

Examining MLB's Strike Zone

An Interactive Qualifying Project Report

Submitted to The Faculty

of

Worcester Polytechnic Institute

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Joseph Flanagan

January 2015

Approved:

Professor Sarah Olson

Abstract

The goal of this project is to examine the public opinion of using technology to automate the MLB strike zone and its possible effects on MLB's run-scoring environment. To examine public opinion, a survey will be administered and an interview with someone inside baseball scheduled for a more nuanced look. To examine the possible effects, data from the 2007-2014 seasons will be analyzed using Markov chains in an attempt to predict the differences between letting umpires call balls and strikes and leaving it to technology. Initial results suggest that automation would increase offense, but that fans are not quite ready to trust the technology.

Contents

1	Introduction	4
1.1	Baseball	4
1.1.1	Important Terms and Acronyms	4
1.2	Automating the Strike Zone	5
1.3	PITCHf/x	6
1.4	Technology in other sports	7
2	The Data	9
2.1	Plate Discipline Data	9
2.2	Trends by Count	9
2.3	Umpire Accuracy	10
3	Methodologies	12
3.1	Markov Chains	12
3.2	Building the Transition Matrix	13
3.3	Simulating a Hypothetical 2014	15
4	Results	16
5	Survey	18
6	Discussion	25
7	Appendix	27
7.1	MATLAB code	27
7.2	Interview transcript	29

List of Figures

2.1	Plate Discipline statistics for 2007-2014	9
2.2	Relative Zone, Z-swing, and O-swing percentages by count for 2014	10
3.1	A generalized transition matrix	13
3.2	The calculated state space for 2014	14
3.3	The state space for a hypothetical 2014 with an automated zone	15
4.1	Markov-predicted strikeout and walk rates	16
5.1	Those who indicated watching some amount of baseball opposed a fully automated zone	19
5.2	Just as strongly, however, they supported a semi-automated zone	20
5.3	Survey respondents think umpires today average 10-15 missed calls per game	21
5.4	Generally, though, they believe that 5-10 missed calls is an acceptable number	21
5.5	Those who demanded very few missed calls were the only group to support automating the strike zone	22
5.6	They were also the only group to support the change if it meant more offense	23
5.7	Those who are willing to accept more bad calls were strongly opposed to a fully auto- mated zone	23
5.8	In general, their responses were more in line with the general survey results	24

List of Tables

2.1 Umpire accuracy by count	11
--	----

Chapter 1

Introduction

1.1 Baseball

The strike zone is at the heart of Major League Baseball. Any hitter who wants to stick around must learn to be disciplined about which pitches are in the strike zone and which are out, and any pitcher who wants to last must learn to paint its edges. Any pitch that passes through the strike zone without the batter making contact is a strike, while any pitch that does not cross the zone and is not swung at is a ball.

“The count” in baseball refers to counting balls and strikes. After 4 balls, a batter draws a walk, and after 3 strikes he has struck out. The count is represented as the current number of balls, followed by a hyphen and the current number of strikes. For example, if there are 2 balls and 1 strike on a batter the count is said to be 2-1. 1 Strike and no balls is an 0-1 count, etc. When the count reaches 3-2, it is said to be “full” because it can get no bigger. Another ball would lead to a walk and a new batter with a fresh count, and another strike would lead to a strikeout and the same thing. Every plate appearance starts with an 0-0 count, and can later progress to 0-1, 0-2, 1-0, 1-1, 1-2, 2-0, 2-1, 2-2, 3-0, 3-1, and 3-2. In all there are 12 possible states for the count to be in.

These calls are made by the home plate umpire, who must decide if a pitch was in the strike zone or not. His job is made more difficult by having to crouch and view the plate over the catcher’s shoulder, often leading to an angle that is off-center, and by the fact that major league pitches generally travel at least 75 MPH and can reach up to 100 MPH! It should be obvious, then, that umpires do not get every call right. Humans can only get so good at tracking tiny objects at incredible speeds; an umpire might last actually “see” the baseball when it is still a good 5 or 10 feet from the catcher [1].

1.1.1 Important Terms and Acronyms

- PA = plate appearance. Every time a batter completes his time at bat (barring the inning ending prematurely on a baserunning out), it is recorded as a plate appearance. All of a player’s counting stats (at bats, sacrifices, walks, etc.) are included in his plate appearance total.
- TBF = total batters faced. This is the total number of plate appearances a pitcher participates in. Total batters faced and plate appearances across MLB are equal every year.
- BIP/IP = “Ball in play/In Play”. Any pitch that is contacted by the batter and not a foul ball is considered “in play.” Sometimes, this term also excludes home runs, but for the purposes of this project it will encompass all fair contact.

