UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE RELATIONSHIP BETWEEN BATTER'S EYES CHARACTERISTICS, THE HITTING PROCESS AND HIT QUALITY AMONG MAJOR LEAGUE BASEBALL PLAYERS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ANDREW GILLINGHAM

Norman, Oklahoma

2019

THE EFFECTS OF BATTER'S EYE CHARACTERISTICS ON HIT QUALITY
AMONG PROFESSIONAL PLAYERS

A THESIS APPROVED FOR THE
DEPARTMENT OF HEALTH AND EXERCISE SCIENCE

BY

Dr. Daniel Larson, Chair

Dr. Christopher Black

Dr. Amanda Wilkerson

Table of Contents

Chapter 1: Introduction **1**
Purpose of the Study 2
Research Questions 4
Hypotheses 4
Significance 5
Limitations 6
Delimitations 6
Assumptions 7
Operational Definitions 7

Chapter 2: Review of Literature **9**
Introduction 9
The Hitting Process 10
Performance Metrics 18

Chapter 3: Methodology **22**
Research Design 22
Sample 23
Instrumentation 24
Data Collection Procedure 25
Data Management and Analysis 27
Model Specification 28

Chapter 4: Results **31**
Introduction 31
Effects on Swing Decision 32
Effects on Exit Velocity 34
Effects on Launch Angle 37
Hit Quality to Runs 39

Chapter 5: Discussion **42**
Discussion 42
Recommendations for Practice 47
Future Research 48
Conclusion 50

References **52**
Appendix A: Tables & Figures **57**
Appendix B: Code **59**

Abstract

Because there are no rules on how MLB stadiums are to be designed, there end up being a variety of differences between the stadiums. This includes the batter's eye. This study seeks to see if there are differences in hit quality i.e. exit velocity and launch angle, that can be attributed to these differences. This descriptive correlational study looked at pitch level data at the major league level from 2017-2018. In this study, we found that there are significant differences between several key features of the batter's eye. These differences have been found to affect the batter's decision to swing or not, the exit velocity of the ball coming off the bat and the deviation from the mean launch angle. The relationship of these hit quality metrics to runs scored were also studied so that differences in batter's eyes could be attributed to a certain number of runs gained or lost throughout a season.

Chapter 1

Introduction

In basketball, a court is going to be exactly 94 feet long and 50 feet wide. In professional football, a field is always going to be 120 yards long and 53 yards wide. However, in baseball, apart from the standard dimensions of the infield, there is no league standard of how an outfield should be constructed, only guidelines (Official Baseball Rules, 2019). This lack of consistency has led to monstrous green walls, hills in centerfield, ivy covered walls, Pesky's pole, and some truly odd dimensions. But with all of the quirks and oddities that each stadium may have, one thing has remained fairly consistent, a large dark green or black area in centerfield, directly behind the pitcher and in direct line of sight of the hitter. This is the batter's eye.

The history of the batter's eye is ambiguous. It isn't known when and where this addition came about, but baseball historians have found evidence of it as far back as the 1880's (Holmes, 2017). Its purpose is to provide a dark, neutral background for hitters to be able to more easily see the ball as it is being thrown. And although it became more common to paint black backdrops when Major League Baseball (MLB) was first starting up, several notable stadiums, Fenway Park, Wrigley Field and Yankee Stadium, excluded a batter's eye when being built (Borzi, 2008). With the ballpark construction boom of the 1960's, stadiums in Los Angeles and San Francisco brought back free-standing batter's eyes, large, dark screens behind the centerfield fence (Holmes, 2017). More recently, teams like the Rockies, Twins and Cubs have implemented aesthetic designs to their batter's eyes with the use of shrubbery or trees. There was even some slight controversy in 2010 when Target Field placed trees in front of their batter's eye that catcher Joe Mauer called “one of the worst backdrops we’ve ever seen” (Gleeman, 2011). The trees were removed that winter and there haven’t been any complaints about the batter’s eye since. Because of the peculiarities of each stadium’s batter’s eye, they may play a role in how hitters see the ball as it comes out of the pitcher’s hand. The reason batter’s eyes are being examined in this study is that while players may have different personal opinions on the different types of backgrounds, it is unclear if there is an objective and significant difference in hitting performance based on the type of batter’s eye. The general research question is therefore, do the differences in backdrops make a quantifiable difference in how well players hit?

Purpose of Study

The idea for this study originated while reading an article on the popular baseball statistics website, Fangraphs.com. Created in 2005 by David Appelman and designed to help people with their fantasy baseball teams, Fangraphs since has expanded to be one of the premier baseball statistics research sites. The post by Jeff Sullivan on September 12, 2018, *There’s Definitely Something Strange About Citi Field*, outlined that since their last major renovation in 2012, the Mets have consistently ranked last in batting average on the balls that they made contact with and put into the field of play (BABIP) at home but in the top third when they are on the road (Sullivan, 2018). Two other teams that had a major difference in home versus road BABIP were the San Diego Padres and the Houston Astros. The 2017 Astros team had historic offensive production leading in eight different offensive categories and went on to win the World Series. So, the problem they and the Mets faced was likely not due to the quality of team but may be due to external factors. Just before the 2017 season, the

Astros renovated their centerfield area, bringing the fence 27 feet closer and completely changing their batter's eye. Not only was BABIP affected by playing in their home stadiums, so was exit velocity, which is how fast the ball is traveling when it leaves the bat. These three teams, the Astros, Mets and Padres, also took three of the bottom four spots with a drop in exit velocity when they played in their respective home stadiums. Although Petco Park (i.e. the Padres home stadium) has not had any recent renovations of their batter's eye, this problem may stem from the location of their ball park in San Diego with a layer of heavy marine air (Sullivan, 2018). While this is usually adjusted for with park factors, they could be coming from biases beyond the scope of Sullivan's testing.

Over the last few years, exit velocity and launch angle have become synonymous with the data revolution that has swept through the MLB. Though they have been talked about constantly, how important are they really in determining player quality? On the surface, it would appear that these qualities of a hitter are very important, as the top hitters will typically have above average exit velocities and launch angles in the mid-teens (Petriello, 2016). That is, the top offensive players will typically be hitting the ball harder than average and their launch angle, or the angle at which the ball leaves the bat parallel to the ground, will be in a range that creates a line drive and not a pop-fly. However, regression analysis revealed only a weak-moderate relationship between exit velocity and batting average, BABIP and slugging percentage (Anelso13, 2016). Looking solely at these two stats to determine how good a player is would be a poor choice, however, ignoring them just because they seem like a fad would also be equally irresponsible because they can be predictive of hitting success.

Although there has been some research on the visuals of hitting, the effects of color, and how players track pitches (Kato & Fukuda, 2002; Davis, 1978), no comprehensive study has looked at what background size, color or texture in a batter's eye is the best for hitters to be able to see the ball.

Research Questions

This paper will address several questions. All of the questions pertain to the visuals of hitting and how players react with different classifications of batter's eye. The first question is, does the batter's eye affect a batter's decision to swing or not? This decision could come as a direct response to being able to see the ball or not. This question looked solely at the decision a batter makes to swing. Non-swinging events like bunting and being hit by the pitch were excluded. Whether or not there is a change in that behavior, the next question examined was does a batter's eye affect exit velocity of balls that are actually hit? Furthermore, does the batter's eye affect the launch angle of batted balls? These two factors play an important part in how well a batter does during an at bat and can directly affect offensive quality and run production for the team. As a final conditional analysis, if the batter's eye characteristics do affect hitting performance, the potential impact on run production is examined. If there are impacts on exit velocity or launch angle, how did that translate to runs across a season for an individual team?

Hypotheses

If the type of batter's eye used makes a difference, then significant differences will occur in exit velocity and launch angle variation, and a change in runs per game in that stadium. Alternatively, if the type of batter's eye does not make a difference, then there will be no significant change in exit velocity, and expected runs will not be affected.

H_0: Characteristics of the batter's eye have no effect on the probability of a batter swinging, or the hit quality of a batted ball.

H_1: The probability of a batter swinging at a pitch will differ based on characteristics of the batter's eye.

H_2: Exit velocity will differ based on the characteristics of the batter's eye.

H_3: Launch angle variability will differ based on the characteristics of the batter's eye.

Significance

The impact of batter's eye design on the game of baseball could be substantial if there is evidence that certain types of configurations impact the playing performance of both hitters and pitchers. At the very least, teams that have a batter's eye that hinders their hitter's performance, could make some cost-effective changes to try and help their own team. Or if a team is building a new stadium, which at least four teams have discussed, the design could incorporate the most hitter friendly elements (Sanders, 2019). While the argument could be made that teams would not want to have hitter friendly batter's eyes because it will help their opponent as well, there has been a general push recently for more offense. Some people find the game of baseball to be "boring" and "slow," and partially because of this, there is a myth that the game of baseball is "dying" (Calcaterra, 2018). Ever since Rob Manfred took over as the MLB Commissioner in early 2015, he has been pushing for more offense and faster pace of play. Although the pace of play rules have been controversial, offensive numbers are at their highest level since the Steroid Era of the 1990's ("Major League Baseball

Batting", 2019). In 2018 the Yankees broke the record for most home runs hit by a team in a season, and in the middle of June there are currently four teams that are on pace to beat that mark (Jaffe, 2019). So, if a team wanted to increase the offensive production of their stadium without having to make drastic changes to their team, they could more easily change their batter's eye.

Limitations

One limitation of this study is the reliance on data collected from outside sources. Because we cannot capture all of the data ourselves, the data was drawn from third party resources. These outside sources, however, are considered to be the most reliable tracking measures in baseball and the data collected by them are used by all of the professional teams in the league. Another limitation of this study is not being able to use exact measures of the characteristics of the batter's eye. For example, there are many different shades of the color green around the league, and therefore, some color classifications involved a level of subjective evaluation. This evaluation also included the dimensions of the batter's eye itself. Because this information is not always publicly available, the color, texture, angular orientation, observed lighting, configuration and overall dimensions were generated by a panel of baseball experts. Lastly, a limitation that could be detrimental could be computing power. Because of the nature of the study, the data set was massive, and it included well over 1 million observations. Therefore, the data were parsed into manageable random samples when necessary.

Delimitations

The delimitations of this study are:

1. This study considered Major League Baseball Statcast Data from the 2017 and 2018 seasons.
2. All pitches from 29 of the MLB stadiums were evaluated.
3. Fenway Park, home of the Red Sox, was not included due to their lack of consistent batter's eye.

Assumptions

The assumptions of this study are:

1. The Statcast data is precise and accurately records the data correctly.
2. Coding for batter's eyes remain consistent and fair for all categories.
3. The there is a normal distribution of pitching talent across the league.
4. Batters are trying to maximize their exit velocity and optimal launch angle for each at bat recorded.

Operational Definitions

1. <u>BABIP</u>- batting average of balls in play. Of balls that were hit into the field of play, how many fell in for a hit. The formula for this is: *(Hits-Home Runs) / (At-Bats – Strikeouts – Home Runs + Sacrifice Flies)*
2. <u>Batter's Eye</u>- the large, dark colored neutral area directly behind the pitcher from the batter's perspective in centerfield. In this study, hitting background may be interchanged with batter's eye.
3. <u>Exit Velocity</u>- the speed of the ball as it leaves the bat.
4. <u>Launch Angle</u>- the angle of the ball as it leaves the bat, where zero degrees represents an initial path that is parallel to the ground.

5. Park Factors- a statistic that tries to take into account different features of a park like elevation, weather and wind patterns. It enables people to see which parks are “pitcher friendly” or “hitter friendly”
6. Pythagorean Expectation- Developed by Bill James, this formula, based to look like the Pythagorean theorem, derives a team’s expected winning percentage based on the run differentials over a season. The formula is: $(Runs\ scored)^{1.83} / ((Runs\ scored)^{1.83}+(Runs\ Against)^{1.83})$
7. wOBA- weighted on base average. Based on linear weights, this statistic is designed to measure a player’s total offensive productivity. It is highly correlated with runs.
8. xwOBA- expected weighted on base average. Designed to measure a player’s offensive productivity based on the exit velocity and launch angle of any given batted ball, sprint speed is used in specific instances.

