Exploring and Modeling Baseball Hitting Probabilites on Balls in Play

Alex Jackson

2022-12-16

Introduction

Baseball is the ultimate statistical sport. Everything in the game seems to fit perfectly in a box and is automatically quantified. Following 'Moneyball,' teams across Major League Baseball began defining the basis on their team building strategy on statistics and models. Through the abundance of statistics in the game, we are able to beat the 'eye test,' which is a revelation that is still fighting its way into other sports. Recently, further advancements in pitch tracking technology have allowed baseball statisticians to find new ways to quantify and simplify complex interactions between the bat and the ball on any given pitch.

One aspect of hitting that teams universally care about is putting the ball in play. It is understood that, generally, good things happen when the ball is put into play. Being put into play is classified as any time the ball is active off a pitch. For the purpose of this study, this means when the batter makes contact with the ball and the ball goes into the field of play, the ball is 'in play.' These balls in play can either result in a hit (a single, double, triple or home run) or an out.

The analysis in this paper seeks to delve into the aspects of a single pitch that benefit the chances of that ball being put into play for a hit by creating a model that attempts to predict the probability of a ball in play becoming a hit. These aspects include game situation variables, as well as pitch characteristics and hit characteristics. The model does not explore any prior information on the batter or pitcher to generalize the resulting information. This modeling is important for two reasons. First, players can better understand aspects of the game they need to be focusing on to improve. Batters always want to increase there chances of getting hits and pitchers always want to find ways to get outs. Additionally, team general managers and coaches are always looking for ways to improve their rosters and an analysis like this may be able to help them look over their team and find the areas that need improvement by seeing which players are good or bad at certain aspects. While this paper focuses on the inference side of modeling, in theory, with enough data and variables, it could have predictive applications as well.

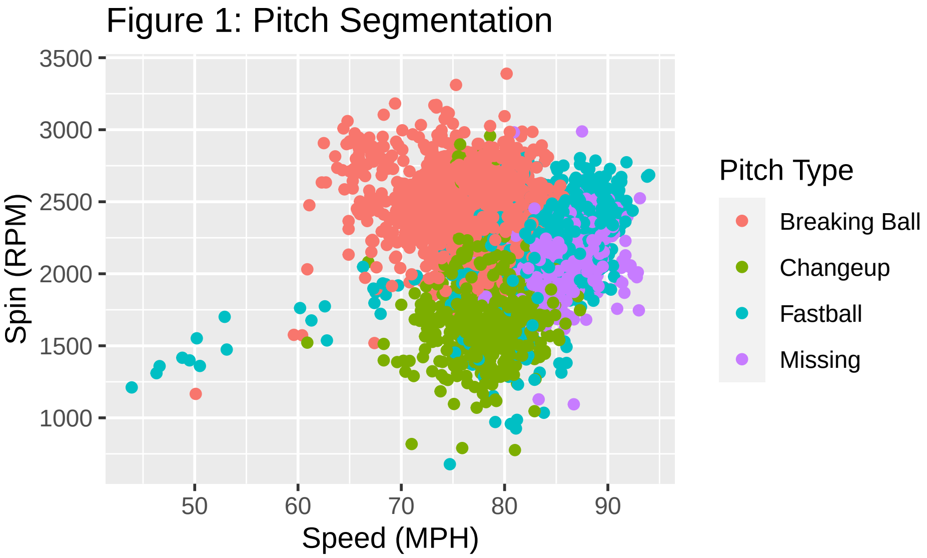
We believe there will be statistically significant evidence of certain aspects affecting the probability of generating hits. Mainly, we believe that bat-contact statistics will prove to have far larger effects than game state or pitching statistics.

The data is collected from the `baseballr` package using their pre-built web scrapping functions of play-by-play data that can reach any game from any year that has recorded play-by-play data. The package scrapes from baseball-reference.com and baseballsavant.com to get the information as well as advanced statistics. However, due to the difficulty of data collection and sheer number of balls put in play in 2022 across the league. We have chosen to look at only balls in play by the Chicago Cubs last season. Due to this, there cannot be sweeping generalizations made about the game of baseball and the league as a whole, but rather we focus on the applicability of the results to the Cubs only.

Exploratory Data Analysis

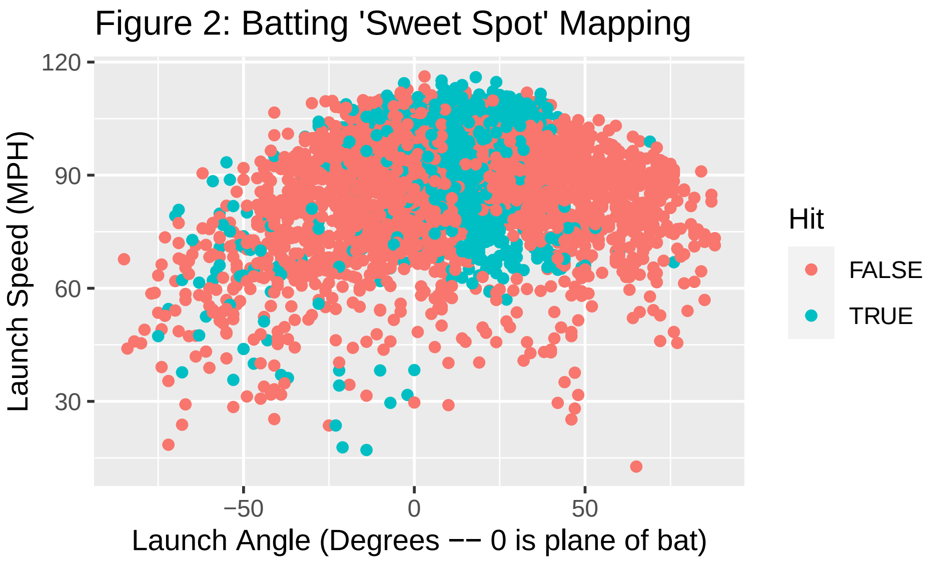
Our exploratory data analysis focuses both on better understanding the data, as well as uncovering potential relationships in the variables for possible interaction terms later on in our model. In exploring the pitching-related variables, it is apparent in Figure 1 that the combination of spin and pitch speed reveal the grouping of pitch types. This makes sense because speed and spin often dictate the classification of pitches. However, different pitchers throw each pitch type differently and thus groupings span large areas.

