

Measuring the Impact of Robotic Umpires

James Zhan, University of Rochester '20
Luke Gerstner, University of Rochester '20
John Polimeni, University of Rochester '20



Figure 1: a) Pitch location and calls from the catcher's perspective. b) Manny Machado believes he has earned a walk, but the pitch was called a strike.

Abstract

On July 10th, 2019 the Atlantic League had the first game ever called by a robotic umpire. While commissioner Rob Manfred said that there was no timeline for when the technology will be used in the majors, many people think the change will happen eventually. The purpose of our paper is to measure the impact that the robotic umpires will have on the outcomes of games and individual at bats. This question is important to the industry because it is necessary to see the possible impact that this new way of umpiring will have on the game. This is useful for teams and the rules committee. The result of this paper helps the rules committee see how this possible change may impact the game. It also is valuable to teams because if robotic umpires' impact one aspect of the game more than the other then the teams may be able to gain an advantage when constructing their rosters over other teams. Initially, we identified all the pitches that were miscalled based on a standard strike zone. Then using "Run Expectancy based on the 24 base-out states" (RE24) we analyzed the impact of miscalled pitches over the duration of a game. RE24 is calculated based on the run expectancy for the end state (RE_{es}), beginning state (RE_{bs}) and the number of runs scored (RS), the equation for RE24 is shown below. Table 1 below also shows the average RE12 for each possible count, we call it RE12 since there are twelve possible counts. We find that there are only three counts where the average RE12 is negative, indicating a count where the pitcher has the advantage, and nine counts that have a positive RE12, indicating the hitter has the advantage.

$$RE24 = RE_{es} - RE_{bs} + RS \quad (\text{Equation 1})$$

Table 1: Average RE24 for each possible count.

Count	0-2	1-2	2-2	0-1	1-1	3-2	0-0	2-1	1-0	2-0	3-1	3-0
RE12	-0.114	-0.102	-0.080	0.030	0.045	0.050	0.053	0.057	0.059	0.078	0.193	0.248

Our data came from Kaggle and it includes information of every pitch from 2015-2018 (N = 2,867,154). Our approach is novel because it analyzes the impact on the game that the robotic umpires. Most other papers on the topic of robotic umpires focus on the number of calls missed but fail to go into how much those calls actually impact the outcome of games. The results of our study can have an impact on how teams build rosters, and how they evaluate and value pitchers. The results of our study show that for every missed call the run expectancy increase by .065 runs. This means that if fifteen pitches are missed over the course of a game for one team, they would expect to score one more run. If the change in umpiring was enforced, it would negatively impact pitchers, especially pitchers that relied more on missed calls. However, pitchers that don't rely on missed calls will be much more valuable and could be the key to success for teams building for the future. Overall, we would also expect to see an increase in offensive production if robotic umpires were used.

1. Introduction

With technology rapidly growing baseball is evolving with it. There are many new technologies that are being used to help players train at the professional level and even at the college level [1,2]. For some, the technology craze is even causing distress and paranoia because of the worry of sign stealing [3]. One of the newest technologies that is being brought into baseball is with the discussion of having robotic umpires. This was already tested out in the Atlantic League at the end of the 2019 baseball season [4]. The driving force of this decision is the repeated concern that umpires, specifically home plate umpires, are impacting the game with their inaccurate calls [5]. There is also no shortage of times that umpires are missing calls which raises further concern for this problem [6, 7]. This is further demonstrated in Figure 1a, which shows the pitches and calls for each pitch in one of Manny Machado's at bats in 2015, which had the most missed calls in a single at bat over the 2015 to 2018 MLB seasons. Machado believed he earned a walk on the sixth pitch of the at bat, but as shown in Figure 1b, the umpire called it a strike as Machado was walking up the first base line [8]. This is just one example of many that existed of umpires missing a call that impacted the at bat. There has been a lot of previous work that has studied how often umpires are missing pitches such as in the Machado at bat, and this has been an argument for why there should be robotic umpires calling balls and strikes. While many people are suggesting using robotic umpires there is little knowledge on how much this would impact the game.