Chapter 2

The purpose of this study was to determine if a stadium's batter's eye makes a difference in hitting performance. Because there are no set rules for a batter's eye, every team will have different backgrounds. These differences may have an effect on how hitters "pick up" the ball, how quickly they are able to react and the quality of the contact that they make with the ball. We will attempt to estimate how much impact these differences have on exit velocity, launch angle and decision making. In addition to this, we will see what effect hit quality has on overall run performance. As teams and players look more closely at launch angles and exit velocities, this study includes research into how much of a factor a batter's eye play in scoring runs, and ultimately winning games.

This chapter will look at previous research in each of the stages of the hitting process, starting with how players first find and focus on the ball in the air, then their decision on whether or not to swing. How they are able to see the ball mid-flight and then intercept the path of the ball by swinging the bat. This section includes studies of physical differences between baseball players' and the general public's eyes. Literature related to the anticipation skills and reaction times of hitters and how color may affect their ability to see the ball was examined. Finally, background for the evaluation in hitting performance through exit velocities, launch angles and ultimately wOBA is presented

With the rise of the data revolution in the MLB, there has been some recent research done, albeit in non-academic settings, on the effects and importance of launch angles and exit velocities. However, through numerous searches in various journal databases, no academic study has been found that looks at how background visuals may impact hitters. For this literature review, a majority of the studies looked at will describe the process of hitting, from the initial anticipation of the pitch, the search strategies used to find the ball out of the pitcher's hand, being able to see the ball mid-flight and the timing needed to make quality contact with the pitch. It is important to understand the process that hitters go through every pitch because the entire process appears to be based on visual cues. Hitters are not just randomly guessing where to swing at every pitch (Sarris, 2016). The databases used were accessed from the University of Oklahoma's library's online resources. The journals and databases searched were, SportDiscus, Medline, Jstor, PsycInfo and Journal of Quantitative Analysis in Sports. A final important source of information is the website, Fangraphs. Although this source, to some, may be considered just a blog, it is one of the most respected and comprehensive baseball statistics websites with many of the writers coming from or going on to work for professional teams. They are widely considered to be some of the best "amateur" data scientists and statisticians. In their articles, they will often lay out the step-by-step process for their testing methods as well as their conclusions. Nonetheless, Fangraphs will mostly be used for defining terms, providing context, framing anecdotes and filling in any gaps that peer-reviewed academic research does not cover.

The Hitting Process

Hitting a baseball has been claimed to be one of the hardest things to do in all of sports (Regan, 1997). With a distance of 60 feet between the mound and home plate, a league average fastball of 93 mph will reach the batter in less than four tenths of a second, or about three blinks of an eye (Shank & Haywood, 1987). In that time, players must locate the ball, perceive the type of pitch, speed and where the ball will end up, on top of reacting in the most appropriate way. To do this, they must utilize both inherent abilities and skills developed through repetition.

One of the first things that a player will do is try and locate the ball coming out of the pitchers' hand as quickly as possible. This is done by using search strategies. Search strategies are the small and quick movements of the eye in order to locate and pinpoint a moving target, in this case a baseball about to be thrown. A 2002 study found that professional players had a narrow search area and were able to set their initial focal point faster than amateurs (Kato & Fukuda, 2002). They were able to ignore the general motion of the pitchers and focus in on an area around the pitchers' elbow in order to pick up on the ball as it is coming out of the pitcher's hand (Kato & Fukuda, 2002). This allows them to maximize their time assessing the pitch. This study provides evidence that players do have an exceptional ability to pinpoint the ball earlier, a common strategy that has been voiced by Kris Davis, Manny Machado and hitting coaches across the league (Sarris, 2016). The search strategies that hitters use extend beyond the initial windup of the pitchers. Although most hitters will be focused on the ball as it leaves the hand, hitters will typically use the first thirty feet or about 200 milliseconds (ms) of the ball's flight, to gather the rest of the information needed to judge how to react to the pitch (Shank & Haywood, 1987). In the 1987 study, Shank and Haywood found that professional hitters and amateurs had similar initial eye movement reaction times but that the pros were able to locate and distinguish pitches 63% faster, likely due to experience. However, there very well could be biological explanations as to why professionals are still better than amateurs at hitting.

Between 1992 and 1995 a team of researchers looked at the eyesight of three hundred eighty-seven professional baseball players, both major and minor league, and found that their eyesight was significantly better than the general public. Tested on visual acuity, distance stereoacuity and contrast sensitivity, baseball players' visual abilities surpassed even what the researchers expected. Visual acuity is the clarity or sharpness of a persons' vision, typically described as 20/20, 20/15 and so on. Stereoacuity is a persons' ability to distinguish differences in distance between objects that are far away from them. Lastly, contrast sensitivity is the ability to distinguish colors from each other even when they look incredibly similar. They found that 81% of the players tested at having at least 20/15 vision and about 15% approached the perceived human limit of 20/7.5 (Laby et. al., 1996). The team also used multiple different tests to evaluate the players on color contrast sensitivity. In all three tests used, mean scores were significantly higher than the general population but in both the Vision Contrast Test and Contrast Sensitivity Viewer, twenty-eight players achieved a perfect score. Although there were more minor leaguers that achieved a perfect score than those in the MLB, the percentage of perfect scores for MLB players was more than double that of the minor leaguers (Laby et. al., 1996). Professionals may have better visual skills than the general population, however, there are also learned abilities that start at a young age. A similar visual acuity test was performed on youth baseball players and similar results were found to that of the Laby study. In this case, kids between the ages of 10 and 18 were tested on their static near stereoacuity. Static near stereoacuity being the ability to distinguish distance of objects a closer distance. This involved them trying to discern differences in distance between different circles as they traveled across an arc. The faster they were able to do so, the better their score. The youth that played baseball or softball had a final mean score of 25.2 seconds of arc while the non-athletes had a significantly higher mean of 56.2 seconds (Boden, Rosengren, Martin & Boden, 2009). Knowing that professional and youth players have exceptional visual skills and are able to quickly narrow their search patterns helps provide evidence that hitting is truly a visual endeavor. Because baseball players rely so heavily on their visual abilities, differences in backgrounds that occur behind the pitcher could have an impact on their ability to see, focus on and hit the ball.

Even though the classic baseball saying “see the ball, hit the ball” would appear to be true, players still have to rely on their anticipatory skills to be able to time up their swing so that they are able to make quality contact and “barrel” the ball up. The term “barrel” meaning hitting the ball squarely in the “sweet spot” on the bat, about 5-7 inches from the end and in the center of the bat. Statcast typically defines a barrel as a ball struck that has a combination of launch angle and exit velocity that will produce a batting average of at least .500 and a slugging percentage of 1.500. Although they are having to rely on some amount of anticipatory skill, players striking any moving object, like a batter hitting a pitch, must start by receiving the necessary information through visual cues. And that is what separates professionals from amateurs. In a 2017 study, batters were shown video of pitchers going through their windup. The video was then cut off 200ms after the pitch was thrown and the batters were asked to predict where the pitch would end up and what type of pitch was thrown. The same batters were then shown similar video but with the video being cut off just 80ms after the pitch was released. The batters were found to be 33% better than just randomly guessing when shown the first 200ms and 16% better when seeing the first 80ms (Morris-Binelli, Müller & Fadde, 2017). Being able to pick up on the initial cues allowed those with experience to better be able to determine what was going to happen during the pitch.

The anticipatory skills of baseball players have been shown to be significantly better than the general population but they are far from perfect. One of the main roles for pitchers is to deceive the hitter; and with strikeouts on the rise, it would appear that they have gotten better at doing just that. Many hitters walk up to the plate with an idea about how the pitcher will throw to them based on previous at bats or pitcher tendencies. Chances are, the batter will be wrong but based on the cues that batters are able to see in the first 200ms they do have the chance to change their timing. In a 2012 study from Japan, researchers looked at people's ability to react to sudden speed changes in moving objects. They showed baseball experts and novices videos of targets moving at a constant speed and asked them to react when they should "hit" the target. The speed of the ball would then suddenly and randomly slow in velocity. The researchers found no significant difference in reaction when the target was moving at a constant speed. But when there was a change in velocity, experts were able to change their timing significantly better than the control group (Nakamoto & Mori, 2012). A hitter's ability to change their timing is one thing, but being able to still keep an accurate swing path while changing timing is a whole other thing. This problem was also looked at in Japan at Waseda University. The team of researchers asked participants to hit a ball off of a tee so they could measure the distance between contact and the "sweet spot" of the bat. Participants were then asked to adjust their swing to intentionally miss the ball and swing just over it. This was to be a proxy for batters trying to adjust their swing during the flight of the pitch in order to make contact. While hitting off the tee, two-thirds of the participants had an average contact area in the sweet spot. However, during the altered swings, only one participant's average "contact" was in the sweet spot (Higuchi, Nagami, Morohoshi, Nakata & Kanosu, 2013). A baseball is only a few inches in diameter, and simply swinging above it proved that it does not take much to alter the accuracy of a batter's swing. In fact, even smaller changes may have a large impact on accuracy. In 2018, researchers at the Tokyo International University looked at how minor head movements at the beginning of a swing can impact the accuracy and contact of a swing. They had batters focus on and track a ball that would force their head to move as they began the loading phase of their swing. After the data had been collected, they had found a significant negative correlation between the amount that the head had to move while tracking the ball and the quality of contact the batter made. They also noted that the point of contact was consistently lower on the barrel the more players had to adjust their head in order to focus on the ball (Akaike, Fumoto & Usuil, 2018).

Stepping out of the realm of baseball and into the greater world of striking sports like badminton, tennis and cricket, a review by Sean Müller and Bruce Abernathy in 2013, found that professionals across striking sports had several things in common when it came to anticipating where and when to strike. They were able to use smaller amounts of motion, or kinematic cues, from the beginning of the ball's path to determine where they needed to make contact. Once the ball was in flight, they were also able to make adjustments faster to be able to strike with more accuracy than control subjects (Müller & Abernathy, 2013). Training anticipation and the ability to make better anticipated judgements is something that a 2011 study looked into. Looking at hitting in baseball as well as striking in soccer and hockey, the researchers attempted to create a training protocol that would train anticipation in amateurs and professionals. One method of training was to show video of players passing the ball/puck and to freeze the video just before a decision was made. Players would then be instructed to highlight areas that the pass should go to and rank them on where they believed the pass will go. More elite players were able to highlight and rank passes that had better chances of scoring than semi-professional athletes according to one test's findings (Williams, Ford, Eccles & Ward, 2011). They also found, similar to the Morris-Binelli, Müller & Fadde study, that more experienced baseball players were better able to predict the location of pitches with minimal visual cues. They concluded that most athletes that were studied had improved upon their anticipatory skills through reviews of video, making predictions, or had spent a considerable time practicing the specific striking actions necessary for their sport (Williams, Ford, Eccles & Ward, 2011). However, a common theme amongst these papers is that the anticipatory skills of these professional athletes is heavily reliant on visual cues. Even in the Gray & Cañal-Bruland study, a wider variety of pitch selection meant that hitting ability decreased because the batters were not able to anticipate the pitch as well.

The point of the batter's eye in baseball stadiums is to make the ball more distinguishable for hitters. It is thought that the differences in color might have the greatest impact on being able to see and hit the ball (Davis, 1978). Although this study appears to be the first to look, in depth, at this question it is not the first to study the effects of color when hitting. A 1978 study at Middle Tennessee State looked at what might happen if the color of the ball was different. The study had twenty people, five baseball players and fifteen non-athletes, watch and swing at normal white, bright yellow and orange baseballs and then rank their visual perception of each type of ball. The final results showed no statistically significant difference in how well hitters performed or in what they preferred. One thing the researcher did note was that there seemed to be a trend that players liked the brightly colored baseball better than the white one but that sample size could be part of the reason that preference did not translate to results in the data (Davis, 1978). One of the takeaways from this study is that there may be more than just good eyesight that helps players focus on the ball as it is moving. Two years later at Montana State University, researchers looked at the effect of background color on hitting, this time with children. Using blue, yellow and white balls as well as white and black backgrounds the researchers tested elementary school aged children on their ability to strike the balls with a wooden racket. The racket was colored differently based on the ideal areas of contact. The "sweet spot" was one color, with the outer edge and handle different colors to make seeing where contact occurred easier. The children were then asked to hit the colored balls they were pitched and analysis was done on the quality of contact that the children made. This study found that although color had a significant impact on the quality of contact, their results were different than what was in line with their previous research. Yellow and white balls fared the best while blue did the worst. They also found that the white background had the best performance which contradicted their previous research (Alomar, 1980).