In Appendix Figure A, we also explore the breakdown of pitch type mixing by batter-pitcher handedness, ball and strike counts. The most noticeable attribute there is the flexible use of changeups and breaking balls. Pitchers on the same side as the batter move away from changeups and lean more heavily on breaking balls. Additionally, pitchers love to lean more on off-speed pitches in “safer” counts (i.e. low balls, high strikes) and turn more heavily to fastballs in tougher counts (high balls, low strikes).



When it comes to batting, the center of our EDA focuses on visualizing evidence of a “sweet spot.” It is immediately apparent from Figure 2 that a line of hits in the 12-24 degree range is almost invariable to the speed at which they are hit. Additionally, as balls reach a certain exit velocity, the hits spread and angle seems to matter less.

In Appendix Figure B, we relate distances to hits. In those graphs, there are two very distinct and interesting shapes. As we explore launch angle, it is apparent there is a very strict distance each launch angle travels. In the 25-50 degree range there is a bit more variability, but we also see a dead-spot in distance, where we can assume the outfields catch most balls. When exploring launch speed, there is a clear maximum distance associated with each speed, however there is much more randomness to whether each ball was caught. We do see the similar dead-spot in distance at around 300 feet.



Methodology

Variable Selection and Alteration

The main method of variable selection is based on domain baseball knowledge. We avoided using any type of automatic selection in order to prevent maximizing or minimizing an arbitrary statistic. Our approach to variable selection involved using variables from the three different aspects of an at-bat: game state, pitching, and hitting.

Target Variable:

- Hit: Created indicator if a hit was recorded regardless of the type of hit (TRUE for yes, FALSE for no).

Game State:

- Ball Count: The number of balls in the at-bat at time of the pitch. Modeled as a factor with levels 0, 1, 2 and 3. Selected because the count of the at-bat may dictate a batter's approach to the next pitch.
- Strike Count: The number of strikes in the at-bat at time of the pitch. Modeled as a factor with levels 0, 1 and 2. Selected because the count of the at-bat may dictate a batter's approach to the next pitch.
- Out Count: The number of outs in the inning at time of the pitch. Modeled as a factor with levels 0, 1 and 2. Selected because the outs going into the at-bat may dictate a batter's approach to the at-bat.
- Is Home Team: Created, set to TRUE if the better is on the home team and FALSE if on the away team. Selected for investigation into home field advantage claims, as seen in a 2013 Bleacher Report Article [https://bleacherreport.com/articles/1803416-is-home-field-advantage-as-important-in-baseball-as-other-major-sports].
- Pitcher-Batter Same Handed: Created, set to TRUE if the pitcher and batter are both righties or lefties, FALSE if they are different. Selected because throwing from the same side as a hitter is generally considered an advantage for a pitcher.

Pitching:

- Pitch Type: Created, the kind of pitch that was thrown. Grouped specific pitch types together into Fastball, Breaking Ball, Changeup or Missing (see missing data paragraph) to simplify the output and account for certain pitchers only throwing certain types of pitches. Selected because pitch type can dictate how a batter swings and makes contact with a ball.

- Pitch Speed: The speed of the pitch in MPH. Selected because certain pitch speeds may be harder to hit, especially when combined with pitch type.
- Spin Rate: The spin rate of the pitch in RPM. Selected because certain pitch spins may cause more movement, especially when combined with pitch_type.
- In Strike Zone: Created variable to indicate whether the pitch was in the strike zone (TRUE for yes, FALSE for no). Created from the given `pitchData.zone` variable which provides the numbered zone in which the ball passed through (1-13, 1-9 is theoretical strike zone). Selected because pitches in the zone are, in theory, more likely to be hittable than pitches outside the zone.

Hitting:

- Launch Speed: Exit velocity for which the ball left the bat (MPH). Selected because exit velocity may cause the ball to travel farther or escape the reach of defenders.
- Launch Angle: Angle at which the ball left the bat, with 0 being the perpendicular plane of the bat (degrees). Selected because the angle at which the ball leaves the bat can either cause an easy out on a popout/flyout, become a home run, or turn into a different kind of hit on the ground.
- Distance: The distance the ball traveled on the fly from home plate (feet). Selected because the distance hit can elude the defenders, or be hit within their fielding range.

Interactions:

Several interactions were chosen based on the findings within the EDA. Mainly, interactions between the pitch type and pitch speed/spin were obvious to include because certain pitches are based on being faster/having more spin than others. Additionally, interactions between ball distance and launch angle/speed were obvious inclusions because distance is a direct factor of speed and angle, as well as external weather factors not included in the model.

Data Missingness

After doing the initial data collection, filtering and variable selection, there was still some missing data issues. As seen in Figure C in the Appendix, while the overall missingness was only at 2.2%, pitch type had a missing rate of 21.29%. This could be for a number of factors that would be impossible to tell which exactly is the issue. Thus, instead of dropping all observations with the variable missing or dropping the variable itself. We chose to create a 4th pitch type as “Missing” and still include the variable in the model. The EDA on pitch types shows in more detail what this looks like.

Launch Speed, Launch Angle, Distance, Pitch Speed, Spin and Strike Zone also all experienced some type of missing data with the most being at 2.21%. In Figure C, we can see most of the missing entries happened in the same at-bats for which we can assume there was some type of mechanical glitch in those moments that did not allow the information to be recorded. Thus, we felt comfortable just dropping those entries and finalizing our dataset at 3,907 balls in play.

Model

We chose to use a logistic regression model to solve for this analysis. Logistic regression using the `glm()` function allows for us to create a model that yields predicted probabilities of a binary event, which is exactly what we are attempting to study. Logistic regression has the added benefit of ease of interpretability, which is especially important because this study focuses on the interpretation and analysis of the variables in the model and not as much on the actual predicted probabilities. If we were more focused on predicted probabilities, there may be better model fits than logistic regression, but more advanced models sacrifice interpretability for predictive results and that is not what we are hoping to achieve.

