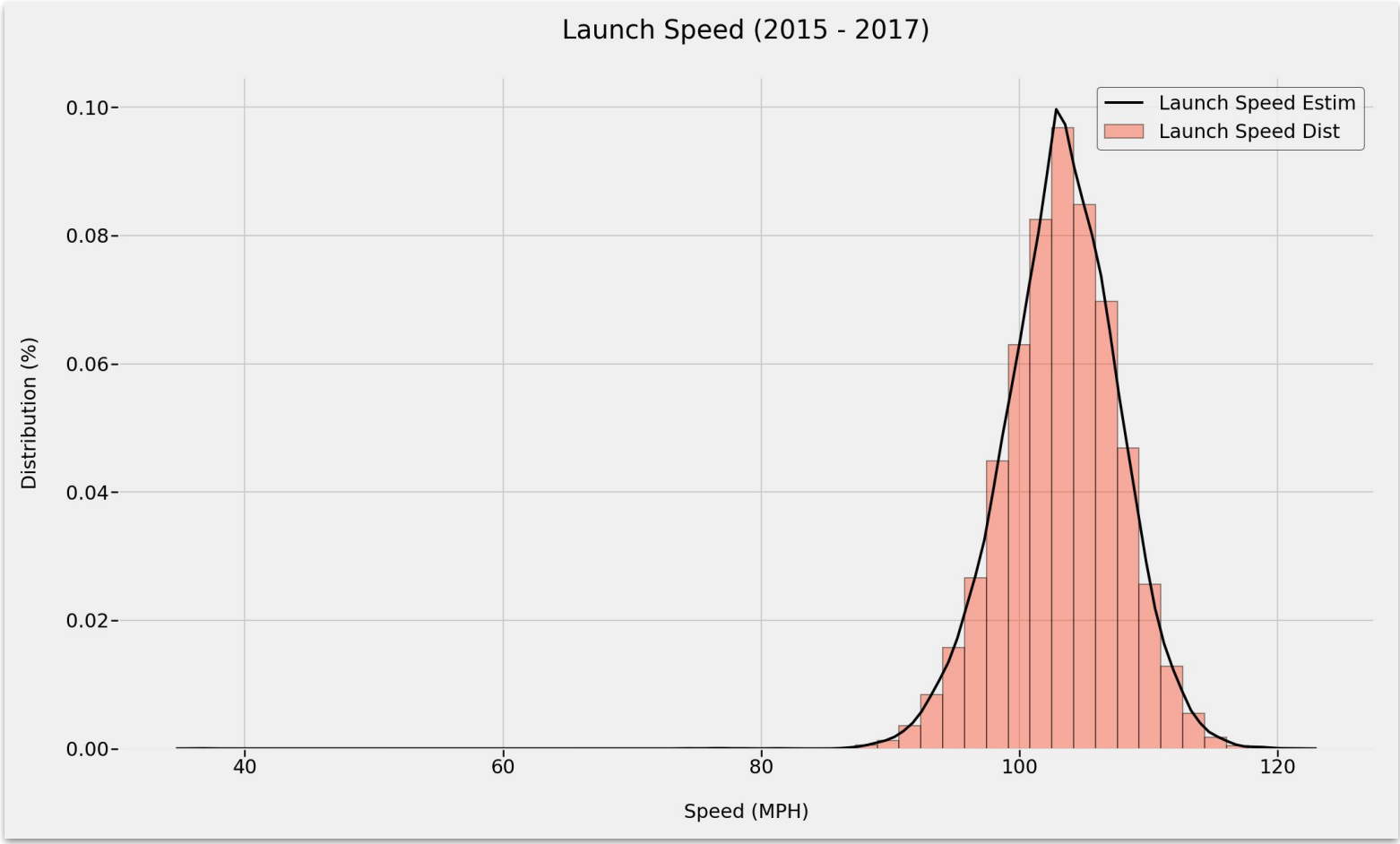


Explore the Data

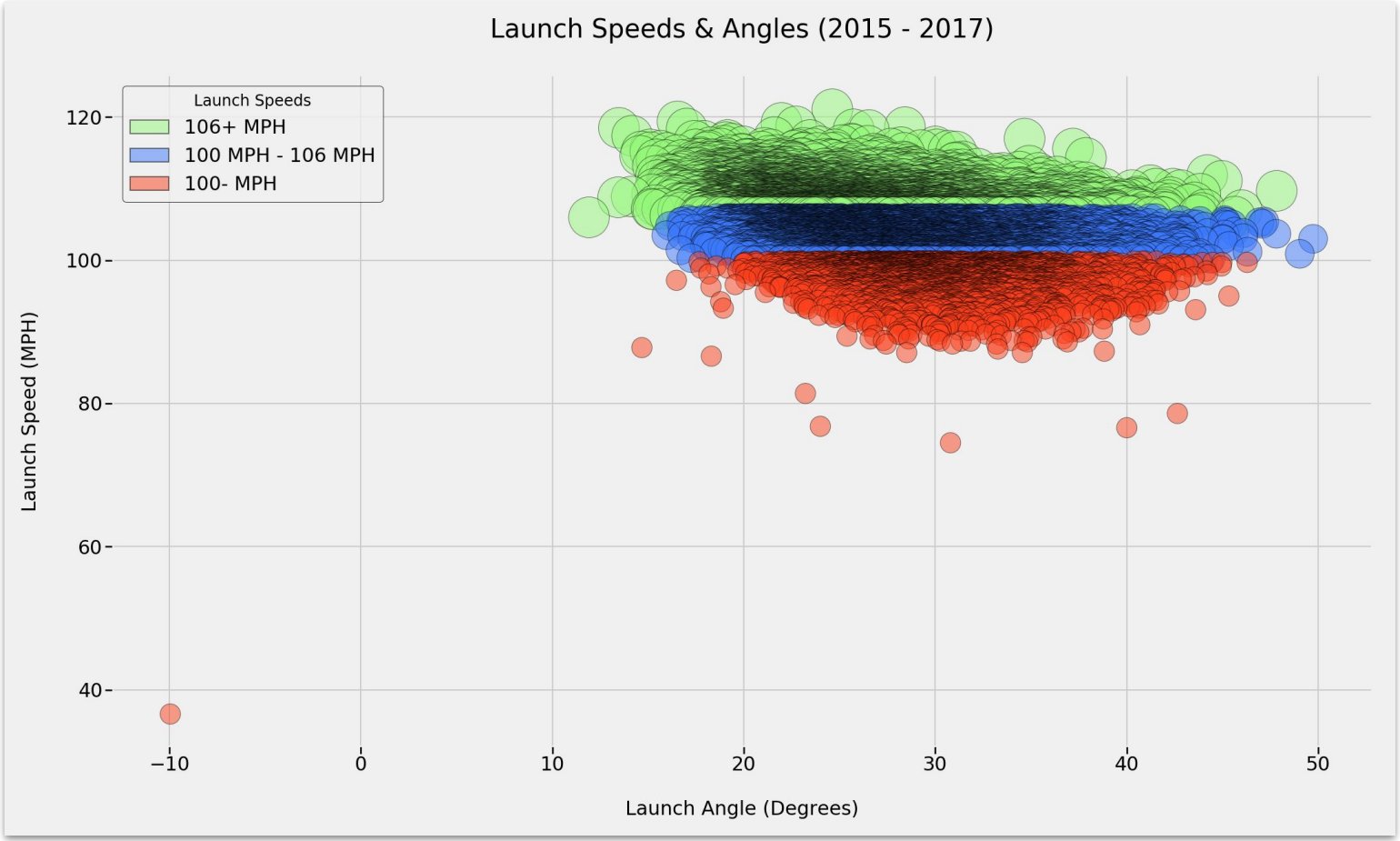
Home Runs by Season



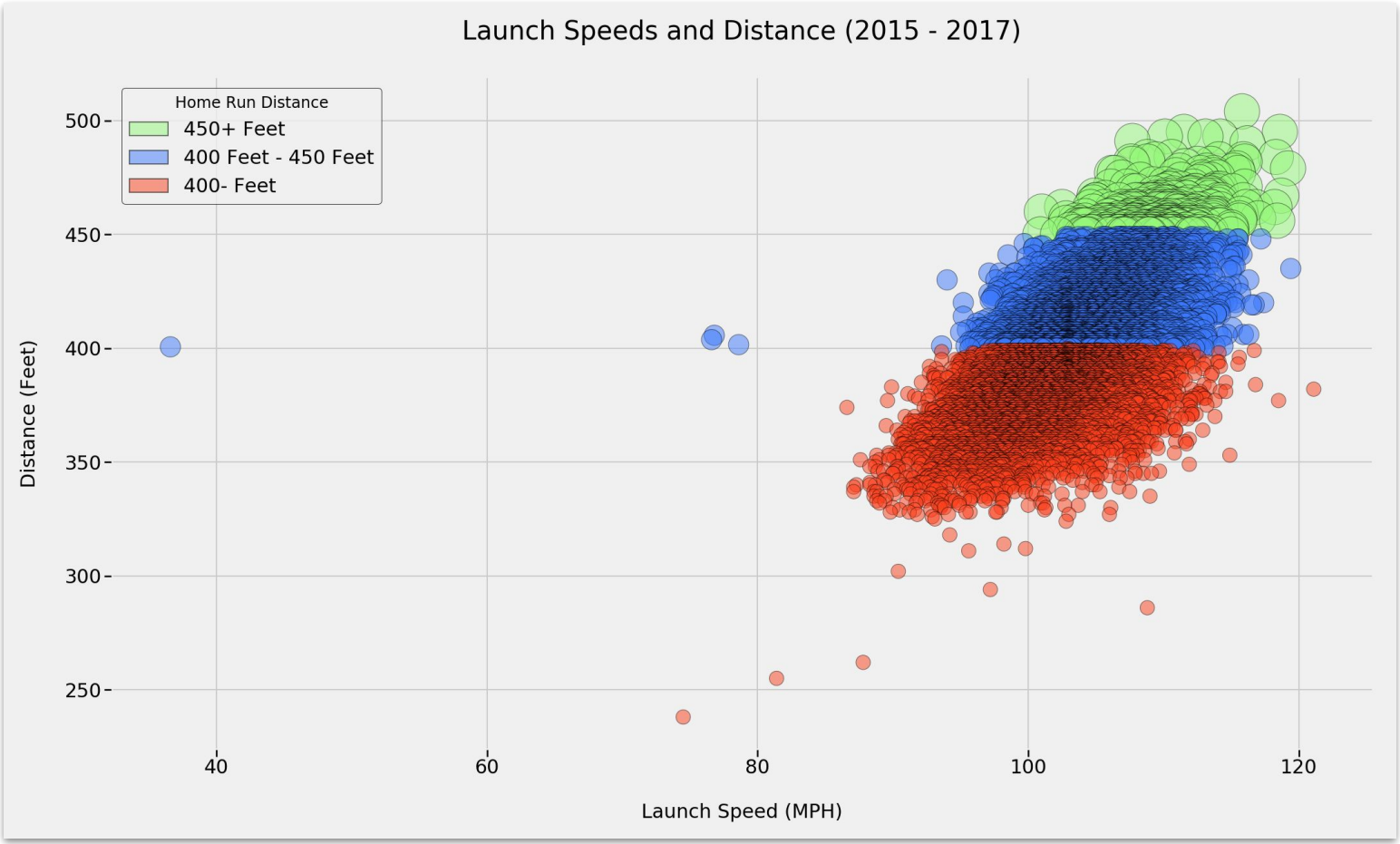