For our analysis we wanted to use a measure to determine the impact of umpires missing calls repeatedly throughout a game. To do this we used the run expectancy based on the 24 base-out states (RE24) metric [9]. Table 2 below shows the run expectancy matrix that is used to determine the final RE24 for an individual at bat. Using this table, we are able to use Equation 1 that was shown in the abstract. It was important to use a more overall encompassing metric such as RE24 because it allows us to analyze each at bat at a more granular level.

Table 2: Run expectancy based on the 24 different possible outcomes.

Runners	0 Outs	1 Out	2 Outs
Empty	0.461	0.243	0.095
1 - -	0.831	0.489	0.214
-2-	1.068	0.644	0.305
12-	1.373	0.908	0.343
--3	1.426	0.865	0.413
-23	1.920	1.352	0.570
123	2.282	1.520	0.736

2. Exploratory Data Analysis

In this section we first discuss how we collected the data and give a brief description of the data set that we worked with in this analysis. Then we provide visualizations of the data to further explain the data set.

2.1 Data Collection & Description

The data that was used for analysis was collected from the MLB Pitch Data 2015-2018 that is available on Kaggle [10]. This dataset contains every pitch from the 2015 to 2018 MLB seasons, and it contains many features for each one of the pitches. This dataset had all the information that was necessary to do the analysis that we present.

Overall, the pitches data set is very large consisting of 2,867,154 rows and 40 features. Most of these features describe the pitch and the others give information that relate it to a specific game. Many useful features are provided in the data set, such as the pitch's location, spin rate, velocity, break angle, how much it broke, and a nasty factor. The nasty factor is a way of describing how well hitters have done against previous pitches that are similar to the one thrown on a scale of 0 to 100 [11]. The most important features for our analysis were the ones that described the pitch location because we used this to ultimately determine if a pitch was accurately called a ball or strike.

2.2 Data Visualization

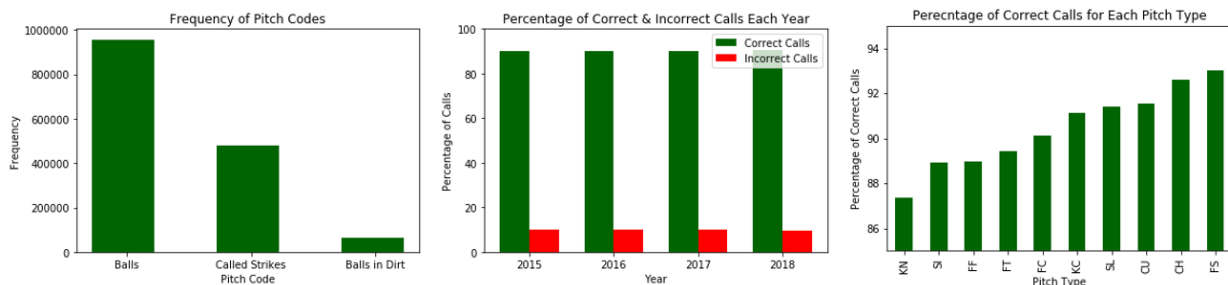


Figure 2: a) The right graph is distribution of pitch codes. b) In the middle the percentage of correct and missed calls for each year is shown. B) The right graph depicts the percentage of correct calls for different pitch types.