- K = strike out. When expressed as K%, it represents the percentage of opposing hitters that a pitcher strikes out, or K/TBF. More strikeouts are better for a pitcher.
- BB = walk. When expressed as BB%, it represents the percentage of opposing batters a pitcher walks, or BB/TBF. Fewer walks are better for a pitcher.
- HR = home run. This is the worst possible outcome of any given plate appearance for the pitcher.
- HBP = hit by pitch. A hit batsman gets first base for free. They are essentially walks, except that runners may not advance after a HBP.
- ERA = earned run average. This is the number of runs, on average, that a pitcher allows per nine innings pitched, not counting runs that scored due to a fielder's error.

The following statistics are known as “plate discipline” stats.

- Zone. This is the chance that a pitcher throws the ball within the defined strike zone.
- Ooz = Out of zone percentage. This is the chance that a pitcher throws the ball outside of the defined strike zone.
- Z-swing. This is the chance that the batter swings, given that the ball was in the strike zone.
- O-swing. This is the chance that the batter swings, given that the ball was not in the strike zone.
- Z-contact. This is the chance that the batter makes contact, given that he swings at a ball in the zone.
- O-contact. This is the chance that the batter makes contact, given that he swings at a ball not in the zone.

1.2 Automating the Strike Zone

Partly because of this and partly because of the unending forward march of technology, it has been proposed by many that the strike zone should be automated, no longer subject to the discretion of fallible humans. Despite all the conversation over how to automate the zone and why it must or must not be done, there is surprisingly little discussion about the possible consequences of such automation. With such a drastic change in how balls and strikes are called, there almost certainly will be a tangible effect on run-scoring. Will batters or pitchers be helped more by the new, ultra consistent zone? What will the magnitude of the change be? These things must be considered before implementing any new technology, as MLB is already experiencing a rather sharp decrease in scoring, and a change that helps pitchers too much could possibly drive away fans.

But how to predict these possible changes? We can't just look back at plate appearances with bad calls and imagine how they might have ended otherwise; once a pitch is called (correctly or incorrectly), it changes the count that the rest of the plate appearance will be predicated on. One of the best examples is a 1-1 count. Imagine that the pitcher throws a fastball just off the edge of the

plate. If the pitch is called according to the rule book strike zone, it is a ball and the count will be 2-1, an advantage to the batter. But if the umpire thinks it was right on the edge and incorrectly calls a strike, the count becomes 1-2, and the batter is on the defensive. Looking at the rest of that plate appearance after the wrong call is not informative, because both the batter and pitcher are likely to change their behavior between 2-1 and 1-2 counts. At 1-2, the batter may now strike out reaching for the next pitch, which he may have comfortably let go if the count were correctly 2-1. Be they batter or pitcher, no MLB player's actions are independent of the count, and it is for this reason that any attempts to reconstruct "what could/should have been" are futile.

This is why I have opted to use Markov chains to model the batter/pitcher interaction. For each pitcher, we know how often he throws the ball in the zone, we know how likely batters are to swing at his offerings, their chance of making contact or whiffing, etc. With Markov chains, we can use these to estimate how many strikeouts, walks, and balls in play will result. Then we can attempt to estimate how these numbers would change if the strike zone were automated. We can assume that any pitch out of the zone that is incorrectly called a strike would now be a ball, and that any pitch taken in the zone for a ball is now a strike. Despite the fact that players may sometimes chase pitches out of the zone for fear of having them called borderline strikes, it is impossible to tell which swings fall in this group and I believe most swings at balls occur because the batter is simply fooled.

Once we have figured pitcher's actual expected rates of strikeouts, walks, and balls in play, and their expected rates with the new strike zone automation, we can compare the two groups and attempt to predict the change in MLB's offensive environment. More strikeouts and fewer walks would almost certainly mean a more pitcher-friendly environment, while more walks and fewer strikeouts would be expected to help batters. If both walks and strikeouts move in the same direction, it will be more difficult to assess any possible future changes in run scoring. Rates of balls in play do not need to be examined, as they are simply the function of walk and strikeout rates (specifically, the rate of balls in play is 1 minus the rate of (walks plus strikeouts)).

Baseball lends itself to Markov analysis much more so than the other major sports, and I am not the first to recognize this fact [2]. Markov chains will be explained in detail in Methodologies.

1.3 PITCHf/x

If the strike zone were to be automated in the near future, the most likely technological candidate is Sportvision's PITCHf/x [3]. First used experimentally in the 2006 playoffs, PITCHf/x cameras are now installed in every major league stadium and provide the data for MLB.com's GameDay application, which displays each pitch visually in real time. By capturing the game from the center field camera, PITCHf/x can record the speed, break, and location of every pitch, accurate down to one mile an hour and one inch.

There's plenty of hard work behind the scenes to enable the technology to be so accurate in the first place. Before even the groundskeepers get out there to do their job, a special crew sets up to ensure that PITCHf/x works properly during the night's game. Called the "registration" process, they place colored or numbered markers on the first and third base lines, with another eight foot pole demarcating home plate. This is captured by cameras high above home plate and first base, and one way out in center field, in order to create the grid that allows PITCHf/x its tracking capabilities.