In total, the model is fit with 12 variables and 4 interaction terms. It is also trained on the same data as described previously with no additional filtering having been made so that as much data as possible is used. Discussion of the model assumptions can be found in the Appendix. The mathematical formula for our final model is as seen here:

$$\begin{aligned}\frac{\pi_i}{1-\pi_i} = \exp(&\beta_0 + \beta_1 x_{1\text{ Ball Count},i} + \beta_2 x_{2\text{ Ball Count},i} + \beta_3 x_{3\text{ Ball Count},i} + \beta_4 x_{1\text{ Strike Count},i}\\
&+ \beta_5 x_{2\text{ Strike Count},i} + \beta_6 x_{1\text{ Out},i} + \beta_7 x_{2\text{ Outs},i} + \beta_8 x_{\text{Is Home Team},i} + \beta_9 x_{\text{Pitch-Batter Same Handed},i}\\
&+ \beta_{10} x_{\text{Changeup},i} + \beta_{11} x_{\text{Fastball},i} + \beta_{12} x_{\text{Missing},i} + \beta_{13} x_{\text{Pitch Speed},i} + \beta_{14} x_{\text{Spin},i} + \beta_{15} x_{\text{Strike Zone},i}\\
&+ \beta_{16} x_{\text{Launch Speed},i} + \beta_{17} x_{\text{Launch Angle},i} + \beta_{18} x_{\text{Distance},i} + \beta_{19} x_{\text{Changeup*Pitch Speed},i}\\
&+ \beta_{20} x_{\text{Fastball*Pitch Speed},i} + \beta_{21} x_{\text{PT Missing*Pitch Speed},i} + \beta_{22} x_{\text{Changeup*Spin},i} + \beta_{23} x_{\text{Fastball*Spin},i}\\
&+ \beta_{24} x_{\text{PT Missing*Spin},i} + \beta_{25} x_{\text{Launch Angle*Distance},i} + \beta_{26} x_{\text{Launch Speed*Distance},i})\end{aligned}$$

The index of observation i corresponds to each ball in play. Thus, π_i is the "success probability" for ball in play i. This means the probability that a ball in play becomes a hit, since the baseline value is out. $\frac{\pi_i}{1-\pi_i}$ is the odds ratio, or the probability of a given ball in play being a hit over the probability of it being an out. The β values are the logistic model coefficients in the order of "Full Model Output" below with β_0 representing the intercept.

Results

Table 1: Full Model Output

Variable	Estimate	exp(Estimate)	Confidence (2.5%)	Confidence (97.5%)	P-Value
Intercept	0.353	1.423	0.884	2.291	0.15
1 Ball Count	-0.002	0.998	0.967	1.030	0.89
2 Ball Count	-0.024	0.976	0.941	1.013	0.2
3 Ball Count	-0.005	0.995	0.947	1.045	0.83
1 Strike Count	-0.027	0.973	0.941	1.006	0.11
2 Strike Count	-0.022	0.978	0.943	1.014	0.23
1 Out	-0.030	0.971	0.941	1.001	0.06
2 Outs	-0.003	0.997	0.966	1.028	0.84
Is Home Team	0.019	1.019	0.994	1.046	0.14
Pitcher-Batter Same Handed	0.007	1.007	0.980	1.034	0.63
Changeup	-0.564	0.569	0.220	1.471	0.24
Fastball	0.105	1.111	0.607	2.033	0.73
Missing	-0.674	0.510	0.168	1.546	0.23
Pitch Speed (MPH)	-0.001	0.999	0.994	1.004	0.74
Spin (RPM/100)	-0.006	0.994	0.985	1.003	0.17
Pitch in Strike Zone	-0.024	0.976	0.948	1.005	0.11
Launch Speed (MPH)	0.001	1.001	1.000	1.002	0.08
Launch Angle (Degrees)	0.002	1.002	1.002	1.003	<0.01
Distance (Feet/10)	0.021	1.021	1.012	1.030	<0.01
Changeup*Pitch Speed	0.005	1.005	0.994	1.017	0.38
Fastball*Pitch Speed	-0.004	0.996	0.989	1.003	0.25
Missing*Pitch Speed	0.004	1.004	0.991	1.017	0.52
Changeup*Spin	0.007	1.007	0.992	1.022	0.34
Fastball*Spin	0.011	1.012	0.998	1.025	0.08
Missing*Spin	0.015	1.015	0.997	1.034	0.11
Launch Angle*Distance	-0.001	0.999	0.999	0.999	<0.01
Launch Speed*Distance	0.000	1.000	1.000	1.000	0.04

Although only launch angle, distance, launch angle∗distance and launch speed∗distance came back significant at an alpha level of 0.05 we still provide interpretations for all variables on a given ball in play:

- Since setting variables like pitch speed and launch speed to 0 would not make sense for baseball, we do not interpret the intercept.
- Holding all other variables constant, when there is a 1 ball count (versus 0 ball count), the predicted odds of a hit are multiplied by a factor of 0.998.
- Holding all other variables constant, when there is a 2 ball count (versus 0 ball count), the predicted odds of a hit are multiplied by a factor of 0.976.
- Holding all other variables constant, when there is a 3 ball count (versus 0 ball count), the predicted odds of a hit are multiplied by a factor of 0.995.
- Holding all other variables constant, when there is a 1 strike count (versus 0 strike count), the predicted odds of a hit are multiplied by a factor of 0.976.
- Holding all other variables constant, when there is a 2 strike count (versus 0 strike count), the predicted odds of a hit are multiplied by a factor of 0.976.
- Holding all other variables constant, when there is 1 out (versus 0 outs), the predicted odds of a hit are multiplied by a factor of 0.971.
- Holding all other variables constant, when there is a 2 strike count (versus 0 strike count), the predicted odds of a hit are multiplied by a factor of 0.997.
- Holding all other variables constant, when the Cubs are the home team (versus being the away team or at a neutral site), the predicted odds of a hit are multiplied by a factor of 1.019.
- Holding all other variables constant, when the batter and pitcher are the same handedness (versus being opposite handedness), the predicted odds of a hit are multiplied by a factor of 1.007.
- Holding all other variables constant, when the pitch is a changeup (versus a breaking ball), the predicted odds of a hit are multiplied by a factor of exp(log(0.569)+log(1.005)∗pitch speed+log(1.007)spin/100). The predicted odds of a hit on a an 80 mph changeup with 2000 rpm are multiplied by a factor of 0.975 versus a breaking ball with the same stats.
- Holding all other variables constant, when the pitch is a fastball (versus a breaking ball), the predicted odds of a hit are multiplied by a factor of exp(log(1.111)+log(0.996)∗pitch speed+log(1.012)spin/100). The predicted odds of a hit on a an 80 mph fastball with 2000 rpm are multiplied by a factor of 1.023 versus a breaking ball with the same stats.
- Holding all other variables constant, when the pitch is classified as Missing (versus a breaking ball), the predicted odds of a hit are multiplied by a factor of exp(log(0.510)+log(1.004)∗pitch speed+log(1.015)spin/100). The predicted odds of a hit on a an 80 mph Missing pitch with 2000 rpm are multiplied by a factor of 0.945 versus a breaking ball with the same stats.
- Holding all other variables constant, a single mph increase in pitch speed multiplies the predicted odds of a hit by exp(log(0.999)+log(1.005)∗Changeup + log(0.996)∗Fastball + log(1.004)∗Missing). The predicted odds of a 90 mph fastball are multiplied by a factor of 0.995.
- Holding all other variables constant, a single 100 rpm increase in pitch spin multiplies the predicted odds of a hit by exp(log(0.994)+log(1.007)∗Changeup + log(1.012)∗Fastball + log(1.015)∗Missing). The predicted odds of a 2000 rpm fastball are multiplied by a factor of 1.006.
- Holding all other variables constant, when the pitch is in the strike zone (versus not), the predicted odds of a hit are multiplied by a factor of 0.976.
- Holding all other variables constant, a single mph increase in launch speed multiplies the predicted odds of a hit by a factor of exp(log(1.001)+log(1.000)∗distance). Since we are 95% confident the interaction term has no effect (log(1)=0), this is the same as saying the predicted odds are multiplied by a factor of 1.001 for every single mph increase.
- Holding all other variables constant, a single rpm increase in launch angle multiplies the predicted odds of a hit by a factor of exp(log(1.002)+log(0.999)∗distance).
- Holding all other variables constant, for every additional 10 feet of distance, the predicted odds of a hit are multiplied by a factor of exp(log(1.021)+log(0.999)∗launch angle + log(1.000)*launch speed).

Discussion

Looking at the results above, there are some interesting things to notice. First and foremost, only two variables and two interaction terms reach statistical significance at alpha level 0.05–all of them relating to batting. What we realized from this is that perhaps the model is being over powered by the batting variables since game state and pitching information are what lead to the eventual batting stats (launch angle, launch speed, distance) and it is those stats the ultimately determine if the ball becomes a hit or not. To test this theory out, we divided the model up into 3 different models, each including only one type of variables. In the Appendix the outputs of these three models can be seen. While only 2 Strike Count became statistically significant, several variables moved closer to the 0.05 threshold and could be considered statistically significant at slightly higher alpha levels.

In the full model results, in it is interesting to see the detrimental effects the count on getting a hit as we move away from the base 0-0 count. But what does stand out is the positive effect of the Cubs playing at Wrigley Field. Even though it is not statistically significant, it is close, with a p-value of 0.14, signifying that home field advantage can actually play a factor for the Cubs. It is also worth noting the large contribution of the pitch being a fastball (versus a breaking ball) to getting a hit. This shows that the Cubs were a much better fastball hitting team last year than they were against other pitch types. The pitch type results have the potential to aid the Cubs in addressing areas of struggle and aid opponents in game planning against the Cubs in the future. The estimates of the batting variables, while significant, are a bit more difficult to use practically because they vary so heavily on each other. It is interesting to notice however, that each of the three main hitting variables strengthen the probability of a hit with additional increments and distance (by 10 feet intervals) is a rather strong booster of hit probabilities, feeding into the notion of baseball that the unwritten goal of each at bat is to hit the ball the furthest for a home run.

This information as a whole may be relevant for the Cubs to look at certain scenarios from the season last year and find ways they could improve their approach to the at-bat based on the predicted probabilities above. With the abolishing of the the shift (a defensive strategy) [https://www.nytimes.com/2022/09/09/sports/baseball/mlb-bans-shift.html], putting the ball in play has never been more pertinent and accounting for the shift could have been an interesting incorporation to this study had the information been available.

This model certainly had its limitations. The linearity model assumption violation as discussed in the Appendix and the variable domination as discussed above are two of them. Additionally, the model struggles to account for the nature of baseball defense. Infielders and Outfielders are positioned certain distances from home plate and that relationship is clearly not linear (many hits come in between the infielders and outfielders, but get caught if they are hit further).

In future modeling, it would be very interesting to use adding league-wide data to make the analysis applicable to baseball as a whole and not just the Chicago Cubs. The inclusion of defensive metrics and batter/pitcher specific information could improve the the prediction capabilities of the model, or using a more advanced modeling technique for hit prediction could improve this as well. An interesting area of possible exploration could be using a ordinal logistic regression model or multilevel model that accounts for multiple types of hits (singles, doubles, triples, home runs) and multiple types of outs (e.g. flyout, double-play).