Distribution of Exit Velocities



Launch Speeds (Velocity) vs Launch Angles

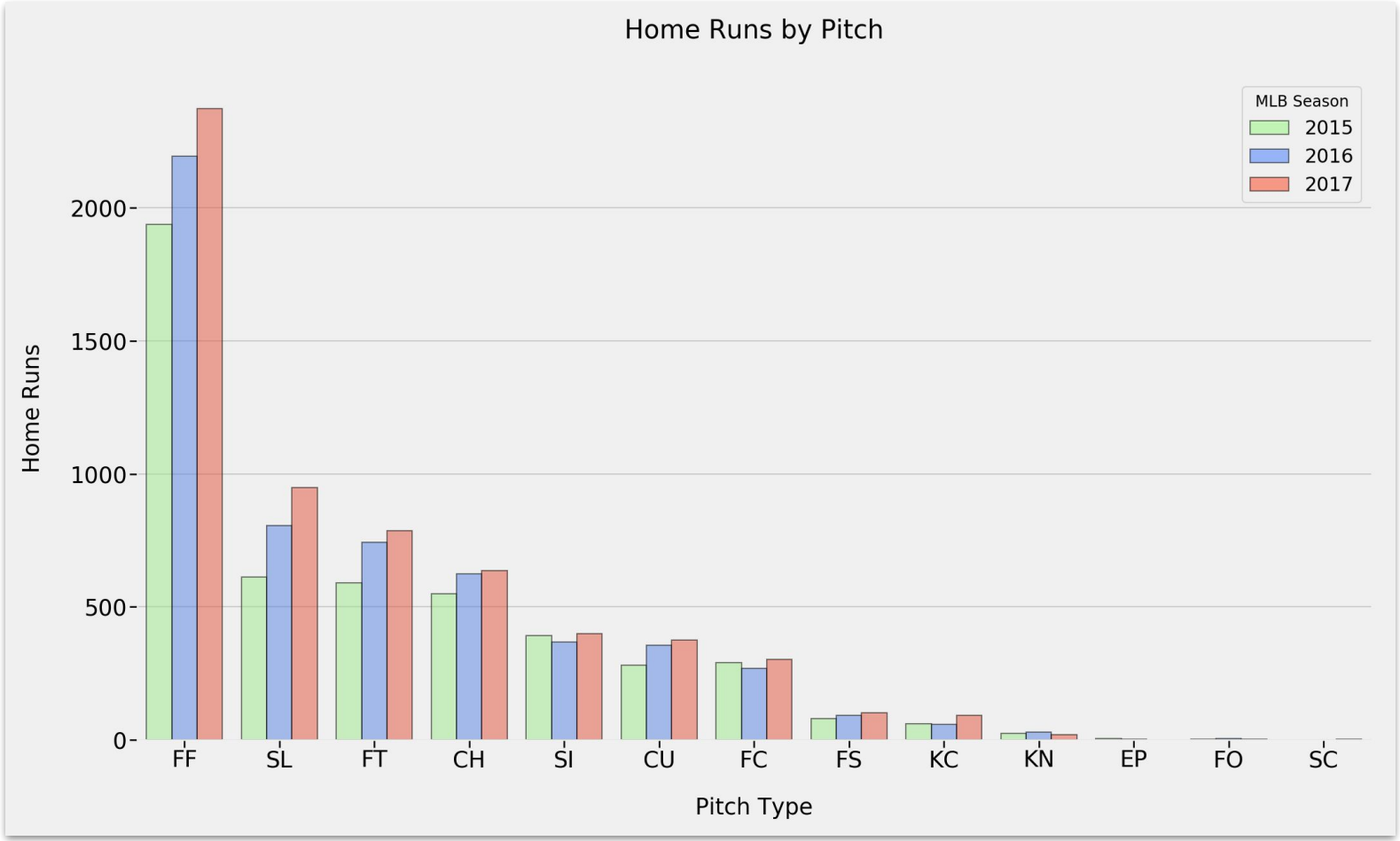


Launch Speeds (Velocity) vs Home Run Distance