The center-field camera is also used for the extremely important job of "sizing" the batter. For the software to accurately call the pitch, there needs to be a different strike zone for Jose Abreu than for Jose Altuve, who is nearly a foot shorter than Abreu. While the players take batting practice before the game, the PITCHf/x crew sizes each of them, marking where the top and bottom of the zone should

be for their natural batting stances. This particular zone is then remembered by PITCHf/x for each subsequent plate appearance by that player.

According to Kurt Meyer [4], a broadcast engineer for SportVision, “a guy stands at home plate with the eight-foot pole and marker, and then the software takes about 20 minutes to snap the grid into place. That tells each of the computers where home plate is in relationship to the three cameras, so they’re all on the same page. You’re telling the computer to look for a certain object between parameters of speed ... a blob traveling between the mound and the plate.”

The pitch-tracking system sets those parameters of speed from 40 mph to 120 mph. That’s certainly a wide enough range to capture every pitch, but it also prevents the system from picking up a trash bag picked up by a gust of wind or a beach ball that’s gotten loose from the stands. Occasionally, the system will accidentally pick up a third baseman charging towards the plate as the batter squares to bunt. In this case, the crew must intervene to drag the grid on the software back to its intended dimensions.

In addition, due to miscalibrations or just plain computer glitches, sometimes the system simply misses a pitch. For this reason, it is likely that there will always remain an umpire behind home plate as a technological failsafe of sorts.

1.4 Technology in other sports

When a free kick is awarded in soccer, the defending team is supposed to remain 10 yards from the ball’s location until it is kicked, and the kicking team is not supposed to move the ball from the initial location appointed by the ref. Scrounging for any miniscule advantage, it is common for defenders to encroach closer than that, and for offensive players to try to nudge the ball toward the net. This is simple human nature, analagous to children craning their necks to copy a test problem, knowing that the teacher cannot keep her eye on all of them at once. In the same way, it is difficult for a referee to control all of the players on the field and prevent them from, technically, cheating.

The similarity in baseball is the fuzziness at the edges of the strike zone. Pitchers and catchers want to slowly expand the umpire’s strike zone by throwing a series of pitches successively further off the plate. They must hit their spots precisely and frame the ball in such a way that it looks good to the umpire to be successful, though. In addition, much as a referee cannot actually see the spot on the ground where the ball should be or the imaginary line 10 yards away, an umpire cannot see the strike zone, nor can he actually even see the ball the entire time. With these inherently human difficulties come inevitable mistakes. These mistakes may be only tiny, but in a game of inches like soccer or baseball, specifically given the fact that the strike zone is measured in inches to begin with, they can make a tangible difference.

To solve this problem in soccer, vanishing spray [5] has now been introduced at the highest levels of competitive play, making its first World Cup appearance in 2014. Vanishing spray, or foam, is applied from an aerosol can to an athletic field to provide a visual marker. Invented in 2000 by Heine Allemagne, it was first used in professional competition in the 2001 Brazilian Championship and soon adopted in Brazilian play. With this invention, the referee can mark the ball’s location and then spray a line 10 yards away that all the defenders must stay behind during the free kick. After about a minute, the spray disappears, playing no part in the rest of the match until it may be needed again. It has been considered a success so far, having no real drawbacks.

Baseball has given us another recent example of integrating technology into the sport with the use of instant replay. Introduced for the 2014 season, the replay system allows managers to challenge a call that they believe to be incorrect. To prevent abuse, a manager who loses a challenge (does not get the

call overturned) may not challenge again in that game. Starting in the 7th inning however, managers can request the umpires to review a close play regardless if they have already used a challenge. Instant replay can be characterized as a success so far. Data is not yet readily available on the number of calls upheld and overturned, but instant replay allowed some number of incorrect calls to be corrected and did not intrude too much on the game. In this way, instant replay has been a successful model of baseball adopting technology to improve the game.

Chapter 2

The Data

2.1 Plate Discipline Data

To obtain the plate discipline data, I created a custom leaderboard at www.fangraphs.com [6] in order to pick and choose which stats I wanted to examine. I selected the years 2007-2014, because those are the years for which there exists complete plate discipline statistics. I chose Zone, O-swing, Z-swing, O-contact, and Z-contact to display and clicked the “League Stats” option to view these percentages for MLB as a whole for each of the last 8 seasons. Then I imported the data into Microsoft Excel, as seen in Figure 2.1.

Season	Zone% (p)	O-Swing%	Z-Swing%	O-Contact	Z-Contact
2007	49.60%	28.80%	62.00%	68.20%	88.60%
2008	50.40%	28.00%	62.30%	67.40%	88.90%
2009	50.50%	28.00%	61.20%	67.00%	88.80%
2010	50.60%	28.30%	61.00%	66.10%	88.80%
2011	50.30%	28.90%	62.00%	66.10%	88.80%
2012	49.50%	29.20%	61.80%	64.90%	88.30%
2013	49.30%	29.40%	62.10%	64.70%	88.40%
2014	48.90%	29.80%	62.70%	64.50%	88.40%

Figure 2.1: Plate Discipline statistics for 2007-2014

Over the last 8 seasons, Zone, O-swing, Z-swing, and Z-contact have remained essentially stable, with only minor fluctuations. O-Contact is the only stat to display any trend and it has been trending downward over the selected period.