Figure 2a exhibits the frequency of each of the pitch codes that are used in the analysis. To get the best representation of how often umpires are missing calls we considered only the pitches that are called directly by the umpire, this excludes swinging strikes, intentional balls, foul balls and other pitches whose outcome is not determined by the umpire. Figure 2b shows the percentage of calls that were correct and incorrect for each of the years that the data was collected. It shows that over the course of each of the year’s umpires were correctly calling pitches around 90 percent of the time, and incorrectly calling pitches around 10 percent of the time. This shows that over the year’s umpires have been relatively consistent in how they are calling pitches. Figure 2c shows the percentage of correct calls for knuckleballs (KN), sinkers (SI), four-seam fastballs (FF), two-seam fastballs (FT), cutters (FC), knucklecurve (KC), sliders (SL), curveballs (CU), changeups (CH), and splitters (FS). As seen in Figure 2c, the most common pitch type that was missed was knuckleball, with a correct call rate of only 87.4 percent. This agrees with intuition because a knuckleball moves a lot before crossing the plate and can be tough for an umpire to determine if it was actually a ball or strike. Sinkers (SI), four-seam fastballs, two-seam fastballs and cutters were the next four most commonly missed pitches. This may be surprising at first, but it is common to avoid throwing fastballs, and other hard pitches right down the middle of the plate and try to throw them more towards the corners of the plate. This could be one reason that more fastball type pitches are being miscalled. Also shown is that splitters are called correctly the most out of any pitch at an impressive 93.0 percent. This could be caused by pitchers often throwing this pitch outside the strike zone and trying to get hitters to chase it.

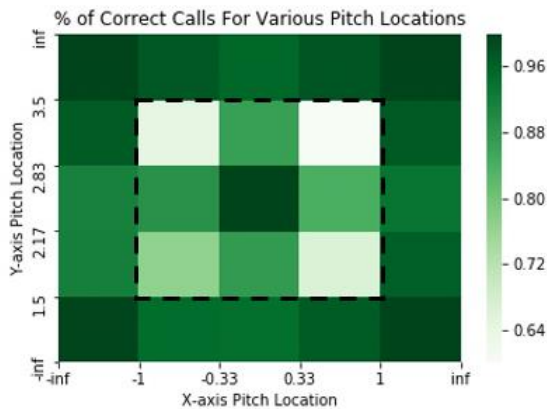


Figure 3: The percentage of correct calls inside the strike (the dashed area) and outside the strike zone.

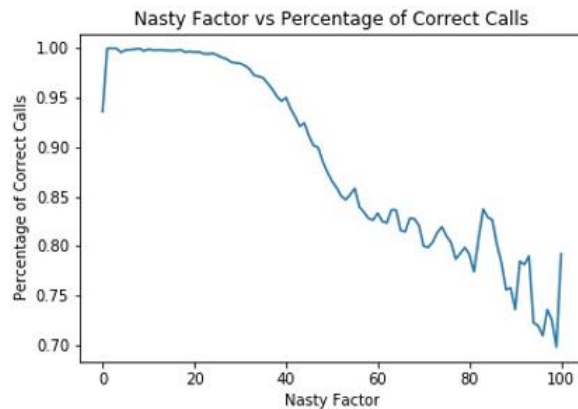


Figure 4: The percentage of correct calls as the “Nasty Factor” increases.

We also wanted to discover how the percentage of correct calls differs based on where the pitch is located. Figure 3 shows a heatmap representing the percent of calls that were correct in 25 different locations. As expected, pitches that are in the middle most region are called correctly the most often. The pitches that were called incorrectly were the pitches that were in corners of the strike zone, most specifically the top two corners. This finding is consistent with the idea that the hardest pitches to call correctly are those that are right near the border of the strike zone.

As baseball continues to evolve and change people are creating more and more unique measurements to find patterns in the data. One new metric is the nasty factor, and it is interesting to see how the percent of missed calls changes as the nasty factor increases. Figure 4 shows that as the nasty factor rating increases the percentage of correct calls decreases. Since the nasty factor considers other features about how a pitch moves it is reasonable that the nastier a pitch is the more often it is incorrectly called.

3. Methods

In this section we first present our method for how we labeled if a pitch was called correctly by the home plate umpire. We then describe our reasoning for using the RE24 metric throughout our analysis.

3.1 Labeling Correct Calls

One of the most important parts for this data set was to label the pitch if it was actually a ball or strike. The original data contains information about if the pitch was originally called a ball or strike. However, these labels aren't always the correct call. As with many human tasks there can be error in an umpire calling a pitch a ball or strike. Our method of labeling whether the pitch was actually a ball or strike relied on a consistent strike zone. We used the width of home plate, 17 inches, plus a half inch on each side to account for the width of the baseball, and we used a constant height of the pitch. If the pitch was in this region then we labeled it as a strike, regardless of what the original call was, and if the pitch was outside this region by any distance the pitch was labeled a ball. Overall, we found that 10 percent of pitches were called incorrectly from umpires.