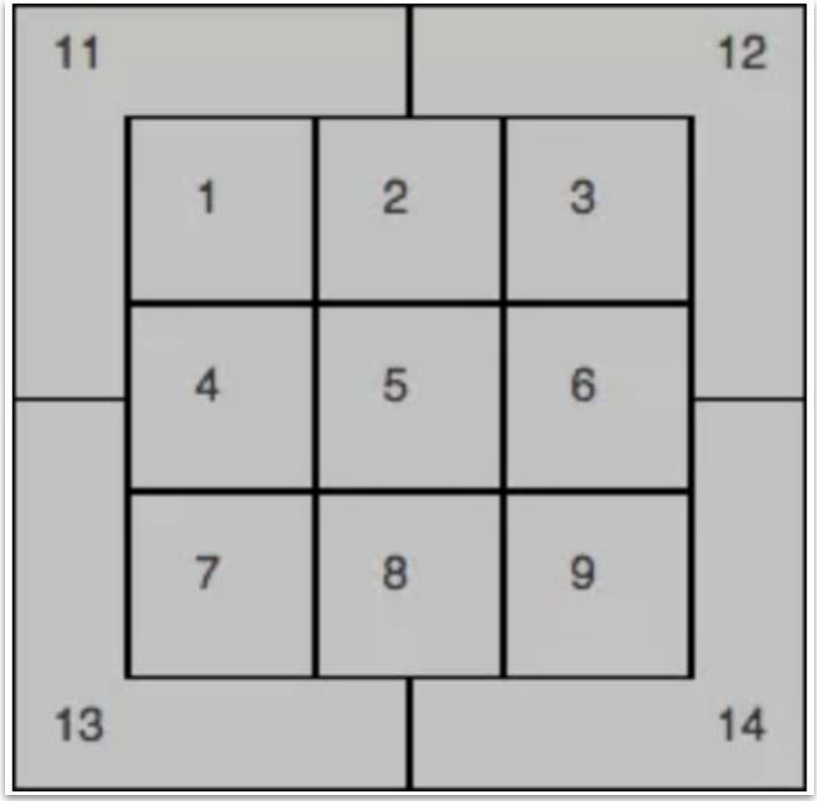


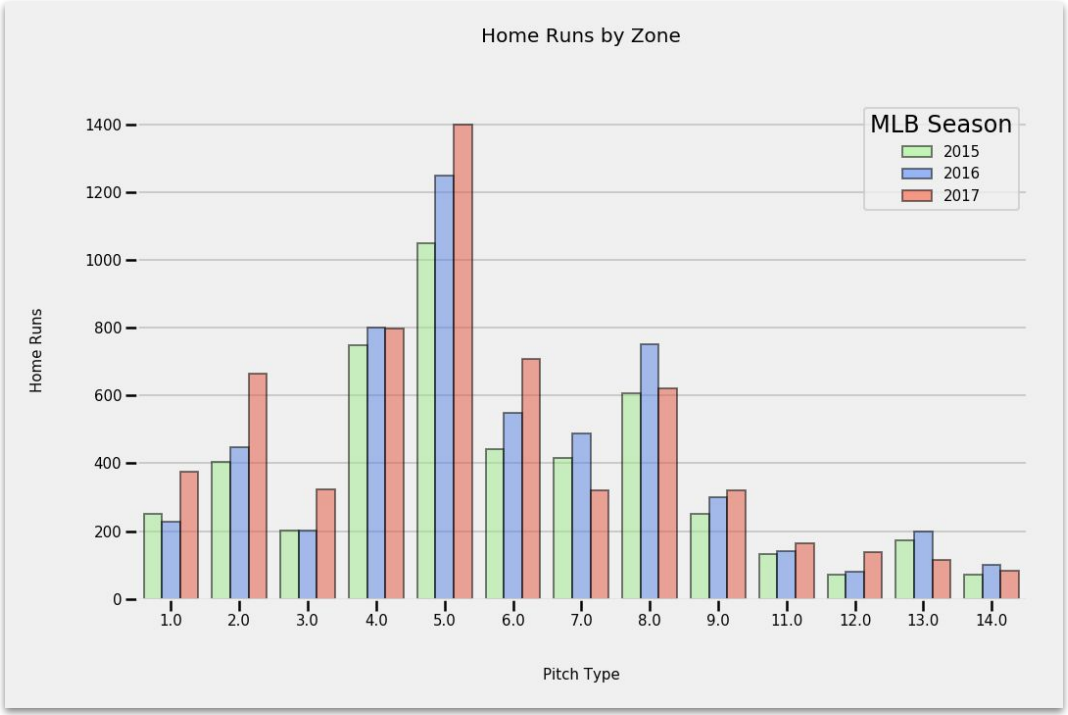
Home Runs

Home Runs by Pitch Type

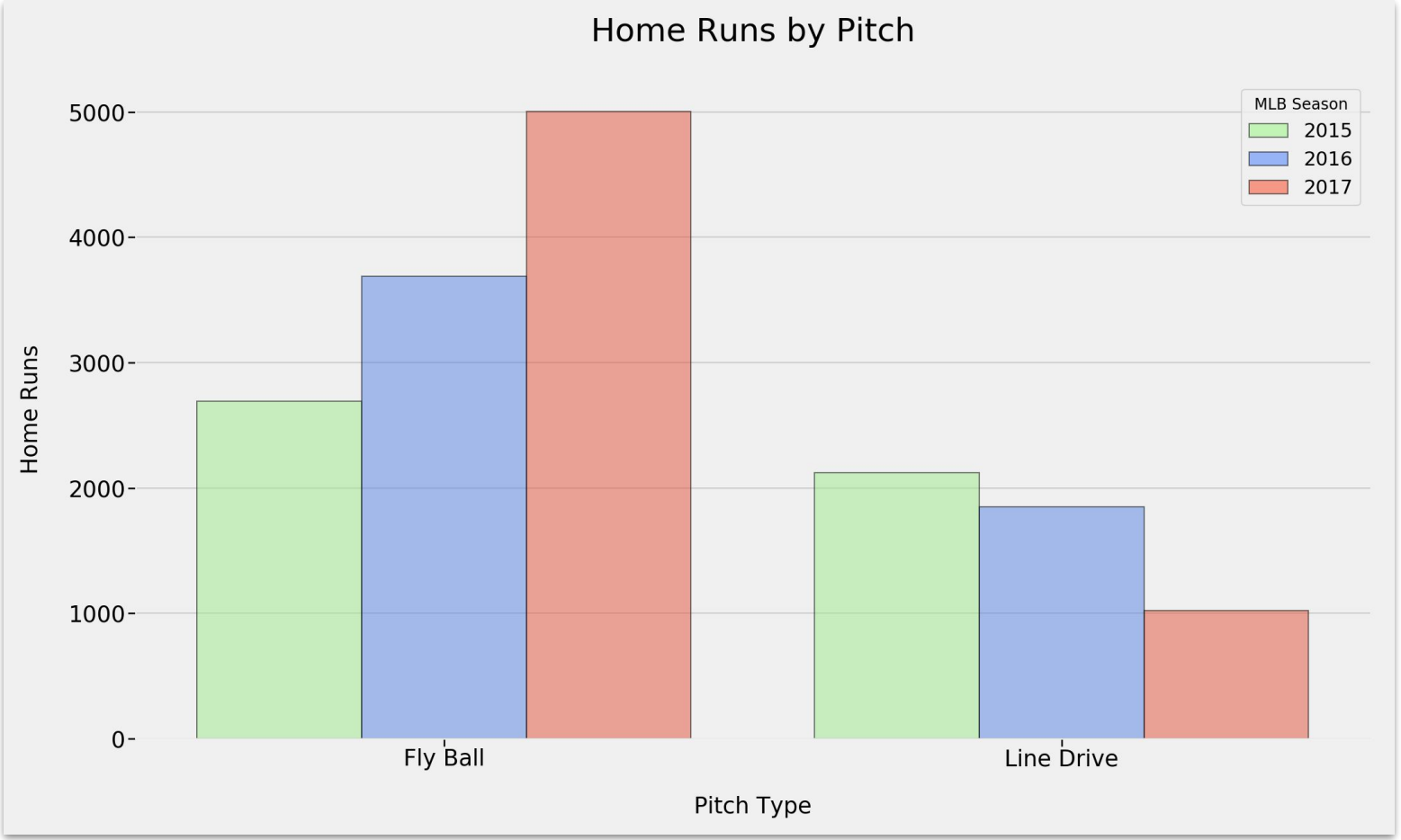


Home Runs by Zone





Home Runs by Batted Ball Type (more on this later)



Modeling

Gradient Boost Regressor

Boosting:

- When boosting a model the model is building multiple simple models and learning from these models to be more approximate when predicting. These simple models are referred to as weak model or weak learners.