Appendix

Figure A: Extended Pitching EDA

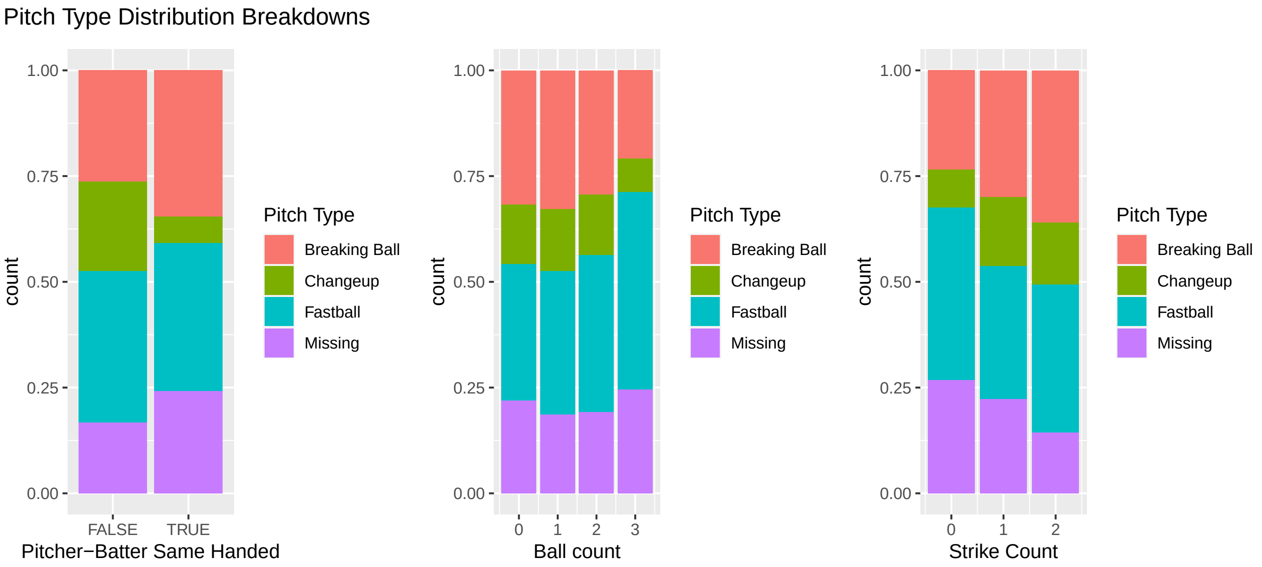
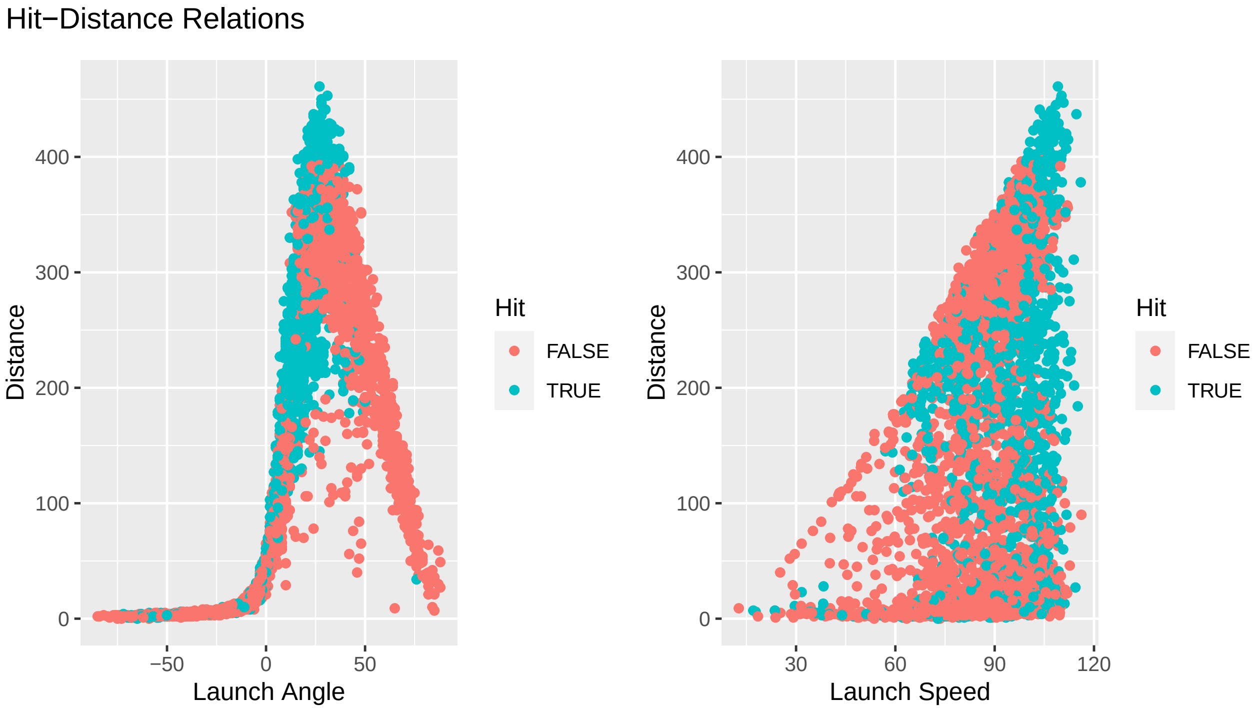