2.2 Trends by Count

No MLB player’s actions are independent of the count. A pitcher is much more likely to throw a ball when he is up in the count 0-2 than when he is behind 3-0. In the former case, he ideally wants to get the batter out on a tough pitch and knows he can afford to throw a ball in the situation. In the latter situation, throwing another ball means a baserunner, so the pitcher has more incentive to throw one in the zone.

Count	Zswing+	Oswing+	Zone+
0-0	61	52	110
0-1	112	98	88
0-2	131	117	62
1-0	91	81	113
1-1	119	115	99
1-2	133	136	75
2-0	91	79	121
2-1	123	121	115
2-2	136	153	95
3-0	19	17	129
3-1	114	105	128
3-2	138	170	122

Figure 2.2: Relative Zone, Z-swing, and O-swing percentages by count for 2014

Much like pitchers throwing in the zone, batters adjust their swing tendencies by count. A hitter is much less likely to swing in a 3-0 count than one in which he has 2 strikes. At 3-0, he would like to force the pitcher to prove that he can throw strikes, and knows he can afford to take a strike to make the count 3-1. With 2 strikes, though, the batter must be wary of striking out and will swing more often.

These basic trends are known to every serious baseball fan from simply watching the game, but thanks to BaseballSavant [7], I was able to confirm these tendencies with data and quantify their affects. Using the PITCHf/x search, I went through the year 2014 count by count and determined Zone, Z-swing, and O-swing for each. I then divided these numbers by the overall Zone, Z-swing, and O-swing data from FanGraphs and multiplied by 100 to get relative rates, Zone+, Z-swing+, and O-swing+, for each count, such that 100 is average and 101 is 1% more than average (Figure 2.2). I did not repeat this process for the other years because it was extremely time consuming and labor intensive, and 2014 is the year I am most interested in anyway. That said, I used these same relative rates for the 2007-2013 seasons because it is certainly more accurate than making no adjustments for the count.

Looking at the first row, for example, we see 61 for Z-swing+, 52 for O-swing+, and 110 for Zone+. This means that, in 0-0 counts, batters swung 39% less often at pitches in the zone than they did in all counts. They swung 48% less frequently at pitches outside the zone than they did overall. And pitchers were 10% more likely to throw the ball in the strike zone than they were when considering all counts.

2.3 Umpire Accuracy

The last data set required is how often umpires screw up the call, according to PITCHf/x. Because PITCHf/x is the system most likely to be used if umpires were replaced and because it is an objective, unchanging strike zone, when I refer to umpires getting a call “wrong” or “right” it is in reference to the PITCHf/x data. If the umpire’s call and PITCHf/x’s determination differ, the umpire is considered to have made a mistake and missed the call.

Again, I wanted count by count data, so I returned to BaseballSavant’s [7] PITCHf/x search. For each count for the year 2014, I recorded the number of called balls (by umpires) on pitches outside

the strike zone (according to PITCHf/x), the number of called balls in the strike zone, the number of called strikes outside the zone, and the number of called strikes in the zone. I then calculated the rate of missed calls on pitches out of the zone and pitches in the zone for each count. Table 2.1 summarizes my findings.

Table 2.1: Umpire accuracy by count

Count	Pitches	Zone	Balls	Rate incorrect	Out of Zone	Strikes	Rate incorrect
0-0	183137	79507	3524	.075	102910	17345	.202
0-1	90814	31846	1304	.166	58968	3871	.097
0-2	45276	11182	333	.248	34094	844	.040
1-0	71976	31874	927	.075	39382	5946	.205
1-1	72042	28456	779	.135	43586	3348	.123
1-2	66190	19652	472	.230	46538	1351	.052
2-0	24461	11438	291	.065	12303	2153	.237
2-1	36363	16690	362	.122	19673	1775	.149
2-2	55425	20927	335	.186	34498	1299	.075
3-0	7950	3716	108	.033	3514	1033	.325
3-1	14900	7608	153	.085	7292	901	.188
3-2	32425	15711	148	.134	16714	691	.093

Once again, I did not repeat this process for the other seasons, as 2014 is the most important and it is simply not worth the effort with my software capabilities. The two most important columns in this table are the ones labeled “Rate incorrect”. The first, next to the “Balls” column, is the ratio of called balls in the strike zone to called pitches in the strike zone. If an umpire were to call 1000 0-0 pitches that are inside the PITCHf/x strike zone, we would expect him to incorrectly call 75 of them balls, for a success rate of 92.5%. Consequently, the other “Rate incorrect” column on the far right is the chance of making a bad call on a pitch out of the zone. If an umpire were to call 1000 0-0 pitches that are outside the PITCHf/x strike zone, we would expect 202 incorrect strike calls, a success rate of less than 80%.