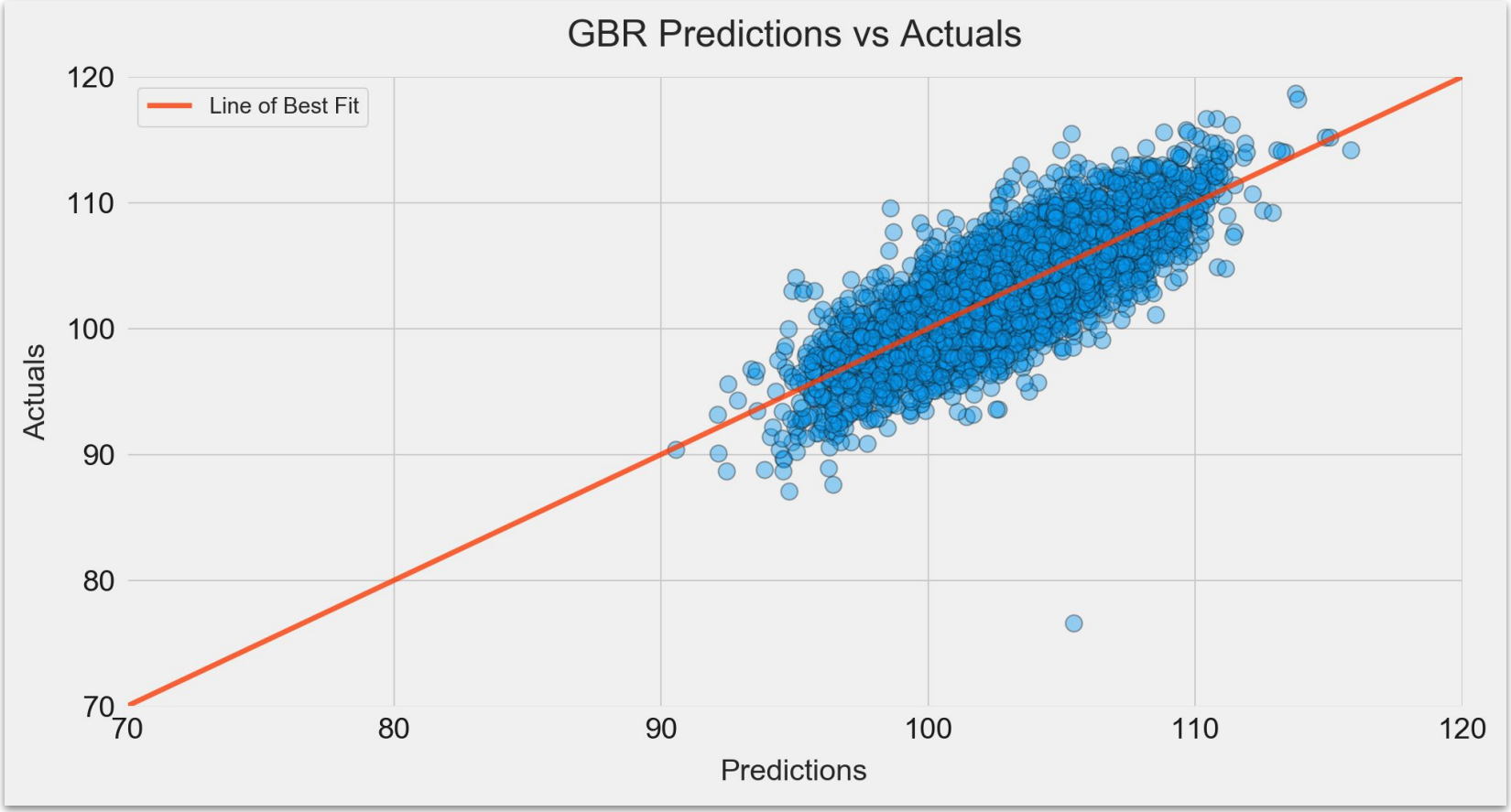
Gradient Boosting:

- Creates weak learner models and sequentially trains itself on the residuals or errors in order to give more importance to the less accurate predictions and once completed uses what was learned from these predictions to combine with the strong predictions to have a better overall approximation.