Other strategies that may help can be found in the sport of cricket. Baseball and cricket batters may have different types of swings but they are trying to do essentially the same thing. Cricket batters have to find and focus on a ball as it is hurled towards them but with the added difficulty of possibly having to hit it off a bounce. Researchers at Edinburgh University performed a study on cricket batsmen to see what cues they look for while trying to hit. They found that the types of deliveries were what most hitters were trying to pick up on. Although some tried to look for the spin of the ball the batters that did the best seemed to be able to watch the ball for longer. Any ball that had bounced within 200ms of making contact with the bat was significantly less likely to be hit or hit well (Renshaw & Fairwether, 2000).

Though there have not been any studies that have shown the effectiveness of the batter's eye on the hitting process, the studies that have been reviewed show that excellent vision is almost required to be a great hitter. The studies discussed in this chapter give an overview of what a player goes through during an at-bat. They highlight how important visual cues and contrast differences can be while trying to hit a moving object. Although baseball players have incredible anticipatory skills, they still have to draw their conclusions based on what they see during the pitchers' motions and release. These reasons are why there might be a basis for the batter's eye causing differences in hit quality in different parks in the MLB.

Performance Metrics

A player's hitting ability can be measured in many different ways, from a traditional batting average (BA), to the first generation of "advanced" metrics of on base percentage (OBP) or slugging percentage (SLG), to the neopostmodern stats of batting average on balls in play (BABIP), weighted runs created plus (wRC+) and expected weighted on base average on contact (xwOBAcon). In order to better translate the results of the study to a meaningful difference in runs, these metrics must have some impactful link to the outcome of the game. There has been little peer-reviewed work on the utility of recent metrics, so much of the sources will come from worked published to Fangraphs and their associated websites.

One peer reviewed study was found that shows a strong relationship between runs scored in a season and a player's wOBA. In 2012, three researchers wanted to build a model that could predict a player's runs scored in a season based on their offensive ability. They used both "traditional" metrics as well as sabermetric statistics to see which would do a better job at predicting runs. In the end, they found that the best predictor of season runs for a single player was wOBA (Beneventano, Berger & Weinberg, 2012). They found that a player's wOBA had an R-squared value of 0.896, meaning that almost 90% of the variance in a player's runs could be explained by their wOBA (2012). Another study was done to discern the importance of exit velocity. Posted as a part of the community research of Fangraphs in 2016, user Anelso13 regressed exit velocity with BABIP, strikeout rates, home runs, slugging percentage and the metric, isolated power. They found that in each instance exit velocity had either a weak or moderately weak relationship with the metrics tested (Anelso13, 2016). This study shows that although exit velocity is an important factor when looking at hit quality and player skill, it is not the end all be all measurement of a player's contribution.

The other key component to hit quality is launch angle. Measured as the angle at which the ball comes off of the bat with zero degrees being parallel to the ground, when paired with exit velocity these two measures can be used to calculate an expected weighted on base percentage. Launch angle is important because it determines the flight of the ball. When fans see a fly ball, they should know that the ball was hit with an angle of greater than about 30 degrees (Zimmerman, 2018). Groundballs will typically have negative launch angles and line drives, what players are usually trying to hit, will be between 5 and 25 degrees. These classifications of hits are important because they have vastly different outcomes. In 2017 batters only had a .205 batting average on fly balls, but had a .712 slugging percentage, indicating low chance of getting on base but a high chance of scoring (Olney, 2017). On groundballs, players had a better batting average at .245 but had a much worse slugging percentage at .266. The best numbers came from line drives where they collectively hit .682 and slugged .989 (Olney, 2017). By far the best offensive production came from line drives, which is the reason for the ongoing trend of players changing their swing so that they can hit more line drives and fly balls.

As previously mentioned, exit velocity and launch angle are two of the components that comprise a player's expected weighted on base average (xwOBA). On groundballs a player's sprint speed is sometimes taken into consideration, however it is only a very small portion of a player's xwOBA. xwOBA can be calculated by mapping out all of the combinations of exit velocity and launch angle. Next the wOBA is calculated from all of the batted balls that correspond with that exit velocity and launch angle (Zimmerman, 2018). If done correctly, areas should emerge having higher values than others. Typically, these areas will have higher exit velocities and a launch angle that is considered to be a line drive. By mapping out previous batted ball data like this, players now know how they should try and hit the ball in order to maximize their results.

Chapter 3

Advanced statistical analysis has been used in the MLB for over twenty years now, making baseball one of the most analyzed sports in the world (Lewis, 2011). But there appears to be a void of research when it comes to the specific background that hitters encounter while batting. Although the information about how players process the speed, motion, spin and direction of a ball play a significant role in their ability to hit, the backdrop on which they are trying to make out these factors could also play a significant role in the quality of their contact. This chapter will go into detail of the data collection and testing methods used to analyze the data. It will also present the programs and data management systems chosen to store and organize the data.

Research Design

The design of this study is descriptive and correlational. The data was collected from the 2017 and 2018 MLB seasons. No interventions were undertaken. This design allows us to test the research hypotheses and ultimately determine whether or not background visuals have a relationship to hitting performance at the major league level. The variables of the batter's eye assessed were: color, texture, angle, height, width and the presence of lights in the immediate area. A panel logistic regression was applied to see if there was a change in initial swing decisions. A panel multiple linear regression model using the same characteristics was applied to the exit velocity and launch angle of the players at different stadiums to tell us more about the relationship that may exist between batter's eyes and hitting quality. Finally, in order to examine the relationship between hit quality and runs, a multiple linear regression model to predict wOBA was estimated, and the expected impact of the batter's eye characteristics on runs was generated using those xwOBA values.

Sample

This study uses batted ball data from each stadium (excluding Fenway Park) in order to determine the quality of hitting. The reason that Fenway was excluded was because they did not maintain a permeant batter's eye. During day games, a tarp is placed over two centerfield sections but it is removed during night games. Its exclusion was determined because the characteristics that it exhibited were very unique and could be found in other stadiums. Every pitch over the last two years with complete data (2017-2018) was recorded by Statcast, and used in this study because it is the most accurate, publicly available data. The main reason for only using the last two seasons, is that there have not been any major renovations to any stadium's batter's eye over this time period. Before 2017 there had been several significant changes to the batter's eye in Wrigley Field and Minute Maid Park. Wrigley Field made adjustments to the glass and shrubs in front of their batter's eye, while Minute Maid Park completely renovated their centerfield by moving the fence closer and changing the wall to a large, green rectangle. Although two seasons may initially sound like a small data set, there were almost 750,000 pitches thrown in the 2018 season alone. Two years should reflect the current state of pitching strategy and batting trends, and provide a consistent environment for each stadium. Also, it is worth mentioning that the Twins have redone their batter's eye to make it larger, a different shade of green and will use real plants to cover the wall. This renovation was completed before the 2019 season began.

Instrumentation

In 2015, MLB installed Statcast systems in all major league parks. Statcast is a system of cameras and radars that are set up around a stadium that will record virtually everything happening on the field. The reason that teams are now able to know what exit velocities, launch angles, spin rates and catch probabilities are, is because of Statcast. The system is able to track all of the movements of both the ball and the players and records data that was not possible a decade earlier. Statcast was considered to be one of the best systems when it was first installed but it is not perfect (Schifman, 2018). When analyzing the different methods of pitch tracking, the 2017 version was found to be the most accurate, but was still influenced by different stadiums. The error in velocity was found to be 0.07 mph with horizontal and vertical pitch location error being 0.35 inches and 0.72 inches respectively (Schifman, 2018). A baseball is between 2.87 and 3 inches so the errors found in this study are at most a quarter of the length of a baseball. However, there are some specific instances when Statcast would miss a batted ball event. If a ball is struck as a weakly hit ground ball or a popfly that goes straight into the air, there is a chance that the cameras or doppler will disregard it and not count it as a hit. This has been corrected for by MLB's Senior Data Architect, Tom Tango (Tango, 2017). Although there are some minor measurement errors with Statcast, the data is still considered to be very reliable. This data is then uploaded on Baseball Savant every night during the season for public use. While the Statcast data is available for all to download, the amount of data that this study used was too much for Excel and SPSS to handle. For that reason, the language R was used to manage, manipulate and calculate all of the data and statistics in this study. R is a coding language that is designed to work with very large data sets, like the one in this study. R was created in the mid 1990's and began rapidly increasing in popularity in the early 2000's. Last year, R was the 3rd most popular language for data scientists behind Python and SQL (Hayes, 2019). Because of its popularity, flexibility and ability to handle large data sets, R was the best option for this study.

Data Collection Procedure

As previously mentioned, the data was pulled from MLB's official website baseballsavant.mlb.com. The site has a built-in search function that allows data to be filtered by various metrics. For this study, data was sorted by stadium and then downloaded into CSV files. This was done for both 2017 and 2018, the most recent years where there have been no significant changes to the batter's eye.

Coding for the batter's eye, however, was significantly different. Because design information on each eye is not widely available, we had to classify and code each characteristic, grouping together backgrounds that share common features. For color we broke them up into groups of black, green, gray and blue. Textures had the widest range of options including grass (artificial and real), ivy (artificial and real), trees and bushes, and no texture (padding, concrete, etc.). The variable Light was included as a dummy variable to indicate if there is a light source near the batter's eye. This light source could be from advertisements, electronic banners or stadium lights. Anything that could be seen as a distraction to the batter. Height and width, while not being subjective, are not publicly available for every stadium. Fortunately, the website Clem's Baseball Blog (andrewclem.com/Baseball), run by Dr. Andrew Clem at Central Virginia Community College, has scaled stadium maps and field dimensions of every stadium currently in use as well as historical ones. The width of some of the batter's eyes were available, so a ratio was able to be found between the map and the real stadiums. After measurements of each mapped batter's eye were taken, the ratio was applied to each width to find an estimate that could be used for each stadium. As for the height, the panel of experts were asked to estimate the height of the batter's eye based on the height of the fence. Dr. Clem's website also had information on the height of each fence, which was crosschecked with data from ESPN. As the panel surveyed each stadium, they were told the height of the centerfield fence and asked to estimate the height of the batter's eye. The mean of the estimations was used as the height. Although this method does suffer from some human subjectivity, it was the most feasible option that did not involve traveling to each stadium for direct measurement.

To counter some of the subjectivity, four experts were selected to help with the process of coding the batter's eye. They were selected based on their experience playing and following baseball as well as their knowledge of the batter's eye and ability to distinguish color. The first expert has coached both softball and baseball for at least 10 years and has followed baseball for over 60 years. The next expert played baseball for 15 years and currently is the manager and bullpen catcher for a major Division 1 baseball team. The third expert has more than 10 years of playing experience and has been involved with a Division 1 club baseball team, albeit a different one than the second expert. The final expert has worked for an independent baseball scouting network for the last four years and also has over a decade of playing experience. As previously mentioned, these experts were chosen to review the subjective characteristics of the batter's eyes and vote on how to classify each of them. The aim was to put each batter’s eye coding through a panel, to reduce some of the subjective biases that may arise with one single judge.

Data Management and Analysis

The data used in this study were collected and stored using Microsoft Excel 2016, and Rstudio where the multiple regressions were used to determine the relationship of the batter’s eye characteristics to hitting performance and decision making. This study used panel regression to test the characteristics of the backgrounds. The major assumptions of a regression are:

1. The linear regression model is “linear in parameters.”
2. There is a random sampling of observations.
3. The conditional mean for observation errors should be zero.
4. There is no multicollinearity.
5. There is homoscedasticity and no autocorrelation.

While we are not randomly selecting stadiums nor selecting random games or at-bats to use, we are using every pitch from 2017-2018 in all of the stadiums.