Figure B: Extended Hitting EDA

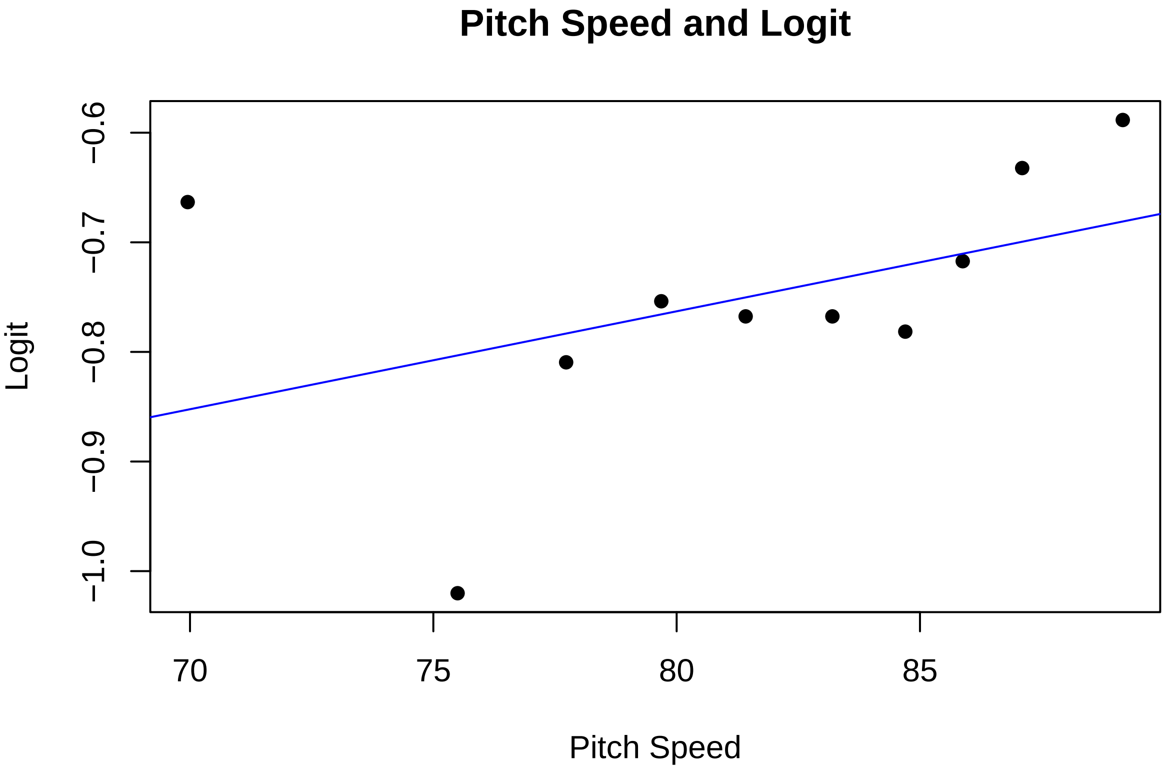


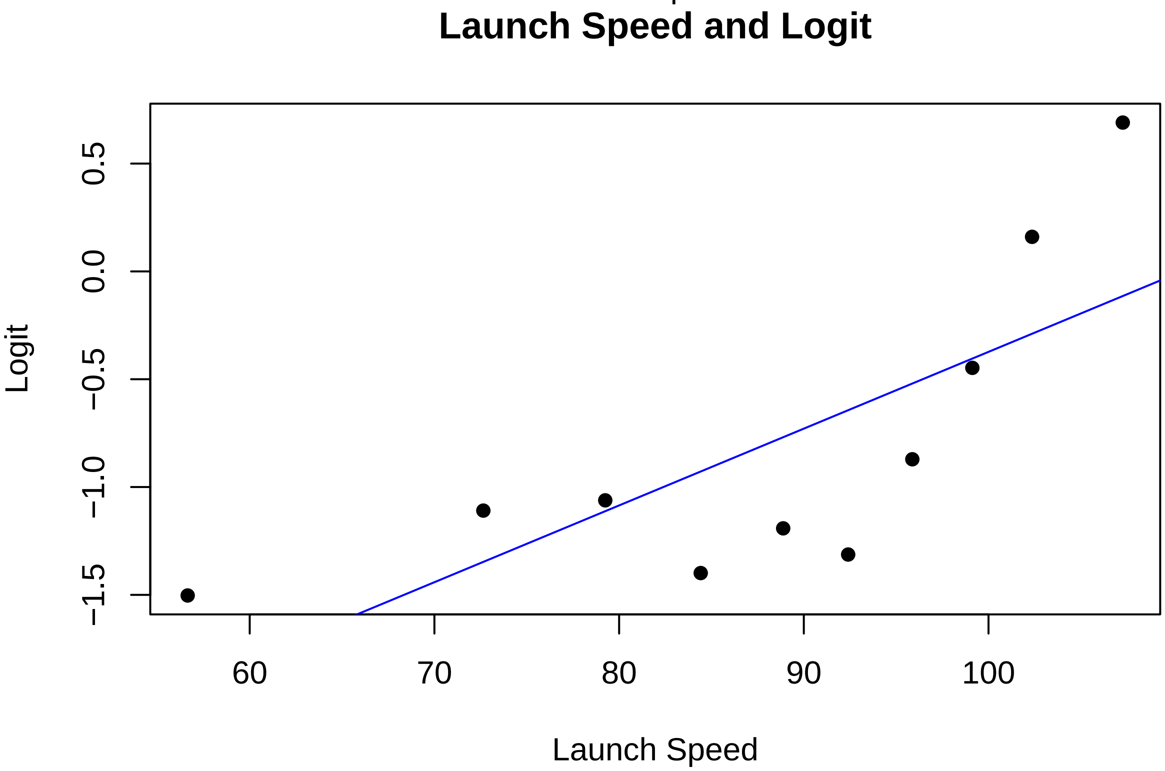
VIF Test

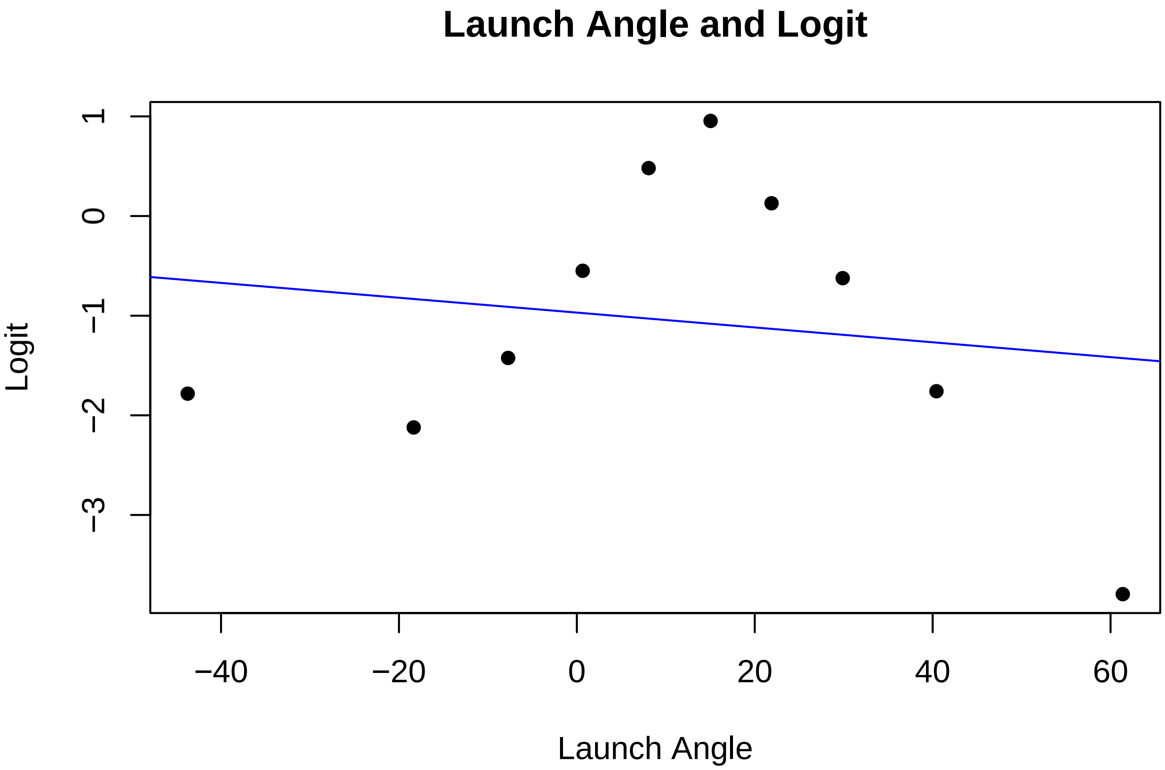
```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

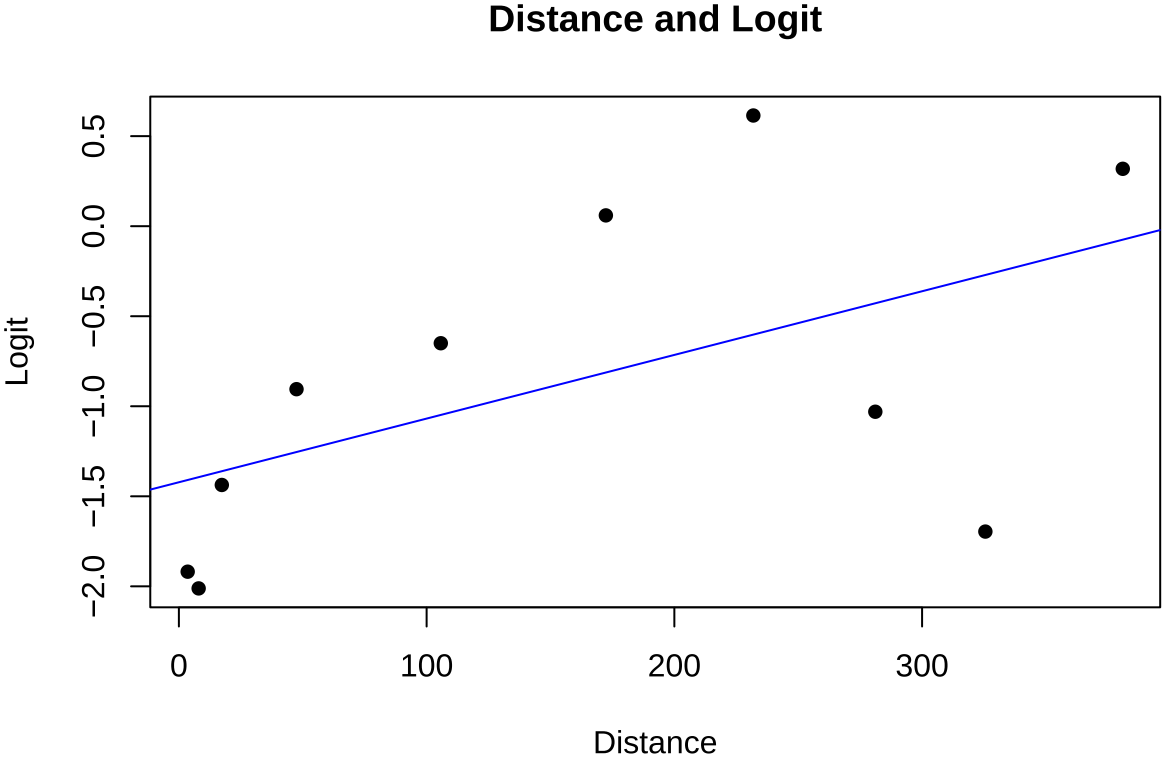
```
##                                GVIF Df GVIF^(1/(2*Df))
## factor(balls)          1.265814e+00  3        1.040068
## factor(strikes)        1.303750e+00  2        1.068559
## factor(outs)           1.015062e+00  2        1.003745
## home_team              1.012565e+00  1        1.006263
## same_side              1.086844e+00  1        1.042518
## pitch_type             2.214972e+08  3       24.597712
## pitch_speed            5.532474e+00  1        2.352121
## spin                   5.880189e+00  1        2.424910
## zone                   1.152856e+00  1        1.073711
## launch_speed           2.118133e+00  1        1.455381
## launch_angle           3.212085e+00  1        1.792229
## distance               8.793036e+01  1        9.377119
## pitch_type:pitch_speed 2.218773e+08  3       24.604741
## pitch_type:spin        4.569039e+05  3        8.776152
## launch_angle:distance  5.739257e+00  1        2.395675
## launch_speed:distance  8.368730e+01  1        9.148076
```

Linearity Analysis









Model Assumptions:

In order to maintain the integrity of the model, all assumptions associated with a logistical model needed to be checked.

1. The response is one variable with two outcomes.