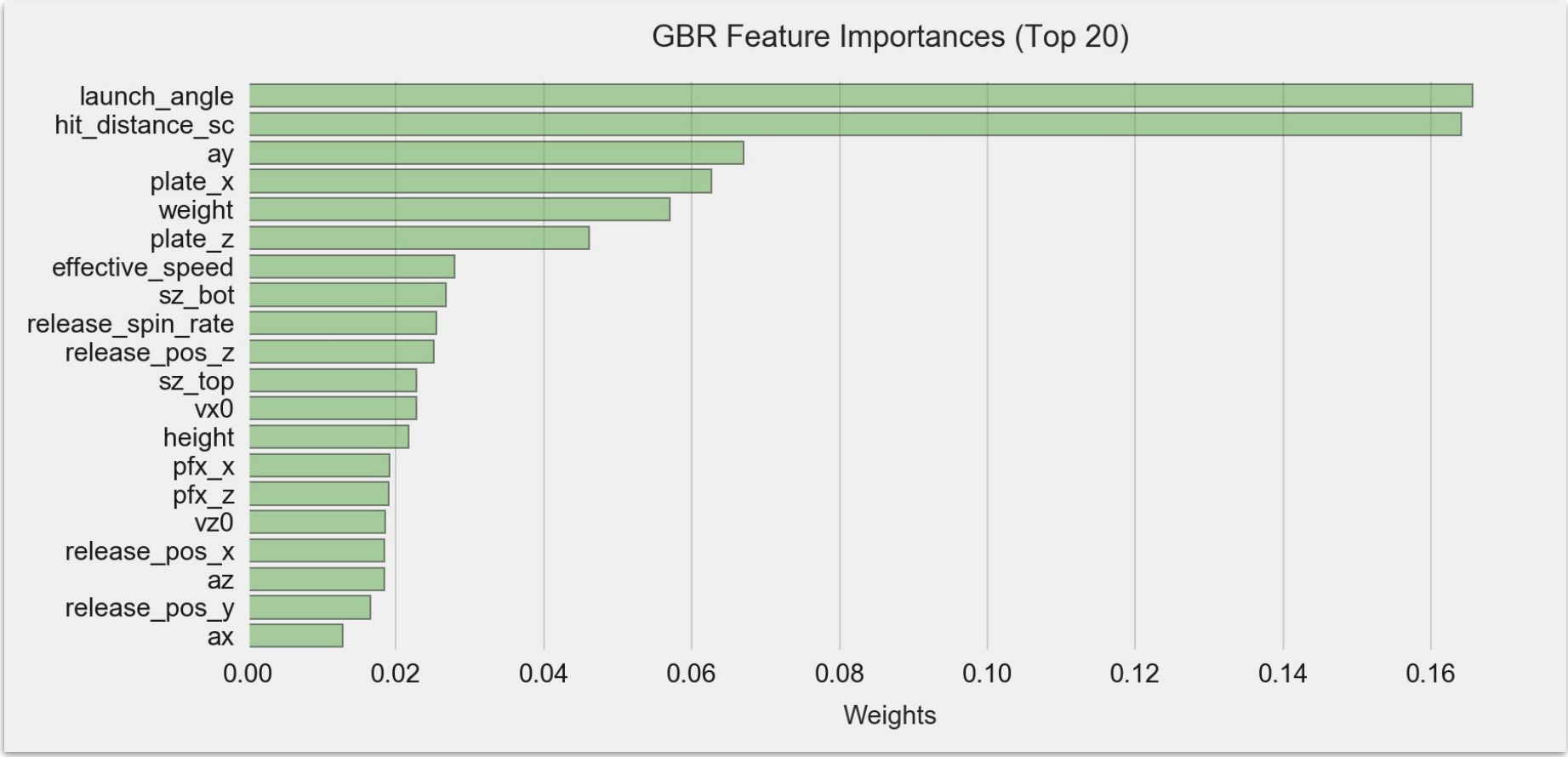
Gradient Boost Scores

Model Scores	
TRAIN SCORE	**TEST SCORE**
0.68038	0.61657

Gradient Boost Predictions



Gradient Boost Feature Importances



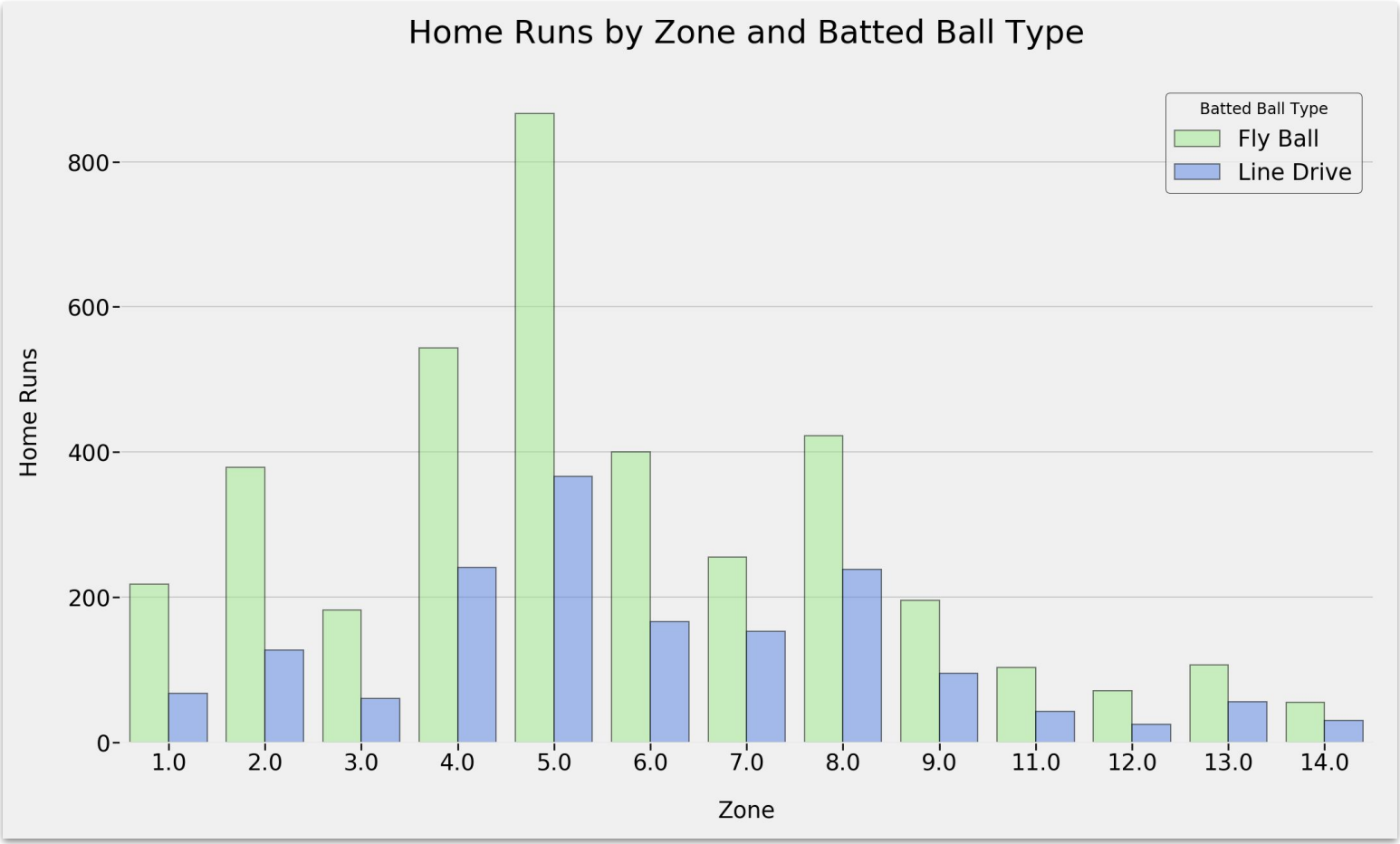
Conclusions and Next Steps

Conclusion

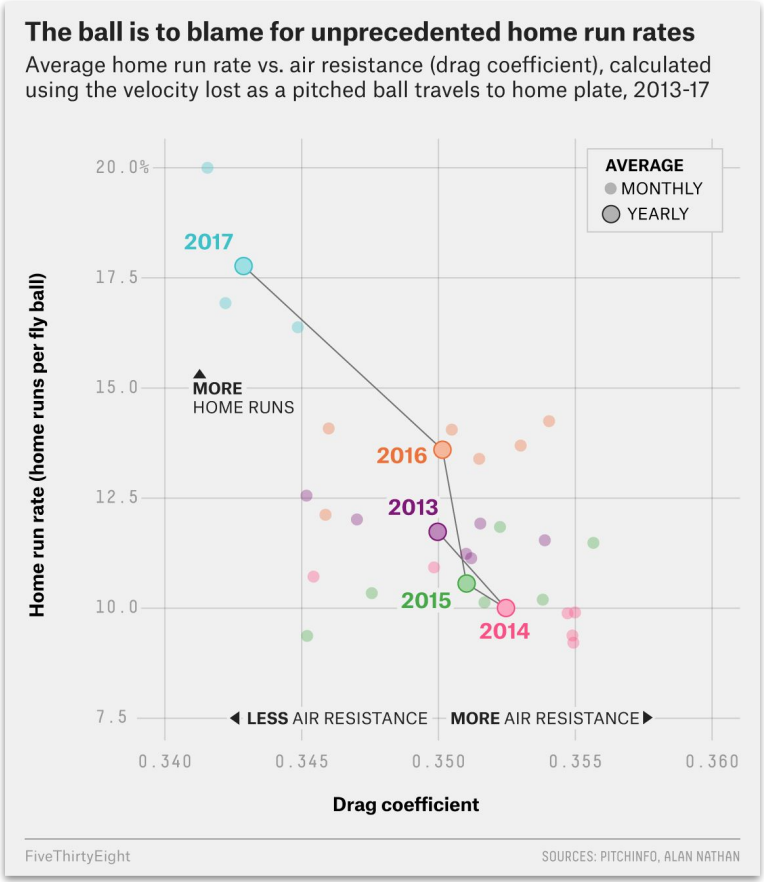
1. Pitching statistics are the most influential in predicting the exit velocity of a home run
2. A batters height and weight are also very influential in predicting the exit velocity of a home run.
3. The baseballs are not influencing a batters home run exit velocity.

However, I am not completely convinced that the baseballs are not more influential based on the changes that were found ….

Next Steps: The Fly Ball Home Run



The Drag Coefficient (Not my plot below)



FiveThirtyEight did a study on the drag coefficient on the baseball measured against air resistance.

This required:

- Temperatures at time of HR
- Ballpark dimensions

"MLB-wide average drag coefficient dropping by about 0.01 from 2015 to 2017.[4] That might not sound like much, but Nathan's calculations show that even a change that small can add up to 5 feet of distance on a well-hit fly ball,[5] which in turn would be enough to make 10 to 15 percent more balls leave the yard in a given season."

https://fivethirtyeight.com/features/in-mlbs-new-home-run-era-its-the-baseballs-that-are-juicing/