One way to test for multicollinearity utilized was to find the variance inflation factor (VIF) for the continuous variables used in the regressions. Although there is debate about what should be the cutoff for maximum allowable VIF, all of the continuous variables passed the multicollinearity assumption using a more conservative VIF of 5 (Henseler, Ringle, & Sarstedt, 2014). The highest VIF was reported by release speed at 3.8. All of the other continuous variables would pass the most conservative level of 2.5. All factors can be found in Appendix A. An annotated version of the programming code used in Rstudio to prepare the data and execute all of the regression estimation is provided in Appendix B.

Model Specifications

The first model that created the batter's decision of whether or not to swing at a pitch. This model is a logistic regression because there are only two choices, swing or not swing. Additionally, a random effects panel estimation was conducted to account for the repeated observations of individual players.

$$Log(Swing)\frac{Y_{it}}{1-Y_{it}} = \beta_{\theta} + \beta_1 Color + \beta_2 Texture + \beta_3 Shape + \beta_4 Light + \beta_5 Height + \beta_6 Width + \beta_7 Ball + \beta_8 PitchType + \varepsilon$$

Apart from the standard characteristics of the batter's eye, two different variables were added to this panel logit equation. Ball and pitch type were added so that they could be controlled for in the model. Batters will often look for certain pitches to hit and will try to swing at balls that are in, or close to, the strike zone. These two variables are probably the most important reasons for a batter to swing or not and to leave them out of the analysis would be detrimental to the study.

To determine if batter's eye characteristics make a difference in performance for balls that are hit, the model used the variables, color, height, width, texture, and a dummy variable for light in the area. The first regression for exit velocity was estimated as follows:

$$ExitVelo_{it} = \beta_\theta + \beta_1 Color_{it} + \beta_2 Texture_{it} + \beta_3 Light_{it} + \beta_4 Angle_{it} + \beta_5 Height + \beta_6 Width + \beta_7 PitchSpeed_{it} + \varepsilon$$

When looking at the deviation from mean launch angle, the same variables were used because we are still looking at the characteristics of the batter's eye and how they affect hit quality. The reason that the deviation from the mean angle is being used is because we are assuming that players are trying to hit the ball about the mean angle, 15 degrees. And that the larger the deviation from that mean, the more variability in the spread of launch angles, and the fewer times players will actually be optimizing their launch angle.

$$Launch\_Angle\_Deviation_{it} = \beta_\theta + \beta_1 Color_{it} + \beta_2 Texture_{it} + \beta_3 Light_{it} + \beta_4 Angle_{it} + \beta_5 Height + \beta_6 Width + \beta_7 PitchSpeed_{it} + \varepsilon$$Texture

Table 1: Coding Values

Variable	Coded 0	Coded 1	Coded 2	Coded 3	Coded 4	Coded 5
Color	Green	Black	Gray	Blue		
Texture	Grass	Ivy	Glass	Trees/Bushes	None	Other
Light	None	Nearby				
Angle	Flat	Angled	Mixed			

The reference variable for each category was the variable coded 0. There is a reference for each variable so that all of the coefficients can be compared. Pitch speed was added to these regressions to control for the speed of the pitched ball. It likely makes a difference in how hard and how well a ball can be hit and would otherwise be a confounding factor. First of all, in terms of the physics of exit velocity, the faster a ball approaches the batter (pitch speed), the slower the resultant exit velocity should be (Neiswender, 2019). In terms of both exit velocity and launch angle variation, higher pitch speeds would likely reduce processing, reaction and adjustment times for hitters. This may result in reduced exit velocities and increased variability in the launch angle outcomes.

Finally, the translation of batter's eye effects to runs was evaluated using a two-stage regression to see how many runs a batter's eye has cost or helped a few select teams. First, in order to find point estimates of the marginal relationships between launch angle, exit velocity and runs, the following prediction equations were estimated:

$$wOBA = \beta_\theta + \beta_1 ExitVelocity + \beta_2 DeviationFromMeanAngle + \varepsilon$$

And:

$$Stadium\ Runs = \beta_\theta + \beta_1 woba + \varepsilon$$

The *wOBA* regression was used to see how exit velocity and the deviation from the MLB mean launch angle, affected wOBA. Although these two factors were not very significant by regressing them directly to runs, by using wOBA as an intermediate step, the impact of these two factors can be better seen. *Stadium Runs* was then used to understand the relationship between wOBA specifically and runs scored in a stadium.

Chapter 4

Introduction

As previously stated, this study worked with pitch level data, at the Major League level, at 29 out of the 30 stadiums in the MLB. All pitches from the 2017 and 2018 season were pulled at the start of the study. Although instances such as bunts and hit by pitches were not included in the final results, over the course of the two seasons there were over a million pitches, across all qualified stadiums, evaluated in the study. The descriptive characteristics of the data among the variables of interest are shown in the table below.

Table 2: Descriptive Statistics

Continuous Variables	(n=1,419,133)		
Variable	Mean	Standard Deviation	Min/Max
Exit Velocity	82.219	14.748	(5.2, 122.2)
Launch Angle	15.878	27.583	(-89, 90)
Height	26.789	5.247	(15.5, 40.25)
Width	131.714	33.906	(72, 224)
Categorical Variables			
Variable	% of Total		
Green	55.1%		
Black	24.1%		
Gray	13.7%		
Blue	6.8%		
Grass	13.7%		
Ivy	10.3%		
Glass	0.0%		
Trees/Bushes	10.3%		
None	41.3%		
Other	24.1%		
No Light	62.1%		
Light Nearby	37.9%		
Flat	65.5%		
Angled	13.7%		
Mixed	20.6%		

Effects on Swing Decision

The first question that was posed was whether or not the batter's eye will affect a hitter's decision to swing. The variables that we were most concerned with were the color, texture, angle, height, width and if there was a light source in the area. Two other variables were added to help create a more robust model. The first was the zone in which ball crossed the plate. This was simply called ball. Although there are 14 zones that Statcast can put the location of the pitch in, this study only looked at clear balls and pitches close to or in the strike zone. After preliminary testing showed that there was a clear division between borderline pitches and clear balls, the variables were re-classified as such. The second added variable was the type of pitch thrown. These additional variables were included because of their perceived importance when hitters are at the plate. Most hitters will look for a certain type of pitch during an at-bat and most will be trying to hit strikes instead of balls.

Because there were so many variables used and the data set was so large (n=1,404,211), there was a problem in computing power. To work around the problem, a random sample of 110,000 observations were selected and tested. The sample used was still run using a panel regression and the number of variables was the largest amount pulled without running into more computing issues. The sample also contained all of the variations of batter’s eye characteristics.

Table 3: Swing Coefficients

Variable	Coefficient		Std. Error	Odds Ratio
Black	−0.841	***	0.083	0.431
Gray	0.766		0.105	
Blue	−0.069		0.121	
Ivy	0.081		0.140	
Trees/Bushes	−0.004		0.144	
None	0.043		0.171	
Other	0.012		0.111	
Angled	0.063		0.125	
Mixed	−0.011		0.089	
Light Nearby	−0.032		0.051	
Height	0.005		0.004	
Width	0.000		0.001	
Ball	1.783	***	0.043	5.947
Curveball	−0.701	***	0.099	0.496
Eephus	−0.761		1.331	
Cutter	−0.265	*	0.111	0.767
Four Seam Fastball	−0.469	***	0.074	0.625
Pitch Out	0.180		1.337	
Splitter	−0.063		0.177	
Two-Seam Fastball	−0.345	***	0.087	0.708
Knucklecurve	−0.426	**	0.154	0.653
Knuckleball	−0.313		0.501	
Sinker	−0.461	***	0.100	0.631
Slider	−0.285	***	0.082	0.752
Sigma	0.279	***	0.062	
Intercept	−0.841	***	0.208	
Observations	110,000			
Log-Likelihood:	-6666.622			

Note: *p<0.05; **p<0.01; ***p<0.001

The inclusion of the final two variables made up a vast majority of the significant factors on whether or not a batter will swing. These are not as important for determining how the batter's eye affects a batter's decision to swing but were used to understand what might cause a batter to swing at a pitch. The regression used to answer this question was a panel logit regression. Because of how the data were coded, the resulting coefficients were used to calculate the change in odds that a player will not swing at a pitch. Negative coefficients mean that they were more likely to swing while positive coefficients mean that players were less likely to swing. For the characteristics of the batter’s eye, only the color black had an effect on a batter’s swing probability. When the batter’s eye is black, batters are 2.3 times more likely to swing. There was no other characteristic of the batter’s eye that made a significant difference in the probability of a batter swinging at a pitch. The only other factors that were significant were pitch location and the type of pitch. It is unclear on if a change in swing probability is good or bad from the hitter’s perspective but more in-depth discussion of this will occur in the next chapter.

The largest impact on a hitter's decision to swing was the point at which the pitch crossed the plate, and the type of pitch thrown. When a ball was thrown clearly out of the strike zone, batters were 5.9 times less likely to swing. In this regression model, change-ups were used as the reference variable. When compared to the change-up, the two pitches that had the biggest impact on a batter’s decision to swing were curveballs and four-seam fastballs. Curveballs actually caused batters to swing slightly more than twice as much as change-ups. And when a fastball was thrown, batters were 1.6 times more likely to swing compared to a change-up.

Effects on Exit Velocity

The next question posed was what effect, if any, the batter’s eye characteristics have on exit velocity. In Table 4, we can see that the average exit velocity of each stadium is fairly stable.

Table 4: Average Exit Velocity & Launch Angle by Stadium

Team	Exit Velocity	Std Dev.	Launch Angle	Std. Dev.
ARI	82.868	15.828	12.837	30.244
ATL	82.006	15.499	14.229	29.804
BAL	82.792	15.498	16.789	29.527
CHC	82.350	15.744	16.098	27.504
CIN	80.710	15.930	15.619	29.017
CLE	82.580	15.689	15.293	28.951
COL	82.553	15.690	12.217	29.577
CWS	82.096	15.539	17.625	27.909
DET	82.764	15.349	18.733	27.402
HOU	82.322	15.364	15.017	30.733
KC	82.984	15.278	17.229	27.254
LAA	83.274	15.559	16.198	28.756
LAD	82.170	15.541	17.858	27.661
MIA	81.485	15.853	14.399	29.991
MIL	81.776	15.855	14.454	30.424
MIN	83.163	14.939	16.719	28.993
NYM	81.200	15.464	17.054	30.824
NYY	82.438	15.837	16.246	29.287
OAK	83.006	15.304	16.637	28.908
PHI	81.955	15.803	16.632	28.384
PIT	81.841	15.567	14.990	28.718
SD	81.994	15.522	13.586	31.504
SEA	82.502	15.637	16.712	28.417
SF	81.887	15.756	15.697	27.711
STL	81.817	15.965	15.727	28.069
TB	81.479	15.652	17.482	28.986
TEX	82.696	15.733	17.058	28.902
TOR	82.405	15.686	15.269	30.931
WSH	82.477	15.761	16.129	28.565

The fastest mean exit velocity was found in Anaheim at Angel Stadium, coming in at 83.27 mph (SD = 15.559) while the slowest was in Cincinnati at 80.71 mph SD = 15.930). The standard deviations of the exit velocities across the stadiums ranged from 14.9 in Minnesota to 15.9 in St. Louis.

After testing the model, it was found to have an adjusted R-squared value of 0.236 and a p-value of $p<0.001$. Although it only explains 23.6% of the variance in exit velocity, it does show that there is a significant relationship between the batter's eye types and exit velocity. The reference variable for color was green. Based off of the results in Table 5, it would appear that it is also the best color to have for the fastest exit velocity. Every color change tested had a negative coefficient compared to the reference with blue being the largest decrease at 1.28 mph. Meaning that if two batter's eyes were created equal, but one was blue and the other green, the hitters seeing the blue batter's eye would expect their mean exit velocity 1.28 mph slower. However, blue is not a common color in the MLB with only two teams, the Rays and Yankees, have a blue batter's eye. Black was the next worst at a decrease of 0.843 mph with gray being slightly better at 0.787 mph. The next characteristic to have an effect, and the only texture to be significantly different, was ivy. Grass was the reference variable for comparing textures of the batter's eye. Batter's eyes that used ivy instead of grass could were seen to be 0.419 mph slower simply because of that. As previously stated, there was no classification difference between real or artificial surfaces so this includes both natural and fake ivy. Unlike color and the last significant characteristic who had p-values near 0, ivy was significant at the α= .05 level. It is still worth noting that it is the only characteristic to affect speed and not be at the 0.001 level. The final variable to affect exit velocity was the width of the batter's eye. With a coefficient of 0.006 it may not seem like a very large change but it means that for every foot increase in width, the exit velocity is increasing by 0.006 mph. The widest batter's eye is at Comerica Park in Detroit while the narrowest is PNC Park's in Pittsburgh. With that difference we would expect to see a change in exit velocity by 0.912 mph between the two stadiums if everything else were equal.