2. The observations are assumed to be independent. For purposes of the study, the observations (individual balls in play) should be assumed to not impact each other. The approach does not assume anything about the pitcher or batters and thus the plays are independent at the start of each pitch.
3. We do experience multicolinearity as seen in the Variable Inflation Factor test above, mainly in pitch type and distance. However, the covariates are included in the interaction terms and are thus exempt from this check.
4. Looking at the linearity plots of the continuous variables above, they noticeably struggle maintaining proper linearity. Due to the nature of baseball, this is a tricky assumption to avoid violating and thus we proceed, but under caution that this assumption may be violated.
5. At nearly 4,000 observations, there is sufficient sizing to the data to satisfy the sample size requirement.

Model Splits

Table 2: Game State Model Output

Variable	Estimate	exp(Estimate)	Confidence (2.5%)	Confidence (97.5%)	P-Value
Intercept	0.344	1.411	1.352	1.473	<0.01
1 Ball Count	0.011	1.011	0.975	1.048	0.56
2 Ball Count	-0.008	0.992	0.951	1.034	0.7
3 Ball Count	0.029	1.029	0.974	1.088	0.31
1 Strike Count	-0.032	0.968	0.932	1.006	0.09
2 Strike Count	-0.045	0.956	0.918	0.995	0.03
1 Out	-0.023	0.977	0.943	1.013	0.2
2 Outs	-0.013	0.987	0.953	1.023	0.49
Is Home Team	0.022	1.023	0.993	1.053	0.13
Pitcher-Batter Same Handed	0.001	1.001	0.972	1.031	0.96

Table 3: Pitch Model Output