3.2 Decision to use RE24

To best analyze the impact of miscalled pitches based on the standard strike zone, we quantified this impact through a metric based off of RE24. RE24 is a measurement of how well hitters are impacting the game at an at-bat level. This metric is context dependent meaning it factors in the situation of each at bat, for example the number of runners on base as well as the number of outs at the time of the at bat. RE24 is calculated based on the run expectancy for the end state (RE_{es}), beginning state (RE_{bs}) and the number of runs scored (RS), the equation for RE24 is shown below.

$$RE24 = RE_{es} - RE_{bs} + RS$$

4. Results

In this section we present the results of our findings using the run expectancy metric RE24 that has been described previously. We then present the results for individual pitchers in which we suggest different groupings of pitchers and how they may be impacted by robotic umpires.

4.1 RE24

For our analysis, we were focused on how a missed call would impact the game, which required us to break RE24 down to a pitch-by-pitch level, we'll call this RE12, for the 12 unique valid pitch counts. To calculate RE12, our first step was to isolate all at bats that ended with a

particular pitch count, and then take the average RE24 for each pitch count. For example, to calculate the RE12 for an 0-0 pitch count, we looked at all at bats that ended on a 0-0 pitch count, and then calculate the RE24 for each of those at bats, and then average those RE24 values to get our final RE12 result. Through this methodology, we were able to create a metric that measured how well hitters were impacting the game at a pitch-by-pitch level. We ran this methodology on MLB pitch data from the 2015-2018 seasons. Table 3 below shows the average RE24 for each possible count.

Table 3: Average RE24 for each possible count.

Count	0-2	1-2	2-2	0-1	1-1	3-2	0-0	2-1	1-0	2-0	3-1	3-0
RE12	-0.114	-0.102	-0.080	0.030	0.045	0.050	0.053	0.057	0.059	0.078	0.193	0.248

Using these values, we were able to quantify the impact of a missed call. For example, if an umpire misclassified the first pitch of an at bat as a ball, then the hitter would be expected .023 more runs for the outcome of the at bat. Aggregating this for all missed calls and for all at bats from the 2015-2018 MLB seasons, we were able to conclude that missed calls on average, the run expectancy increased by .065 runs per game. This means that if fifteen pitches are missed over the course of a game for one team, they would expect to score one additional run.

One important thing about the calculation for average run expectancy per missed call is that the distribution of missed calls is unbalanced. From the dataset, there is about 3 times more pitches that were incorrectly called as balls than pitches that were incorrectly called as strikes. Because of this, we see a positive average run expectancy for each missed call. Isolating the average run expectancy for miscalled balls and miscalled strikes, we see that the effects are very similar. For each miscalled ball, the average run expectancy increases by about 0.12 runs while for each miscalled strike, the average run expectancy decreases by about 0.10 runs. If the number of miscalled pitches were more evenly distributed, then the effect of miscalled pitches wouldn't be as significant. However, because there is a lot of miscalled pitches in favor of the hitter, then miscalled pitches in the long run would increase runs.