Chapter 3

Methodologies

3.1 Markov Chains

A Markov chain [8] is a sequence of random variables with the Markov property, that is, given the current state, the past and future states are independent. It is a random, memoryless mathematical system that transitions from one state to another on some state space. Changes between the system's states are called transitions, and the probabilities associated with various state changes called transition probabilities. A transition matrix containing all these probabilities becomes the state space. Since we are using Markov chains to simulate plate appearances, our states are the states of the count in MLB.

The states that the count can be in at any time are 0-0, 0-1, 0-2, 1-0, 1-1, 1-2, 2-0, 2-1, 2-2, 3-0, 3-1, and 3-2. For our purposes, we add the following states: strikeout, walk, and ball in play. In reality, all these events signify the end of a plate appearance and the start of a new count, but we will treat them as states of the count nonetheless to act as placeholders. This gives us a total of 15 different states that the count can be in, and causes the transition matrix in Figure 3.1 to be 15x15 for each season.

With 225 total entries, the transition matrix for each season will be fairly large, but it is made much more manageable by the natural laws of baseball. Since the state of the count is updated on every single pitch, it is impossible to change the count by more than 1 strike or more than 1 ball at a time. For example, an 0-1 count can only transition to 1-1 (after a ball is thrown), 0-2 (after a strike is thrown), or a ball in play. In addition to this basic rule, 2-strike counts can transition to themselves again if the pitch is fouled off, all 3-ball counts can lead to a walk, all 2-strike counts can transition to strikeouts, and any count can become a ball in play. All told, the transition matrix for MLB only has 43 nonzero entries. For any given count, there are a minimum of 3 and a maximum of 4 states it can transition to.

Once we have this transition matrix for each season, how do we use it to predict strikeout and walk rates? Well, taking one iteration of the transition matrix simulates throwing one pitch. Taking two iterations of the matrix simulates throwing two pitches. The vast majority of plate appearances do not make it to 10 pitches, so taking 10 iterations would generally be enough to simulate a plate appearance. To be extra safe, however, we can just take the limit of the matrix instead, since a plate appearance is guaranteed to end by the time an infinite number of pitches have been thrown. The limit of a transition matrix is the point at which the probable outcomes stabilize at a stationary distribution [8]. In other words, the probabilities in the matrix will no longer change, no matter how many more iterations are taken. To find the limit, we raise the transition matrix to successively higher powers

that approach infinity. The limit of this matrix should look quite different from the transition matrix itself. Its only nonzero values will be found in the K, BB, and IP columns. This is because, after an infinite number of pitches, it is impossible to arrive at a count like 0-1 or 3-2. The only possible states are these “results” that never transition away from themselves.

After finding the limit of the transition matrix, we are only actually interested in the top row, the intersection of a 0-0 count with K and BB. These represent the chances for a random plate appearance starting at 0-0 to end in a strikeout or a walk. Since all plate appearances begin with a 0-0 count, these are the chances that any given batter will strike out or walk.

3.2 Building the Transition Matrix

	0-0	0-1	0-2	1-0	1-1	1-2	2-0	2-1	2-2	3-0	3-1	3-2	K	BB	IP	
0-0		A		B											D	
0-1			A		B										D	
0-2				E		B							C		D	
1-0					A		B								D	
1-1						A		B							D	
1-2							E		B				C		D	
2-0								A		B					D	
2-1									A		B				D	
2-2										E		B	C		D	
3-0											A			B	D	
3-1												A		B	D	
3-2													E	C	B	D
K															1	
BB															1	
IP															1	

Figure 3.1: A generalized transition matrix

Each of the 43 nonzero transition probabilities is marked with either a 1 or a capital letter in Figure 3.1, meaning that it is possible to transition from the count indicated by the left column to the one indicated by the top row. A 1 represents the intersection of either K, BB, or IP with itself. Since these are not technically counts, once a plate appearances reaches one of these states it stays there, with probability 1. Which letter appears at the intersection informs how to find the transition probability for that particular state change.

A means to sum the chance of swinging strikes, called strikes, and fouls. For example, the probability of transitioning from 0-0 to 0-1 is marked A, because in that situation a foul ball still nets the pitcher a strike.

B means the chance of a ball being thrown. For example, the probability of transitioning from any 3 ball count to a walk is marked B, because a walk can only occur when a pitch is called a ball.

C means to sum the chance of swinging strikes and called strikes. For example, the intersection of all 2 strike counts and the “strikeout” count are marked C, because a strikeout can be the result of either a called or swinging strike.