Table 5: Exit Velocity & Launch Angle Coefficients

Variable	EV Coeff		Std. Error	DEV Coeff		Std. Error
Black	−0.843	***	0.106	0.416	***	0.116
Gray	−0.787	***	0.134	0.647	***	0.148
Blue	−1.286	***	0.160	0.193		0.176
Ivy	−0.419	*	0.176	0.778	***	0.194
Trees/Bushes	−0.264		0.184	0.925	***	0.202
None	0.351		0.216	0.645	***	0.238
Other	−0.084		0.140	0.523	***	0.154
Angled	−0.106		0.157	0.607	***	0.173
Mixed	−0.122		0.110	0.125		0.122
Light Nearby	0.090		0.066	0.204	***	0.073
Height	−0.007		0.006	0.012	*	0.007
Width	0.006	***	0.000	0.001		0.001
Release Speed	0.138	***	0.004	−0.086	***	0.004
Constant	67.819	***	0.465	30.512	***	0.510
Observations	400,002			400,026		
R^2	0.237			0.042		
Adjusted R^2	0.237			0.042		
F Statistic	123,245.200	***		17,121.590	***	

Note: *p<0.05; **p<0.01; ***p<0.001

Effects on Launch Angle

Unlike exit velocity, the mean launch angle at each stadium was a lot more volatile. It is generally accepted in the baseball community that batters want to hit the ball in a range between ten and thirty degrees and average between ten and twenty degrees (Ochart, 2018). This will give players more line drives and fly balls which tends produce a higher wOBA than groundballs and is highly correlated with more runs and offensive production. Although there is a range of 6.5 degrees between means of the lowest and highest stadiums, they all lie in this optimal zone for launch angle. So instead of using straight launch angle we decided to look at the mean angle for the entire MLB and then use the deviation of each stadium from that mean.

The panel regression provided evidence for many more characteristics affecting launch angle than there were affecting exit velocity. The overall model fit was a very low adjusted R-squared of 0.041 but the model was significant with a p-value of $p<0.001$. In total, there were nine factors that affected the deviation from the mean of launch angles. Black and gray backgrounds increased the deviation by 0.417 and 0.648 degrees respectively. Though blue had the worst effect on speed, it was not a significant variable in affecting the deviation from the league mean. Same as with exit velocity, grass was the reference variable when looking at texture. Compared to grass batter's eyes, all of the other textures had significantly larger deviations from the mean. Trees and bushes had the largest deviation at 0.922 degrees, almost a full degree more. This was followed by ivy, neutral texture and "mixed" textures. The degree of deviation of these textures ranged from 0.521 degrees to 0.778 degrees more than a grass covered background. Angled batter's eyes were also found to have a significant deviation from the mean when compared to flat (non-angled) batter's eyes. Angling the batter's eye had a substantial effect on deviation in launch angle at 0.605 degrees. The presence of light in the area was also found to increase the spread of launch angles by 0.203 degrees. So, just by placing a scoreboard or advertisement sign too close to the batter's eye, teams are hurting the run scoring environment of the stadium. The final characteristic that was significant was the height of the batter's eye. For every foot taller the batter's eye is, there is an increase in 0.012 degrees in deviation from the mean. Like with characteristics impacting exit velocity, there were a few variables that had a higher alpha level. Trees/bushes and the presence of a light source both had alpha levels of 0.001 and height was just significant with an alpha of 0.05.

Hit Quality to Runs

Although a variety of the characteristics of the batter's eye were statistically significant, with such a large dataset it is important to evaluate how meaningful the differences would be, as the effects appear to be very small. After all, the largest change for speed was 1.2 mph and for launch angle was 0.922 degrees. To try and give a better perspective on what the true impact is, we further analyzed the relationship between exit velocity, launch angle and runs scored. A comparative statistics approach (estimated marginal changes, all else being equal) was used to predict the impacts of proposed batter's eye design changes on performance outcomes (runs scored).

Table 7: Exit Velocity & Deviation vs. wOBA

	Dependent variable:	
	woba_value	
Exit Velocity	0.009	***
Dev. from Launch Angle	−0.0001	***
Constant	−0.276	***
Observations	250,846	
R^2	0.137	
Adjusted R^2	0.137	
F Statistic	39,791.880	***
Note:	$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$	

The first step was to find the relationship between wOBA and runs scored. To do that, the data were grouped by home team and inning. Due to the nature of the data, it was near impossible to pull the data for a specific team. This obstacle was resolved by grouping the data again, this time by the top and bottom of the inning. Because the home team always bats in the bottom of the inning, grouping the data by stadium and only using values from the bottom of the inning, we can be sure that we are only getting data from the home team in their home stadium. By doing that we were able to calculate the wOBA of each home team in their home stadium. Another problem with the data is that there is no obvious way to get the final score of each game.

Table 8: wOBA vs. Runs Scored at Home

	Dependent variable:	
	woba_value	
Observed wOBA	3071.736	***
Constant	−446.401	***
Observations	29	
R^2	0.637	
Adjusted R^2	0.623	
F Statistic	47.415***	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Fortunately, Fangraphs' leaderboard keeps track of the runs scored by the home team in each stadium. So, that data were pulled and used to regress wOBA with the runs scored by the home team to generate an xwOBA for prediction. The results show that wOBA alone accounts for 62.37% of the variance in runs scored. This correlation is also significant with a p-value of $p<0.001$. Next, we regressed exit velocity and deviation from the mean angle with the calculated wOBA to find that both variables were significant at the $p=0.001$ level. The two variables only explain 13.76% of the variance in wOBA with an adjusted R-squared value of 0.137 but it is also a significant model with a p-value of $p<0.001$. From these two regressions we are able to see how many runs could be gained or lost by each characteristic. And while some might understand the effect of just reporting exit velocity and deviation from the mean launch angle, reporting the differences in runs will give a more complete picture of the magnitude of this study. The practical marginal effects of the batter's eye characteristics to runs will be demonstrated more in the next chapter.

Chapter 5

The data revolution that started with Billy Beane and the Oakland Athletics is in full swing across the league and most sports. In 2015 when Statcast was installed in every MLB stadium, it fundamentally changed how the game of baseball was analyzed. Now, scouts and managers are able to see how players actually hit the ball. They can see how hard a player can hit and at what angle. They can see how fast players can run and how efficient they are at reading fly balls. Pitchers are able to know where they release the ball and if there is a difference in their release of a fastball or a curveball. They can see how fast each pitch spins and the amount of break that each pitch has. Most teams and players have embraced the use of technology in the game and it has led some of the best offences since the Steroid Era of the late 1990's and early 2000's ("Major League Baseball Batting", 2019). How the game is played has changed drastically, and it has only taken a few seasons.

In this push for better data, teams are trying to find the slightest edge over each other. This study seems to be right in line with the ever-expanding research that is going into baseball. With significant results being shown across all of the research questions, the batter's eye is one part of their park factors that they are able to adjust to try and influence the game. Park factors is a number that will typically explain how "pitcher-friendly" or "hitter-friendly" a park is. This includes many characteristics that cannot be changed like elevation and wind patterns and, stadium designs that might take a while to adjust like fence heights and distances. And finally, there are characteristics that could potentially be changed easily, like color and light sources of the batter's eye.

Discussion

Significant differences were found among the different characteristics of the batter's eye, but the interpretation of these differences can still be debated by people involved with baseball as well as those involved with this study. The first research question, does the batter's eye have an effect on the batter's decision to swing, has probably the biggest room for debate on the real effects that a batter's eye has. The biggest question being, is it better to have more swings, or fewer?

According to the results of this study, the color black, angled and mixed angled, light sources, and width of the batter's eye all have an effect that makes batters swing less. Mixed batter's eyes have a tendency to make batter's swing more. After controlling for zone and pitch type, how does the aggressiveness of swinging make a difference in the game? The answer to this question is unclear. Fewer swings could mean that a hitter is able to see the ball better and will wait longer to choose the pitch he wants, but the same could be said for swinging early. If a batter sees the ball well, he could like pitches earlier in the count and swing more at them. Anecdotally, there are players that are aggressive on the first pitch, players like three-time All-Star and former MVP, Jimmy Rollins or 2017 MVP, Jose Altuve, players who are objectively good, and aggressive at the plate. However, it is also possible to rarely swing at a pitch and be patient while still being a good hitter like Hall of Famer, Wade Boggs or six-time All-Star and former MVP Joe Mauer. Some research has been done to try and answer this age-old question, however, like most baseball research, this was done outside of academia. Over the last few seasons, the rate at which a majority of players have been swinging at the first pitch has increased, with some players swinging at 15% more first pitches than in previous years (Sullivan, 2015). However, other research has shown that taking the first pitch has a slightly higher wOBA than swinging at the first pitch (Gentile, 2013). While there is still more evidence that shows that batters, in general, are more successful if the fraction of first-pitch swings is relatively moderate (Albert, 2017).

Another relatively gray area is how the characteristics affect the launch angle. The results show that there are significant effects on the deviation from the mean launch angle. One of the major factors was texture. With every different texture being significantly different from the reference, it would appear that it has the largest impact on launch angle. Of the parts of the hitting process that were examined, the texture may be affecting the search strategies that hitters use. As previously mentioned, Joe Mauer had a huge problem with the trees that were planted at Target Field because of the shadows that they cast (Gleeman, 2011). The reason for this hatred may be because they distracted from his search strategies and so he was unable to pick up the ball until much later in the pitchers' delivery or the balls' flight path. Although we know that batters have excellent search strategies, any disruptions may play a role in changing their strategy (Kato & Fukuda, 2002).

Originally, this study was going to look specifically at the actual launch angle, not the deviation. While those results were similar to the final results, the question of whether or not they were actually effective arose. If an angled batter's eye increased the mean by a degree or so, while it would have an overall effect on run production, as long as the average stayed in the effective launch angle zone, the differences between a league average baseball hit at a degree difference would be negligible. Also, assuming that the stadium mean stayed within this effective zone, the argument could be made that although the difference is statistically significant, the impact of a few runs over the course of one or two seasons is inconsequential.

The findings for exit velocity are probably the easiest to translate to real changes. Color was seemed to have the largest impact on exit velocity. Based on previous research it appears that color shouldn't have much of an effect on hit quality but our results provide evidence to the contrary. Of the studies examined that focused on color, one found that color differences didn't make a difference and the other found that light colored balls on a light-colored background improved striking ability (Davis, 1978; Alomar, 1980). One thing that should be noted is that these studies were done 40 years ago. The technology these studies used, while fine at the time, is outdated at this point. It would be interesting, however, to redo these studies today using modern tracking systems like Statcast. While the results of these studies may not be completely accepted into conventional wisdom, what is generally accepted in baseball is that increased exit velocity leads to more offensive production. However, it is not necessarily a perfect correlation (Petriello, 2016). Over the last few years, there have been hundreds of pitches that were hit over 100 mph that were still caught. They could have been hit directly at fielders or fielders could have robbed a would-be home run. On the other end of the spectrum, there have been plenty of hits that have fallen in because they were hit hard enough to leave the infield but softly enough to fall in front of the outfielders. These are typically found between 50-60 mph (Zimmerman, 2018). The main difference between these hits is that the harder a ball is struck, the higher the xwOBA is and the more likely it is to produce more runs. An increase in exit velocity will probably lead to more home runs over the course of a season. Although the slowest home run hit during the Statcast era was 89 mph, the average exit velocity of a home run is 102.5 mph. So, if there is an increase in exit velocity, the probability of more runs being scored will increase dramatically. According to the Pythagorean expectation, a formula that predicts expected winning percentage based on run differentials, it takes an increase of about twelve runs, to increase winning percentage by one game, if runs against are held constant. Using this we can see how changes in exit velocity and launch angle can affect expected winning percentage.