4.2 Individual Pitchers

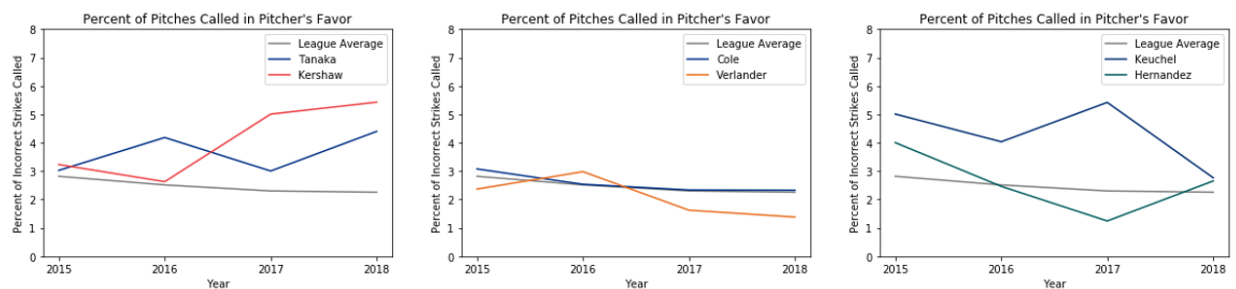


Figure 5: a) The left graph describes pitchers who have above league average calls in their favor. b) The middle graph shows pitchers who are at or slightly below league average for pitches called in their favor. c) The graph on the left shows pitchers who have had their best seasons when they get more calls in their favor.

Figure 5a shows the percent of pitches that were incorrectly called strikes for different pitchers has changed over the past four seasons. It also compares the pitchers to the overall league average. The results shown are for some of baseball top pitchers that have received a lot of

benefits from the fact that umpires miss calls. For example, Clayton Kershaw, a starting pitcher for the Los Angeles Dodgers, had an incredible 5.4 percent of his pitches in 2018 called as strikes when they were actually balls. This compares to a league average of just 2.3 percent in the same year. Kershaw was getting nearly twice as many of his pitches called as strikes when they were balls compared to the entire league average. Also shown are the results of Masahiro Tanaka, a starting pitcher for the New York Yankees, during the 2015 to 2018 seasons. He also saw well above the league average of calls go in their favor over the past few seasons.

Ultimately it is hard to know exactly why these pitchers are getting more calls in their favor. It could potentially be from the reputation they have earned after having multiple good seasons. Or

possibly it could be because their catchers are very good at framing the pitches that they are throwing. However, if robotic umpires were used these pitchers would not be rewarded with as many calls in their favor and this could have an impact on their ability to perform as well as they currently are.

In Figure 5b the percent of pitches that were incorrectly called as strikes is shown for Gerrit Cole, a pitcher that just signed a record-breaking deal with the New York Yankees. As suggested by his deal that was reported for 9 years \$324 million, he is one of the top starting pitchers in the entire league [12]. However, as seen in Figure 5b Cole has not fared as well in getting a lot of pitches called in his favor. In fact, for most years he is extremely close to league average and is even below league average at times. Also shown is Cole's former teammate, Justin Verlander of the Houston Astros. Verlander has also been right around league average or even slightly below in 2017 and 2018, but he remains to be one of the most dominant and reliable pitchers. This suggests that not all of the top pitchers rely on an advantage in getting calls to go in their favor.

Figure 5c shows specifically how Dallas Keuchel's percent of pitches incorrectly called strikes has changed over the past few years. In 2015, Keuchel had one his highest value of pitches that were called in his favor He saw this percentage decrease in both 2016 and then increase again in 2017 and it decreased again in 2018. It is also interesting to note that Keuchel was an all-star in only 2015 and 2017 and not in either 2016 or 2018, Keuchel was also the Cy Young Award winner in 2015 and finished fifth in MVP voting [13]. A similar phenomenon is found in Felix Hernandez's recent years. As shown in Figure 5c, Hernandez had the most calls go in his favor in 2015, and then this metric decreased in 2016 and 2017 while finally returning to league average in 2018. The last season Hernandez was in the top of the league in Cy Young voting was 2015 and this was also his most recent year as an all-star [13]. While there were more than likely other factors that contributed to whether or not Keuchel and Hernandez made the all-star team, the amount of calls they got in their favor could have helped them have better seasons statistically.

5. Discussion

In this section we first discuss the importance of our findings and how it relates to pitcher's value. We then examine how the change to robotic umpires could impact the value of catchers, and finally we explore the overall impact this could have on the game.