Variable	Estimate	exp(Estimate)	Confidence (2.5%)	Confidence (97.5%)	P-Value
Intercept	0.633	1.883	1.104	3.211	0.02
Changeup	-0.814	0.443	0.149	1.318	0.14
Fastball	0.164	1.178	0.590	2.351	0.64
Missing	-0.327	0.721	0.202	2.573	0.61
Pitch Speed (MPH)	-0.002	0.998	0.992	1.003	0.42
Spin (RPM/100)	-0.006	0.994	0.983	1.004	0.22
Pitch in Strike Zone	0.010	1.010	0.979	1.043	0.52
Changeup*Pitch Speed	0.008	1.008	0.995	1.022	0.22
Fastball*Pitch Speed	-0.003	0.997	0.989	1.006	0.53
Missing*Pitch Speed	0.001	1.001	0.986	1.016	0.87
Changeup*Spin	0.007	1.007	0.990	1.024	0.41
Fastball*Spin	0.004	1.005	0.990	1.020	0.55
Missing*Spin	0.013	1.013	0.992	1.035	0.23

Table 4: Bat Model Output

Variable	Estimate	exp(Estimate)	Confidence (2.5%)	Confidence (97.5%)	P-Value
Intercept	0.119	1.126	1.013	1.252	0.03
Launch Speed (MPH)	0.001	1.001	1.000	1.002	0.14
Launch Angle (Degrees)	0.002	1.002	1.002	1.003	<0.01
Distance (Feet/10)	0.021	1.021	1.012	1.030	<0.01
Launch Angle*Distance	-0.001	0.999	0.999	0.999	<0.01
Launch Speed*Distance	0.000	1.000	1.000	1.000	0.04

Figure C: Visualizing Missing Data

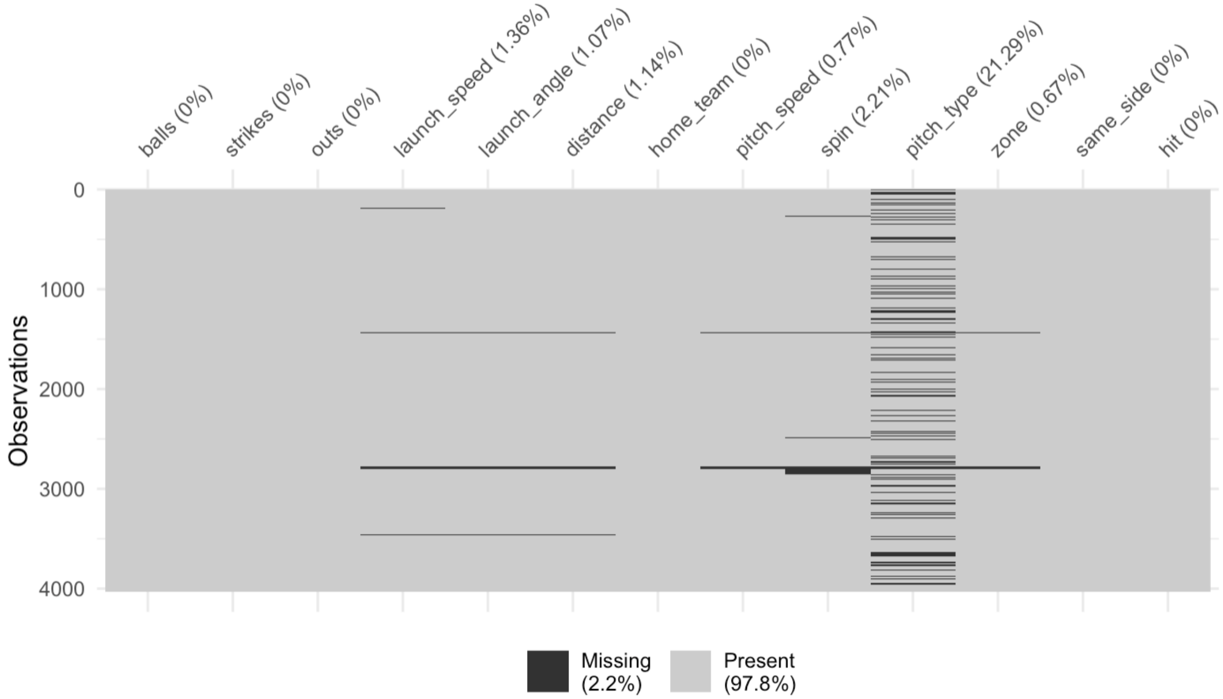


Figure 1: Figure C: Missing Data

Coding and Data Sources

https://billpetti.github.io/baseballr/index.html

https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm

https://bleacherreport.com/articles/1803416-is-home-field-advantage-as-important-in-baseball-as-other-major-sports

https://www.nytimes.com/2022/09/09/sports/baseball/mlb-bans-shift.html