To put this to the test, a few stadiums have been selected to partake in a hypothetical renovation. The first is Tropicana Field in Tampa Bay. The panel of experts coded this batter's eye as being a flat, blue, non-textured, no light sources nearby and about twenty-two feet tall. The Rays are a very analytically minded team. They have to be playing in one of the smaller markets and in a division that houses the Yankees and Red Sox, two of the most resource rich teams in all of baseball. They have decided that instead of spending a lot of money on a free agent they will give their stadium an offensive boost in a different way. Instead they will paint their batter's eye green. Based off of the results, they should see an increase in exit velocity of 1.2 mph with their deviation from the mean launch angle remaining the same. If we take their differences in exit velocity and use the regression to predict wOBA, we will see that there is a difference of 0.0011 points. This change in points is then plugged into the second regression to see how many runs it will increase by. After multiplying by the coefficient 3,071.73 and subtracting the intercept 446.4, we see that there is a point estimate difference of 34 runs over the course of the season. The Rays also believe that they have the best pitching staff in the league and that all of those runs will contribute directly to their team. If this were to happen, according to the Pythagorean expectation, they would expect an increase in winning percentage by 0.022 points, or about three wins.

Although this is a hypothetical situation, it is a stretch to assume that the Rays would be the only team to benefit from the boost in exit velocity. So, although they may not actually reap all of the benefits of the change in the batter's eye, an increase of 34 runs of total offense is a substantial change over the course of a season. But the Rays do have a very good pitching staff, so it might benefit them more than a team with a weaker staff.

The Cleveland Indians are another team that is considered to be a "small market" team. They have found recent success in being able to develop undervalued players better than almost anyone else. They are also one of the few teams that had a batter's eye that was in the trees/bushes category, as rated by this panel. After seeing this study and realizing that type of texture had the largest significant effect on the deviation from the mean launch angle, they decided to carefully rip the trees out of centerfield and re-plant them across the street; leaving a vertical wall that can be covered by green artificial turf. All of the other characteristics remain the same for the Indians, except now, instead of trees the texture of their batter's eye is grass, the best texture to have. This time, instead of adding to the exit velocity, we will subtract from the deviation from the mean launch angle and hold exit velocity constant. We then find that there is an increase in calculated wOBA of 0.0001 points. After plugging this into the second regression we find that the change in wOBA has resulted in an increase of run production by 0.307 runs per season. According to the Pythagorean expectation if the Indians were to add an additional 0.307 runs while holding their opponents constant they would expect virtually no change in winning percentage and no difference in the win column.

This does appear to make some sense though, and it relates to the nature of launch angle. Because launch angle is a range, that means there is more than one "optimal" launch angle. A batter can get a home run off of a pitch that is 18 degrees off the bat or 35 degrees off the bat. As long as the average angle stays within the optimal range, the effects are going to be significantly smaller than that of launch speed.

Recommendations for practice

If someone were building a new stadium or renovating their current batter's eye and wanted to build one that would help offenses the most, they would want a batter's eye that is green, with either artificial or real grass, that is not angled, is as wide as possible, and does not have any light sources near it. This may not be the most aesthetically pleasing batter's eye, but it would have the largest positive impact on exit velocity and the least deviation from the mean launch angle. On the other hand, if a team wanted to try and suppress runs in the stadium, they should have a blue ivy or bushy batter's eye that is angled, as narrow as possible with a scoreboard or advertisement directly above it. If the team wanted to suppress speed, they would want blue ivy, while if they wanted to increase the deviation in launch angle, they could have blue bushes or trees. But this example batter's eye, apart from being unique, would allow a team to suppress the run production in the stadium as much as possible.

Future research

Though the research questions may have been answered, this project is far from answering everything discussed. As mentioned in passing before, there are still plenty of research opportunities on several aspects of this project.

The first one pertains to another way of testing the questions posed in this study. It arose as MLB’s annual Home Run Derby nears. The Home Run Derby is an event that happens around the All-Star Game every year in which some of the best home run hitters gather to try and hit as many home runs as they can over a set amount of time. The batters are obviously swinging for the fences every time, and the pitches thrown are nothing more than batting practice speeds. So, if the data were available, it could be interesting to see how exit velocity and launch angle are being affected when many other variables are inherently being controlled for. Also using this data, we would also know that players are trying to maximize their exit velocity and hit optimal launch angles.

Future testing can be done to see if the batter's eye has an effect on a hitter's decision to swing at certain types of pitches in specific zones. For example, although black batter’s eyes show an increase in swings, do they increase swings on sliders? To take this one step further, analysis may also be done on just the pitches to see how they can affect a batter's decision to swing. It may also be noteworthy to see what the change is by making fastballs, instead of changeups, the reference condition because they are the most common pitch.

There can also be testing on launch angles. Although there is a range of "optimal" launch angles, is there one angle that is generally better than the rest? There has been some research that has tried to boil the range down to one angle; however, like most of baseball, the answer depended on the type of hitter (Zimmerman, 2018). The answer that Zimmerman came up with could be different depending on the era played. Although the 2018 season reflects current trends very well, this question could be re-examined in the future or if there are major rule changes imposed by the commissioner. In addition to that, further testing on batter's eye characteristic's effect on launch angle could be done to determine what might happen at each stadium if the average angle changed by a significant amount.

One question that could be interesting as future research revolves around the idea of plate discipline. Is it better to swing more or be patient at the plate? This question stems from the first research question posed in this study. Although it is just beyond the scope of this project, if the data are available, it would make sense to try and see how plate appearances end when batters swing only once versus multiple times. This project may go a similar course as this one where they try and relate it back to runs scored, or it could be more similar to the Gentile (2013) study that looked specifically at increases in wOBA. However, while this study may capture the general trend of plate appearances, it is worth noting that every player is different and some excel swinging early in the count while others can excel by waiting and taking pitches.

The final idea for future research comes from looking at the descriptive statistics for each of the stadiums. The two stadiums that had the lowest mean launch angle were Chase Field in Phoenix, Arizona and Coors Field in Denver, Colorado. These two stadiums had the lowest launch angle by almost a full degree and were off of the MLB average by over three degrees. What is striking about them is that they are also the only teams to store their game balls in a humidor. Arizona made this decision because of the dryness of the air and climate and, Colorado because of their extreme elevation. Although this could just be a coincidence, it is striking that the only stadium in the National League West Division with an above average launch angle, is Dodger Stadium. So, it could make for an interesting study to see if this trend has continued further past and if so, what is causing the launch angles to be so drastically low in Arizona and Colorado?

Conclusion

At the beginning of this study, four questions were posed. Do characteristics of a batter's eye affect a batter's decision to swing? Do the characteristics affect exit velocity? Do they affect launch angle? And do they affect run production in a stadium? The answer to all of these questions turned out to be yes, to some degree. According to the data, at least one color, one varied texture, the angle, light sources and the height of the batter's eye all appear to affect the batter's decision to swing. Although the characteristics were not the largest factor in a batter's decision, they were significant. All of the colors, one texture, and the width of the batter's eye also appeared to affect the exit velocity of the ball. While color, all of the textures, presence of light and the height had a significant impact on the deviation from the mean launch angle. Finally, because the batter's eye affected these measures of hit quality, it is safe to say that the batter's eye did have an effect on run production in the stadiums tested. No team had a "perfect" batter's eye, and the few examples did show that there were potential runs not scored in the stadium simply based off of the characteristics studied. So, the next time you go to a baseball game, look towards centerfield. What do you see? Because what you see will likely have an impact on the game you are about to watch.

References

Akaike, K., Fumoto, M., & Usui, S. (2018). The Effect of Initial Head Movement on Baseball Hitting Accuracy. *Advances in Exercise and Sports Psychology,24*(2), 27-38.

Albert, J. (2017, March 29). Swinging on the First Pitch. Retrieved from https://baseballwithr.wordpress.com/2017/03/29/swinging-on-the-first-pitch/

Anelso13. (2016, September 13). Is Exit Velocity Important? Retrieved March 12, 2018, from https://community.fangraphs.com/is-exit-velocity-important/

Beneventano, P., Berger, P., & Weinberg, B. (2012). Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics. *International Journal of Business, Humanities and Technology,2*(4).

Boden, L. M., Rosengren, K. J., Martin, D. F., & Boden, S. D. (2009). A comparison of static near stereo acuity in youth baseball/softball players and non–ball players. *Optometry - Journal of the American Optometric Association,80*(3), 121-125. doi:10.1016/j.optm.2008.06.009

Borzi, P. (2008, June 1). Swing Science: The Fine Points. *The New York Times*. Retrieved March 12, 2018.

Calcaterra, C. (2018, October 25). Oh no, Baseball is Dying Again! Retrieved from https://mlb.nbcsports.com/2018/10/25/oh-no-baseball-is-dying-again/

Davis, D. R. (1978). *The effects of yellow, orange and white baseballs upon the visual perception and hitting effectiveness of college baseball players* (Unpublished doctoral dissertation). Middle Tennessee State University.

Gentile, J. (2013, August 23). Swinging or taking on the first pitch. Retrieved from https://tht.fangraphs.com/swinging-or-taking-on-the-first-pitch/

Gleeman, A. (2011, January 31). Hitters complained, so the Twins are getting rid of the center field trees at Target Field. Retrieved from https://mlb.nbcsports.com/2011/01/31/hitters-complained-so-the-twins-are-getting-rid-of-the-center-field-trees-at-target-field/

Gray, R., & Cañal-Bruland, R. (2018). Integrating visual trajectory and probabilistic information in baseball batting. *Psychology of Sport and Exercise,36*, 123-131. doi:10.1016/j.psychsport.2018.02.009

Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Dodeling. *Journal of the Academy of Marketing Science,43*(1), 115-135. doi:10.1007/s11747-014-0403-8

Higuchi, T., Nagami, T., Morohoshi, J., Nakata, H., & Kanosue, K. (2013). Disturbance In Hitting Accuracy By Professional And Collegiate Baseball Players Due To Intentional Change Of Target Position1,2. *Perceptual and Motor Skills*. doi:10.2466/30.23.25.pms.116.2

Holmes, D. (2017, March 22). What is the batter's eye? Retrieved from https://www.detroitathletic.com/blog/2017/03/22/what-is-the-batters-eye/

Jaffe, J. (2019, June 27). We've Reached the Point of "Too Many Homers". Retrieved from https://blogs.fangraphs.com/weve-reached-the-point-of-too-many-homers/

Kato, T., & Fukuda, T. (2002). Visual Search Strategies of Baseball Batters: Eye Movements during the Preparatory Phase of Batting. *Perceptual and Motor Skills,94*(2), 380-386. doi:10.2466/pms.2002.94.2.380

Laby, D. M., Davidson, J. L., Rosenbaum, L. J., Strasser, C., Mellman, M. F., Rosenbaum, A. L., & Kirschen, D. G. (1996). The Visual Function of Professional Baseball Players. *American Journal of Ophthalmology,122*(4), 476-485. doi:10.1016/s0002-9394(14)72106-3

Lewis, M. (2011). *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton.