0.000	0.490	0.000	0.392	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.118
0.000	0.000	0.405	0.000	0.401	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.194
0.000	0.000	0.180	0.000	0.000	0.451	0.000	0.000	0.000	0.000	0.000	0.000	0.175	0.000	0.195
0.000	0.000	0.000	0.000	0.480	0.000	0.341	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.180
0.000	0.000	0.000	0.000	0.000	0.439	0.000	0.337	0.000	0.000	0.000	0.000	0.000	0.000	0.224
0.000	0.000	0.000	0.000	0.000	0.207	0.000	0.000	0.376	0.000	0.000	0.000	0.192	0.000	0.224
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.502	0.000	0.312	0.000	0.000	0.000	0.000	0.186
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.472	0.000	0.278	0.000	0.000	0.000	0.250
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.242	0.000	0.000	0.289	0.207	0.000	0.262
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.609	0.000	0.000	0.351	0.041
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.502	0.000	0.255	0.244
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.281	0.217	0.198	0.304
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Figure 3.2: The calculated state space for 2014

D means the chance of a ball in play being hit. Every real baseball count (0-0 to 3-2) is marked D where it intersects the ball in play (IP) column.

E means the chance of a foul ball being hit. For example, the 4 counts that can transition back to themselves are all marked E, as the only way to not change the count is to hit a 2 strike foul.

Using the plate discipline data from FanGraphs [6] and the trends by count data from BaseballSavant [7], I was able to calculate the entire state space for each season in Microsoft Excel. Transition probabilities marked by the same letter are not equal (because they are more realistic) because of the adjustments we have made by count, but they are usually in the same ballpark. The state space for 2014 can be seen in Figure 3.2.

Here is how to find the probability of each possible result of a pitch, using Zone, O-swing, Z-swing, O-contact, and Z-contact.

Swinging strikes: for a swinging strike to occur, the batter must swing and must not make contact.

$$(Zone) * (Z_{swing}) * (1 - Z_{contact}) + (OOZ) * (O_{swing}) * (1 - O_{contact})$$

Called strikes: for a called strike to occur, the pitch must be in the zone and the batter must not swing.

$$(Zone) * (1 - Z_{swing})$$

Balls: for a ball to occur, the pitch must be outside the zone and the batter must not swing.

$$(OOZ) * (1 - O_{swing})$$

Fouls/Balls in play: for a foul ball or a ball in play to occur, the batter must swing and make contact. It is possible to find the chance of a ball in play in a season, and thus the chance of a foul ball, by taking the total number of batters faced in a season and subtracting out the strikeouts, walks, and hit batsmen and dividing by total batter faced.

Balls in play:

$$BIP = \frac{(TBF) - (K) - (BB) - (HBP)}{TBF}$$

So to find the chance of a foul ball we can calculate the chance of a contacted ball and subtract out the chance it is in play.

0.000	0.416	0.000	0.467	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.118
0.000	0.000	0.373	0.000	0.433	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.194
0.000	0.000	0.180	0.000	0.000	0.467	0.000	0.000	0.000	0.000	0.000	0.000	0.159	0.000	0.195
0.000	0.000	0.000	0.000	0.413	0.000	0.407	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.180
0.000	0.000	0.000	0.000	0.000	0.401	0.000	0.375	0.000	0.000	0.000	0.000	0.000	0.000	0.224
0.000	0.000	0.000	0.000	0.000	0.207	0.000	0.000	0.396	0.000	0.000	0.000	0.173	0.000	0.224
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.429	0.000	0.385	0.000	0.000	0.000	0.000	0.186
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.430	0.000	0.320	0.000	0.000	0.000	0.250
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.242	0.000	0.000	0.314	0.182	0.000	0.262
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.489	0.000	0.000	0.471	0.041
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.449	0.000	0.307	0.244
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.281	0.190	0.225	0.304
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Figure 3.3: The state space for a hypothetical 2014 with an automated zone

$$Fouls = (Zone) * (Z_{swing}) * (Z_{contact}) + (OOZ) * (O_{swing}) * (O_{contact}) - BIP$$

3.3 Simulating a Hypothetical 2014

So what would the 2014 transition matrix look like if the strike zone were automated? Let's create a new one for a hypothetical season in some alternate universe where PITCHf/x was used to automate the strike zone before the 2014 season. We can assume that any result which includes a swing will be unchanged. Hitters may sometimes knowingly chase pitches out of the zone because they are unsure if the umpire will make the correct call, but I think this is a negligible amount and that they are usually just fooled. This leaves balls (all balls are called by the umpire) and called strikes as the only possible areas of change. Since there are only two, our job is even easier because calculating the change in one will also necessarily be the change in the other.

Using the BaseballSavant data [7] on umpire accuracy and the 2014 transition matrix, we can recalculate the expected chances of a called strike or called ball for each count if every call were according to PITCHf/x. To find the new chance of a called strike, just take the actual chance of a called strike, add the chance of an incorrectly called ball (because PITCHf/x would call it a strike), and subtract the chance of an incorrectly called strike (because PITCHf/x would not make that mistake). Since we are not changing the total number of called pitches, called balls will increase by the exact amount that called strikes decrease in every count. This means that only the transition probabilities marked by D or E are the unchanged from the actual 2014 transition matrix, as those marked A, B, or C all are affected by either called balls or called strikes. This new matrix can be seen in Figure 3.3.