5.1 Importance of Findings

Based on our findings each missed call helps the batting team by .065 runs per game . The main reason that this work is important because changing from traditional umpires to robotic umpires fundamentally changes the game. By normalizing the strike zone, different player values are altered. The biggest reason why this is important is because teams that plan ahead and recognize this change can have a huge competitive advantage. We have seen this for years in baseball, innovation leads to success. The A's were one of the first to gain an edge analytically, the Royals gained an edge through their early use of bullpens and the Astros have gained an advantage through player development. The way to success in baseball is to find an advantage against your competitors and exploit it. Through our research, teams may be able to exploit market inefficiencies in measuring player values. Examples of how teams can find value if this change was made can be seen in our results section when we displayed how certain pitchers benefit more than others from missed calls. Pitchers such as Tanaka and Keuchel benefit immensely from missed calls. To best demonstrate how teams can find value if the league moves to robotic umpires is with a hypothetical scenario. If Keuchel was a free agent after the 2017 season he would have been up for a big payday. Any team would love to have an all-star pitcher with a 2.90 regular season ERA and more than 170 IP in the regular and postseason [13]. However, according to our model he was above league average on incorrectly called strikes. If umpiring changed from traditional to robotic, we expect that he would be one of the most impacted pitchers in the game. He has almost twice as many missed strike calls as the average major league pitcher. If the MLB changed to a system that correctly calls those pitches, then he would need to drastically change his game. Teams that recognized this impact in advance would have either stayed away from him as a free agent or valued him much below the market did. The reverse of this example can also be seen in a pitcher such as Felix Hernandez after the 2017 season. Felix was a percent below the league average on missed calls in his favor. So, teams that recognized this mark would place a higher value on him than other teams because he would not be as affected by the change in umpiring as the average MLB pitcher. The same analysis can be done for hitters who receive an unproportionate number of incorrectly called balls that should have been strikes, and how their value would change if robotic umpires were implemented.

5.2 Impact on Catcher's Value

At first glance it seems that changing the strike zone just impacts the batters and pitchers, but the catcher's role in the game is also impacted. In an article measuring value of catchers, posted by The Ringer, the difference between the top pitch framers and worst pitch framers is about 40 runs per season [14]. With the implementation of robotic umpires, the value of good pitch framing catchers would be erased. Pitch framing is becoming more and more of an important stat to front offices, one unnamed GM was quoted in The Ringer article saying "The really bad framers who haven't improved—even with the aforementioned coaching—no longer get major league jobs" [14]. Robotic umpires would completely eliminate this skill and partially diminishing the role that a skilled framing catcher would have. While catchers obviously would still have defensive value in blocking, pitch calling, pitcher management, etc. their value would be expected to diminish with these changes to umpiring.

5.3 Impact on the Game

The robotic umpires that were implemented in the Atlantic League were only in place for a little less than a month, so it is difficult to see how they impacted the game with such a small sample size. However, current players and managers in the Atlantic League are open to the new changes. The manager for the Somerset Patriots, a team in the Atlantic League, Sparky Lyle said, "I didn't know how it was actually going to figure it out, but it's actually been pretty good" [4]. While the jury may still be out on how this change has impacted the Atlantic League, the sentiment towards the change is positive.

6. Conclusion

Our results indicate that one missed call translates to .065 expected runs for the offensive team. While we are confident in our results, we cannot say that this would be the definite impact that removing umpires would have on the game. Players will adapt, we have seen changes in the strike zone before. For example, in 2017 the MLB raised the bottom of strike zone. Players and pitchers both adapted to this change and there was no observed relationship between raising the strike zone and strikeout rates [16]. The change to robotic umpires is much more of a drastic change than the one in 2017. However, it is possible that players adapt to the new playing conditions in a similar way as they have before, and the impact is less drastic than we predicted. Greater insights into how robotic umpires impact the game can be seen if the Atlantic League continues to use them in their upcoming season.

7. Future Work

To continue our work the next step would be to use a more dynamic strike zone for each batter. To determine what a pitch should have been actually called we used a constant strike zone that others have used in their work as well [15]. However, a better approach could be to use a dynamic strike zone that adapts to each hitter's height. This would result in a more accurate determination of if the pitch was correctly called by the umpire. Overall, since we did have an extremely large sample size it is possible that we could achieve even stronger results if we use this approach.