Major League Baseball Batting Year-by-Year Averages. (2019). Retrieved from https://www.baseball-reference.com/leagues/MLB/bat.shtml

Morris-Binelli, K., Müller, S., & Fadde, P. (2018). Use of Pitcher Game Footage to Measure Visual Anticipation and Its Relationship to Baseball Batting Statistics. *Journal of Motor Learning and Development,6*(2), 197-208. doi:10.1123/jmld.2017-0015

Müller, S., & Abernethy, B. (2012). Expert Anticipatory Skill in Striking Sports. *Research Quarterly for Exercise and Sport,83*(2), 175-187. doi:10.1080/02701367.2012.10599848

Müller, S., & Fadde, P. J. (2015). The Relationship Between Visual Anticipation and Baseball Batting Game Statistics. *Journal of Applied Sport Psychology,28*(1), 49-61. doi:10.1080/10413200.2015.1058867

Nakamoto, H., & Mori, S. (2008). Sport-Specific Decision-Making in a Go/Nogo Reaction Task: Difference among Nonathletes and Baseball and Basketball Players. *Perceptual and Motor Skills,106*(1), 163-170. doi:10.2466/pms.106.1.163-170

Nakamoto, H., & Mori, S. (2012). Experts in fast-ball sports reduce anticipation timing cost by developing inhibitory control. *Brain and Cognition,80*(1), 23-32. doi:10.1016/j.bandc.2012.04.004

Neiswender, K. (2019, May 17). Debunking Bat Speed Myths. Retrieved from https://www.drivelinebaseball.com/2019/05/debunking-bat-speed-myths/

Ochart, J. (2018, May 30). Using Swing Plane to Coach Hitters: A Deeper Look. Retrieved from https://www.drivelinebaseball.com/2018/05/using-swing-plane-coach-hitters-deeper-look/

Petriello, M. (2016, January 10). Statcast exit velocity correlates to success. Retrieved from https://www.mlb.com/news/statcast-exit-velocity-correlates-to-success/c-161576128

RotoExperts. (2016, November 26). Slugging Percentage on Batted Balls [Chart]. In *RotoExperts*. Retrieved from https://rotoexperts.com/khris-davis-kendrys-morales-will-continue-mashing-2017/

Regan, D. (1997). Visual factors in hitting and catching. *Journal of Sports Sciences,15*(6), 533-558. doi:10.1080/026404197366985

Renshaw, I., & Fairweather, M. M. (2000). Cricket bowling deliveries and the discrimination ability of professional and amateur batters. *Journal of Sports Sciences,18*(12), 951-957. doi:10.1080/026404100446757

Sarris, E. (2016, September 15). What Can Hitters Actually See Out of a Pitcher's Hand? Retrieved from https://blogs.fangraphs.com/what-can-hitters-actually-see-out-of-a-pitchers-hand/

Schifman, G. (2018, March 27). The Lurking Error in Statcast Pitch Data. Retrieved from https://tht.fangraphs.com/the-lurking-error-in-statcast-pitch-data/

Shank, M. D., & Haywood, K. M. (1987). Eye Movements While Viewing a Baseball Pitch. *Perceptual and Motor Skills,64*, 1191-1197. doi:10.2466/pms.1987.64.3c.1191

Sullivan, J. (2015, September 30). Return of the First-Pitch Swing. Retrieved from https://blogs.fangraphs.com/return-of-the-first-pitch-swing/

Sullivan, J. (2018, September 12). There's Definitely Something Strange About Citi Field. Retrieved from https://blogs.fangraphs.com/theres-definitely-something-strange-about-citi-field/

Tango, T. (2017, April 21). No Nulls in Batted Balls Launch Parameters. Retrieved from http://tangotiger.com/index.php/site/comments/statcast-lab-no-nulls-in-batted-balls-launch-parameters

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2010). Perceptual-cognitive expertise in sport and its acquisition: Implications for applied cognitive psychology. *Applied Cognitive Psychology,25*(3), 432-442. doi:10.1002/acp.1710

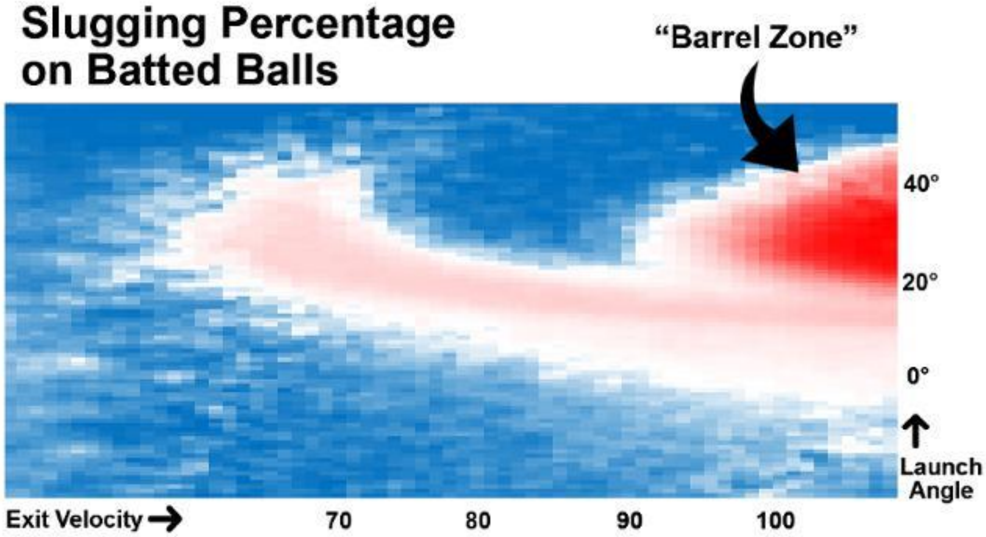
Zimmerman, J. (2018, April 16). Finding a Simple Ideal Launch Angle. Retrieved from https://fantasy.fangraphs.com/finding-a-simple-ideal-launch-angle/

Appendix A: Tables & Figures

Table 10: VIF Coefficients

Variable	VIF	
Black	2.773	
Gray	2.906	
Blue	2.073	
Ivy	3.915	
Trees/Bushes	4.172	
None	15.128	***
Other	5.064	
Angled	4.027	
Mixed	2.971	
Light Nearby	1.377	
Height	1.413	
Width	1.323	
Release Speed	3.853	
Ball	1.008	
Curveball	1.986	
Eephus	1.011	
Cutter	1.551	
Four Seam Fastball	4.889	
Pitch Out	1.003	
Splitter	1.134	
Two-Seam Fastball	2.713	
Knucklecurve	1.241	
Knuckleball	1.032	
Sinker	2.031	
Slider	2.208	
Exit Velocity	1.296	
Dev. From Opt. Angle	1.296	

The variable for texture "None" did have a significantly higher VIF than the maximum allowable level. However, it is a categorical variable and there is no reason that we could tell for it to have such a high VIF.



A heat map like this can be used to calculate xwoba (RotoExperts, 2016).