I also created a hypothetical matrix like this for every season 2007-2013, using the 2014 data, but because of that extrapolation only the 2014 results will be focused on. With all 16 transition matrices created (one real and one hypothetical for each of the 8 seasons), I turned to MATLAB to find the limits of these gigantic matrices for me. The code I used can be found in section 7.1.

Chapter 4

Results

As stated, I used MATLAB to find the limits of these matrices and extract the values we are interested in, as detailed in the section about Markov chains.

Ideally, in the limit of the real matrix these values would be the actual strikeout and walk rates in MLB for each year. For the limit of the hypothetical matrix, these values would, in theory, be the strikeout and walk rates MLB could have expected had it implemented a fully-automated strike zone in that year.

In Figure 4.1, the columns labeled “K” and “BB” contain these values from the limit of the first matrix each year, the one using actual data. The columns labeled “K(pfx)” and “BB(pfx)” contain the values from the limit of the second matrix each year, the one that attempts to account for umpires making incorrect calls. (pfx) stands for PITCHf/x, as these strikeout and walk rates are the ones we would expect if PITCHf/x replaced umpires for calling balls and strikes.

For reasons previously mentioned, I will only be looking at 2014 for analysis. The first thing to notice is how close the Markov-predicted rates (20.9 K% and 7.6 BB%) are to the rates actually seen in MLB (20.4 K% and 7.0 BB%). This is a good sign. The model predicts a 4 percentage point decrease in strikeouts (to 16.9%) and a corresponding 4 percentage point increase in walks (to 11.0%) if the strike zone had been automated in 2014.

To a baseball fan, these seem like pretty drastic changes; what kind of effect would they have on run-scoring? Luckily, we have a handy statistic to estimate this: Fielding Independent Pitching (FIP). FIP is an ERA estimator based on only 4 inputs: home runs, strikeouts, walks, and innings pitched. In this hypothetical 2014 with an automated strike zone, there is no reason to think that home runs or innings would be changed significantly. So we can simply take the 2014 hypothetical strikeout and

'Year'	'K'	'BB'	'K(pfx)'	'BB(pfx)'	K (MLB)	BB (MLB)
2007	20.4%	7.0%	16.1%	11.2%	17.1%	8.5%
2008	20.1%	7.0%	15.9%	11.2%	17.5%	8.7%
2009	21.0%	6.9%	16.3%	11.4%	18.0%	8.9%
2010	21.4%	6.8%	16.7%	11.2%	18.5%	8.5%
2011	20.8%	6.7%	16.5%	11.0%	18.6%	8.1%
2012	21.3%	6.9%	17.0%	11.2%	19.8%	8.0%
2013	21.2%	6.9%	17.0%	11.0%	19.9%	7.9%
2014	20.9%	7.0%	16.9%	11.0%	20.4%	7.6%

Figure 4.1: Markov-predicted strikeout and walk rates

walk rates, multiply them by batters faced to get strikeout and walk totals, and then recalculate FIP with these new inputs. 2014's FIP would be expected to increase from 3.80 to 4.72, or nearly a full run per game! So for anyone who has been missing the high scoring games of the late 1990s and early 2000s, automating the strike zone could be a welcome breath of fresh, run-producing air.

Chapter 5

Survey

The Survey

1. During the MLB season, about how many hours a week do you spending watching or listening to baseball?
2. How many major league baseball games do you typically attend during the season?
3. An average MLB game features about 300 pitches, approximately half of which are called by the umpire. How many of these 150 calls do you think umpires get wrong, on average?
4. Regardless of your answer to #2, what do you think is the highest acceptable number of these missed calls?
5. Are you in favor of a fully automated strike zone? Y/N
6. Are you in favor of a semi-automated strike zone, in which the umpire is informed by the system, but can still make his own calls? Y/N
7. Do you think switching to an automated strike zone would be more beneficial to hitters, pitchers, or neither?
8. If an automated strike zone were guaranteed to increase offense, would you support the change?
9. MLB has just concluded its first season using instant replay. Has the introduction of replay affected your interest in watching baseball?

Using SurveyMonkey [9], I surveyed 100 people in order to glean some insight into others' perception of the game. How good a job do they think umpires do? Do they think there is room for improvement? Should that improvement come through technology or just better training for umpires? I emailed the survey to WPI students, and to ensure that people would respond, I did some research on how to conduct a survey [10]. I found that online surveys should be no more than 5 pages, especially not without some sort of progress bar. The respondent should not have to scroll very much on each page and a little color to look at while they answer never hurts. I also learned that possible responses to survey questions should be kept simple and, if at all possible, should be choices to click rather than open ended responses, as these can allow user error to weaken your results.