While we were able to provide in depth analysis of our findings and results there is still much more work to be done in determining the impact robotic umpires could have on baseball. We presented an analysis that looked at how the change to robotic umpires could impact the overall game using a run expectancy measure, and we showed how it could impact pitchers. However, in the future it would be beneficial to also consider how individual hitters would be impacted by the change to robotic umpires. For example, hitters that are known to have a good batter's eye could be analyzed to see if they are actually very good at determining what a ball or strike is, or if some of the calls they are getting are because of who they are.

References

- [1] Borzi, Pat. “‘If You Don’t Have It, You’re Behind’: College Baseball’s Tech Arms Race.” *The New York Times*, 25 June 2019, www.nytimes.com/2019/06/25/sports/baseball/college-world-series.html.
- [2] Woodruff, Jay. “5 Technologies That Are Revolutionizing Baseball.” *Fast Company*, Fast Company, 8 Aug. 2019, www.fastcompany.com/90378232/5-technologies-that-are-changing-baseball.
- [3] Trister, Noah. “New Technology Only Adds to Baseball's Culture of Paranoia.” *USA Today*, Gannett Satellite Information Network, 19 Oct. 2018, www.usatoday.com/story/sports/mlb/2018/10/19/new-technology-only-adds-to-baseballs-culture-of-paranoia/38209991/.
- [4] Acquavella, Katherine. “Robot Umpires: How It Works and Its Effect on Players and Managers in the Atlantic League, plus What's to Come.” *CBSSports.com*, CBSSports, 30 Aug. 2019, www.cbssports.com/mlb/news/robot-umpires-how-it-works-and-its-effect-on-players-and-managers-in-the-atlantic-league-plus-whats-to-come/.
- [5] Williams, Mark T. “MLB Umpires Missed 34,294 Pitch Calls in 2018. Time for Robo-Umps?” *BU Today*, Boston University, 8 Apr. 2019, www.bu.edu/articles/2019/mlb-umpires-strike-zone-accuracy/.
- [6] Green, Etan, and David P. Daniels. "What does it take to call a strike? three biases in umpire decision making." *2014 MIT Sloan Sports Analytics Conference*. 2014.
- [7] Williams, Mark T. "MLB MUST EMBRACE TECHNOLOGY TO FIX POOR UMPIRE PERFORMANCE." *Phi Kappa Phi Forum*. Vol. 99. No. 3. Honor Society of Phi Kappa Phi, 2019.
- [8] MLB. “BAL AT TOR - June 19, 2015” YouTube, 31 Oct. 2015, <https://www.youtube.com/watch?v=K7RGZUVniE4&list=WL&index=5&t=3807s>.
- [9] Weinberg, Neil. “RE24.” RE24 | Sabermetrics Library, 30 July 2014, library.fangraphs.com/misc/re24/.
- [10] Schale, Paul. (2018, March). *MLB Pitch Data 2015-2018, Version 19*. Retrieved September 15, 2019 from <https://www.kaggle.com/pschale/mlb-pitch-data-20152018>
- [11] “MLB.com At Bat.” *Major League Baseball*, 2010, www.mlb.com/mlb/gameday/y2010/.
- [12] Hoch, Bryan. “Yankees Gerrit Cole Deal.” *MLB.com*, 2019, www.mlb.com/breaking-news/yankees-gerrit-cole-offer.
- [13] Sports Reference LLC. (2018). *Baseball-Reference.com - Major League Statistics and Information*. Retrieved from <https://www.baseball-reference.com/>.
- [14] Lindbergh, Ben. “Pitch-Framing Is More Important-and More Common-Than Ever.” *The Ringer*, The Ringer, 21 Sept. 2018, www.theringer.com/mlb/2018/9/21/17885820/pitch-framing-strike-zone-jorge-alfaro-tyler-flowers.
- [15] Dan Brooks. (2019). *BrooksBaseball.net: PITCHf/x Tool | Strikezone Maps*. Retrieved from <http://www.brooksbaseball.net/pfxVB/pfx.php>
- [16] Brodie, Adam. "The Influence of the Sinking Strike Zone on Major League Baseball’s Strikeout Epidemic."