Appendix B: Code

Data Prep Code

```
library(tidyverse)

library(magrittr)

library(dplyr)

library(ggplot2)

library(skimr)

library(caret)

library(caTools)

library(plm)


#After downloading the data from baseball-savant you will have to edit your file path.

r2017 <- read_csv("C:\\Users\\NewUser\\Desktop\\Thesis_Data\\2017_total.csv")

r2018 <- read_csv("C:\\Users\\NewUser\\Desktop\\Thesis_Data\\2018_total.csv")


#Sometimes my data would double in columns. There should be 89 columns before binding the .csv files.

total_2018 <- r2018[,-c(90:177)]

total_2017 <- r2017[,-c(90:177)]


#These caused errors when trying to bind the files. May not be necessary for you.

total_2017%<>% mutate(game_date = as.character(game_date),release_pos_x =
as.numeric(release_pos_x),release_pos_z = as.numeric(release_pos_z),zone = as.numeric(zone))

total_2017%<>% mutate(pfx_x = as.numeric(pfx_x),pfx_z = as.numeric(pfx_z),plate_x =
as.numeric(plate_x),plate_z = as.numeric(plate_z))

total_2017%<>% mutate(vx0 = as.numeric(vx0),vy0 = as.numeric(vy0),vz0 = as.numeric(vz0),ax =
as.numeric(ax),ay = as.numeric(ay),az = as.numeric(az))

total_2017%<>% mutate(effective_speed = as.numeric(effective_speed),release_spin_rate =
as.numeric(release_spin_rate))

total_2017%<>% mutate(release_pos_y = as.numeric(release_pos_y))

total_2018%<>% mutate(release_spin_rate = as.numeric(release_spin_rate))


total <- bind_rows(total_2017,total_2018)
```

```
#Add the columns for batter's eye characteristics
total$Color<-NA
total$Texture<-NA
total$Light<-NA
total$Angle<-NA
total$Height<-NA
total$Width<-NA
total$Pitch_ID <- NA

#Gives an ID to each pitch seen by the batter and is unique to each batter
total %<>% group_by(batter) %>% mutate(Pitch_ID = row_number())

#Variables are described in the paper.
total%<>% mutate(Color = replace(Color,home_team == "LAA", 0))
total%<>% mutate(Color = replace(Color,home_team == "OAK", 0))
total%<>% mutate(Color = replace(Color,home_team == "HOU", 0))
total%<>% mutate(Color = replace(Color,home_team == "TOR", 1))
total%<>% mutate(Color = replace(Color,home_team == "ATL", 0))
total%<>% mutate(Color = replace(Color,home_team == "MIL", 1))
total%<>% mutate(Color = replace(Color,home_team == "STL", 0))
total%<>% mutate(Color = replace(Color,home_team == "CHC", 0))
total%<>% mutate(Color = replace(Color,home_team == "ARI", 0))
total%<>% mutate(Color = replace(Color,home_team == "LAD", 1))
total%<>% mutate(Color = replace(Color,home_team == "SF", 2))
total%<>% mutate(Color = replace(Color,home_team == "CLE", 0))
total%<>% mutate(Color = replace(Color,home_team == "SEA", 1))
total%<>% mutate(Color = replace(Color,home_team == "MIA", 1))
total%<>% mutate(Color = replace(Color,home_team == "NYM", 2))
total%<>% mutate(Color = replace(Color,home_team == "WSH", 0))
total%<>% mutate(Color = replace(Color,home_team == "BAL", 2))
total%<>% mutate(Color = replace(Color,home_team == "SD", 2))
total%<>% mutate(Color = replace(Color,home_team == "PHI", 0))
total%<>% mutate(Color = replace(Color,home_team == "PIT", 0))
```

```
total%<>% mutate(Color = replace(Color,home_team == "TEX", 0))
total%<>% mutate(Color = replace(Color,home_team == "TB", 3))
total%<>% mutate(Color = replace(Color,home_team == "CIN", 1))
total%<>% mutate(Color = replace(Color,home_team == "COL", 0))
total%<>% mutate(Color = replace(Color,home_team == "KC", 0))
total%<>% mutate(Color = replace(Color,home_team == "DET", 0))
total%<>% mutate(Color = replace(Color,home_team == "MIN", 1))
total%<>% mutate(Color = replace(Color,home_team == "CWS", 0))
total%<>% mutate(Color = replace(Color,home_team == "NYY", 3))

total%<>% mutate(Texture = replace(Texture,home_team == "LAA", 0))
total%<>% mutate(Texture = replace(Texture,home_team == "OAK", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "HOU", 1))
total%<>% mutate(Texture = replace(Texture,home_team == "TOR", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "ATL", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "MIL", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "STL", 0))
total%<>% mutate(Texture = replace(Texture,home_team == "CHC", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "ARI", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "LAD", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "SF", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "CLE", 3))
total%<>% mutate(Texture = replace(Texture,home_team == "SEA", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "MIA", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "NYM", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "WSH", 0))
total%<>% mutate(Texture = replace(Texture,home_team == "BAL", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "SD", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "PHI", 1))
total%<>% mutate(Texture = replace(Texture,home_team == "PIT", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "TEX", 0))
total%<>% mutate(Texture = replace(Texture,home_team == "TB", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "CIN",5))
```

```
total%<>% mutate(Texture = replace(Texture,home_team == "COL", 3))
total%<>% mutate(Texture = replace(Texture,home_team == "KC", 5))
total%<>% mutate(Texture = replace(Texture,home_team == "DET", 1))
total%<>% mutate(Texture = replace(Texture,home_team == "MIN", 4))
total%<>% mutate(Texture = replace(Texture,home_team == "CWS", 3))
total%<>% mutate(Texture = replace(Texture,home_team == "NYY", 4))

total%<>% mutate(Angle = replace(Angle,home_team == "LAA", 1))
total%<>% mutate(Angle = replace(Angle,home_team == "OAK", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "HOU", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "TOR", 1))
total%<>% mutate(Angle = replace(Angle,home_team == "ATL", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "MIL", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "STL", 1))
total%<>% mutate(Angle = replace(Angle,home_team == "CHC", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "ARI", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "LAD", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "SF", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "CLE", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "SEA", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "MIA", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "NYM", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "WSH", 1))
total%<>% mutate(Angle = replace(Angle,home_team == "BAL", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "SD", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "PHI", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "PIT", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "TEX", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "TB", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "CIN", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "COL", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "KC", 2))
total%<>% mutate(Angle = replace(Angle,home_team == "DET", 0))
```

```
total%<>% mutate(Angle = replace(Angle,home_team == "MIN", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "CWS", 0))
total%<>% mutate(Angle = replace(Angle,home_team == "NYY", 0))

total%<>% mutate(Light = replace(Light,home_team == "LAA", 0))
total%<>% mutate(Light = replace(Light,home_team == "OAK", 0))
total%<>% mutate(Light = replace(Light,home_team == "HOU", 1))
total%<>% mutate(Light = replace(Light,home_team == "TOR", 1))
total%<>% mutate(Light = replace(Light,home_team == "ATL", 1))
total%<>% mutate(Light = replace(Light,home_team == "MIL", 1))
total%<>% mutate(Light = replace(Light,home_team == "STL", 0))
total%<>% mutate(Light = replace(Light,home_team == "CHC", 0))
total%<>% mutate(Light = replace(Light,home_team == "ARI", 1))
total%<>% mutate(Light = replace(Light,home_team == "LAD", 0))
total%<>% mutate(Light = replace(Light,home_team == "SF", 1))
total%<>% mutate(Light = replace(Light,home_team == "CLE",0))
total%<>% mutate(Light = replace(Light,home_team == "SEA", 0))
total%<>% mutate(Light = replace(Light,home_team == "MIA", 1))
total%<>% mutate(Light = replace(Light,home_team == "NYM", 0))
total%<>% mutate(Light = replace(Light,home_team == "WSH", 0))
total%<>% mutate(Light = replace(Light,home_team == "BAL", 1))
total%<>% mutate(Light = replace(Light,home_team == "SD", 0))
total%<>% mutate(Light = replace(Light,home_team == "PHI", 0))
total%<>% mutate(Light = replace(Light,home_team == "PIT", 0))
total%<>% mutate(Light = replace(Light,home_team == "TEX", 0))
total%<>% mutate(Light = replace(Light,home_team == "TB", 0))
total%<>% mutate(Light = replace(Light,home_team == "CIN", 0))
total%<>% mutate(Light = replace(Light,home_team == "COL", 0))
total%<>% mutate(Light = replace(Light,home_team == "KC", 1))
total%<>% mutate(Light = replace(Light,home_team == "DET", 0))
total%<>% mutate(Light = replace(Light,home_team == "MIN", 0))
total%<>% mutate(Light = replace(Light,home_team == "CWS", 1))
total%<>% mutate(Light = replace(Light,home_team == "NYY", 1))
```

```
total%<>% mutate(Width = replace(Width,home_team == "LAA", 192))
total%<>% mutate(Width = replace(Width,home_team == "OAK", 208))
total%<>% mutate(Width = replace(Width,home_team == "HOU", 80))
total%<>% mutate(Width = replace(Width,home_team == "TOR", 144))
total%<>% mutate(Width = replace(Width,home_team == "ATL", 152))
total%<>% mutate(Width = replace(Width,home_team == "MIL", 120))
total%<>% mutate(Width = replace(Width,home_team == "STL", 120))
total%<>% mutate(Width = replace(Width,home_team == "CHC", 96))
total%<>% mutate(Width = replace(Width,home_team == "ARI", 176))
total%<>% mutate(Width = replace(Width,home_team == "LAD", 104))
total%<>% mutate(Width = replace(Width,home_team == "SF", 104))
total%<>% mutate(Width = replace(Width,home_team == "CLE", 104))
total%<>% mutate(Width = replace(Width,home_team == "SEA", 120))
total%<>% mutate(Width = replace(Width,home_team == "MIA", 112))
total%<>% mutate(Width = replace(Width,home_team == "NYM", 120))
total%<>% mutate(Width = replace(Width,home_team == "WSH", 112))
total%<>% mutate(Width = replace(Width,home_team == "BAL", 136))
total%<>% mutate(Width = replace(Width,home_team == "SD", 128))
total%<>% mutate(Width = replace(Width,home_team == "PHI", 120))
total%<>% mutate(Width = replace(Width,home_team == "PIT", 72))
total%<>% mutate(Width = replace(Width,home_team == "TEX", 128))
total%<>% mutate(Width = replace(Width,home_team == "TB", 128))
total%<>% mutate(Width = replace(Width,home_team == "CIN", 152))
total%<>% mutate(Width = replace(Width,home_team == "COL", 136))
total%<>% mutate(Width = replace(Width,home_team == "KC", 128))
total%<>% mutate(Width = replace(Width,home_team == "DET", 224))
total%<>% mutate(Width = replace(Width,home_team == "MIN", 120))
total%<>% mutate(Width = replace(Width,home_team == "CWS", 112))
total%<>% mutate(Width = replace(Width,home_team == "NYY", 144))

total%<>% mutate(Height = replace(Height,home_team == "LAA", 26.25))
total%<>% mutate(Height = replace(Height,home_team == "OAK", 32.75))
```

```
total%<>% mutate(Height = replace(Height,home_team == "HOU", 29.13))
total%<>% mutate(Height = replace(Height,home_team == "TOR", 17.67))
total%<>% mutate(Height = replace(Height,home_team == "ATL", 25))
total%<>% mutate(Height = replace(Height,home_team == "MIL", 40.25))
total%<>% mutate(Height = replace(Height,home_team == "STL", 30.25))
total%<>% mutate(Height = replace(Height,home_team == "CHC", 23.33))
total%<>% mutate(Height = replace(Height,home_team == "ARI", 33))
total%<>% mutate(Height = replace(Height,home_team == "LAD", 30.33))
total%<>% mutate(Height = replace(Height,home_team == "SF", 30))
total%<>% mutate(Height = replace(Height,home_team == "CLE", 15.5))
total%<>% mutate(Height = replace(Height,home_team == "SEA", 16.83))
total%<>% mutate(Height = replace(Height,home_team == "MIA", 21.33))
total%<>% mutate(Height = replace(Height,home_team == "NYM", 27.33))
total%<>% mutate(Height = replace(Height,home_team == "WSH", 20.67))
total%<>% mutate(Height = replace(Height,home_team == "BAL", 27.67))
total%<>% mutate(Height = replace(Height,home_team == "SD", 30.67))
total%<>% mutate(Height = replace(Height,home_team == "PHI", 26.67))
total%<>% mutate(Height = replace(Height,home_team == "PIT", 29.17))
total%<>% mutate(Height = replace(Height,home_team == "TEX", 22.67))
total%<>% mutate(Height = replace(Height,home_team == "TB", 21.67))
total%<>% mutate(Height = replace(Height,home_team == "CIN", 26.33))
total%<>% mutate(Height = replace(Height,home_team == "COL", 28.33))
total%<>% mutate(Height = replace(Height,home_team == "KC", 31.67))
total%<>% mutate(Height = replace(Height,home_team == "DET", 30.67))
total%<>% mutate(Height = replace(Height,home_team == "MIN", 28))
total%<>% mutate(Height = replace(Height,home_team == "CWS", 28))
total%<>% mutate(Height = replace(Height,home_team == "NYY", 27))

total%<>% mutate(Light = as.factor(Light), Height = as.numeric(Height), Width = as.numeric(Width))
total%<>% mutate(launch_speed = as.numeric(launch_speed), launch_angle = as.numeric(launch_angle))
total%<>% mutate(Color = as.factor(Color), Texture = as.factor(Texture), Angle = as.factor(Angle))
total$optimal_angle<- NA
total%<>% mutate(optimal_angle = abs(launch_angle-15.878))
```

```
total$new_la <- NA

total%<>% mutate(new_la = (launch_angle-15.878)^2)
```

Logistic Regression for Swing Decision

```
library(pglm)

#You will need to do some data manipulation but not as much as the launch_speed and launch_angle research questions.

swing <- total %>% select(batter, Pitch_ID, Color, Angle, Texture, Light, Height, Width, pitch_type, release_speed, zone, game_year, description, home_team)

swing %<>% ungroup(batter)

swing %<>% mutate(Pitch_ID = as.integer(Pitch_ID), batter = as.integer(batter))

#I am leaving out bunts, pitchouts and HBP. You can leave them in if you want.

swing$swing_event<-NA

swing%<>% mutate(swing_event = replace(swing_event,description == "ball", 0))

swing%<>% mutate(swing_event = replace(swing_event,description == "blocked_ball", 0))

swing%<>% mutate(swing_event = replace(swing_event,description == "called_strike", 0))

swing%<>% mutate(swing_event = replace(swing_event,description == "foul", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "foul_bunt", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "foul_pitchout", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "foul_tip", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "hit_by_pitch", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "hit_into_play", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "hit_into_play_no_out", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "hit_into_play_score", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "missed_bunt", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "pitchout", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "swinging_pitchout", NA))

swing%<>% mutate(swing_event = replace(swing_event,description == "swinging_strike", 1))

swing%<>% mutate(swing_event = replace(swing_event,description == "swinging_strike_blocked", 1))


swing$ball <- NA

swing%<>% mutate(ball  = replace(ball ,zone >10, 0))
```

```
swing%<>% mutate(ball  = replace(ball ,zone <10, 1))


swing2 <- swing %>% select (Color, Texture, Angle, Light, Height, Width, ball, home_team,
release_speed, pitch_type, Pitch_ID, batter, Pitch_ID, swing_event, )

swing2 %<>% mutate(ball = as.factor(ball))

swing2 <- na.omit(swing2)


swing_test <- sample_n(swing2, 110000)



swing_prob <- pglm(swing_event ~ Color + Texture + Angle + Light + Height + Width + ball +
pitch_type + release_speed,

   data = swing_test,  family = binomial('logit'), model = "random", index = c("batter", "pitch_id")

summary(swing_prob)
```

<u>Panel Regression for EV & LA</u>

```
total2 <- total %>% select(launch_speed,new_la, launch_angle, optimal_angle, release_speed, batter,
Pitch_ID, Color, Angle, Texture, Light, Height, Width)

total2 %<>% ungroup(batter)

total2 %<>% mutate(Pitch_ID = as.integer(Pitch_ID), batter = as.integer(batter))


random_speed <- plm(launch_speed ~ Color + Texture + Angle + Light + Height + Width +
release_speed, data = total2, index = c("batter", "Pitch_ID"), model = "random")

summary(random_speed)

random_angle2 <- plm(optimal_angle ~ Color + Texture + Angle + Light + Height + Width +
release_speed, data = total2, index = c("batter", "Pitch_ID"), model = "random")

summary(random_angle2)
```

<u>Regressions for EV & LA vs. wOBA and wOBA vs. Runs</u>

```
test <- total %>% select(launch_speed, launch_angle, new_la, home_team, woba_value,
estimated_woba_using_speedangle, batter, Pitch_ID)


test %<>% mutate(woba_value = as.numeric(woba_value))

test %<>% ungroup(batter)

test %<>% mutate(Pitch_ID = as.integer(Pitch_ID), batter = as.integer(batter))
```

```
panel_woba <- plm(woba_value ~ launch_speed + new_la, data = test, index = c("batter", "Pitch_ID"), model = "random")

summary(panel_woba)

stargazer::stargazer(panel_woba)

ols_woba <- lm(woba_value ~ launch_speed + new_la, data = test)

summary(ols_woba)


total4 <- total %>% select(launch_speed, launch_angle, woba_value, woba_denom, inning_topbot, home_team, batter)


total4 %<>% mutate(woba_value = as.numeric(woba_value), woba_denom = as.numeric(woba_denom))


total4 <- na.omit(total4)


woba_d <- total4 %<>%group_by(home_team,inning_topbot)%<>%summarize(woba_d = mean(woba_denom))

view(woba_d)


team_woba <- total4 %<>%group_by(home_team,inning_topbot)%<>%summarize(woba_value = mean(woba_value))

view(team_woba)
```