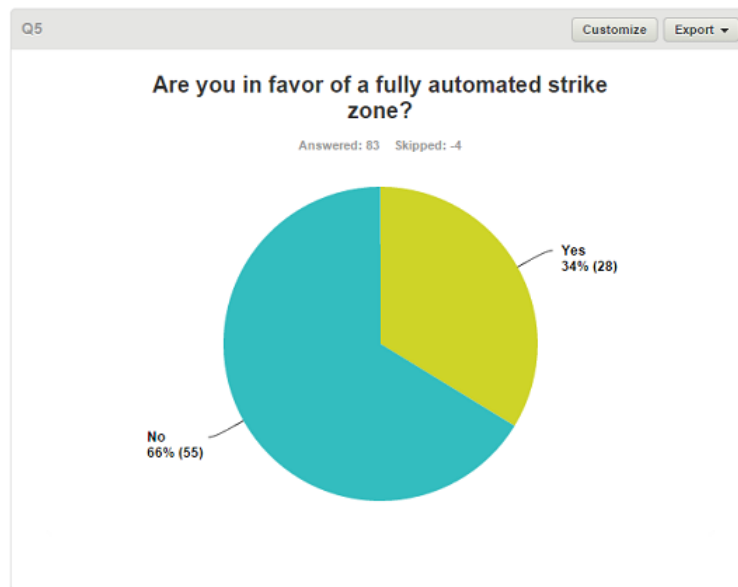


Figure 5.1: Those who indicated watching some amount of baseball opposed a fully automated zone

Due to incomplete responses, I threw out 4 of the surveys and was left with 96 complete responses to analyze. To provide a different and perhaps more informed viewpoint, I also conducted an interview with Mike Callahan, WPI’s head baseball coach and former minor league umpire, who could be considered a baseball insider (transcribed in full in section 7.2).

One aspect of the survey that was particularly interesting to me was examining the opinions of those who would likely proclaim themselves “baseball fans.” One of the possible responses to question 9 is “I don’t watch baseball.” Using this as a proxy for possibly uninformed opinions, we can filter the data and see only the views of those who care at least somewhat about the game. When this is done, we are left with 83 respondents. These respondents were against a fully automated strike zone by a 2:1 margin (Figure 5.1), but in favor of a semi-automated zone by the same margin (Figure 5.2). These are about the same ratios seen in the entire survey, not just the baseball fans.

The questions I most looked forward to seeing the results for were #3 and #4, regarding perceptions of how umpires currently perform and how they (or some new technology) SHOULD perform. Overwhelmingly, respondents felt that there are too many bad calls in today’s game. 28% of respondents believe that umpires miss more than 15 calls per game right now, but only 7% agree that that is an acceptable number. Only 39% of people think umps are making 10 or fewer mistakes per game, but a whopping 79% believe that is how good they should be.

Let’s attempt to quantify this disparity. The possible answers to these questions are all ranges of numbers, like 1-5 or 6-10. But working with ranges of numbers can be confusing and is not necessary. Instead, consider every response to be numerically equal to the lowest number in the selected range (so a person who answered 11-15 is recorded as simply 11). It could also be recorded as the highest number in the range, or the middle one, but the specifics do not matter and this is an inexact process anyway. This gives an average of around 10 missed calls for question #3 (Figure 5.3) and around 5 for question #4 (Figure 5.4). So the average survey respondent believes there are up to twice as many missed calls today compared to what they find acceptable. These figures run contrary to the opinion of Coach Callahan. While he did not put a number on it, he expressed the sentiment that MLB umpires are currently doing an excellent job calling balls and strikes accurately. In addition, he noted that consistency is an important factor as well. If an umpire makes 10 incorrect calls per game, a batter

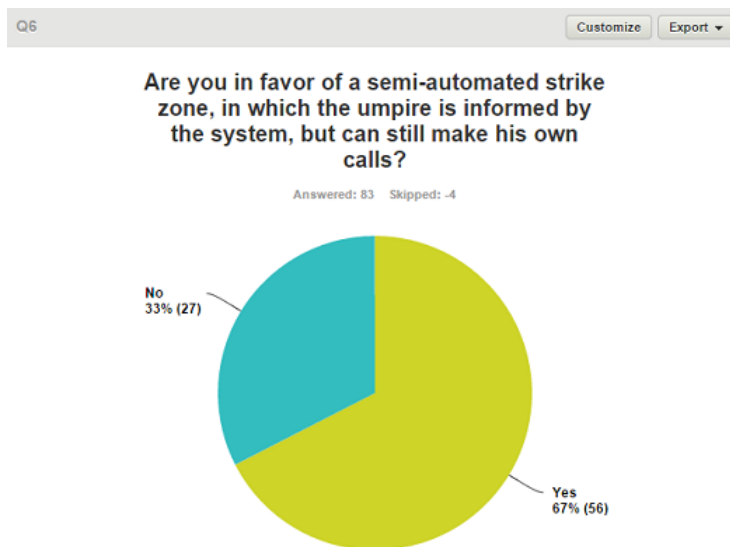


Figure 5.2: Just as strongly, however, they supported a semi-automated zone

would certainly rather they be on similar pitches than randomly distributed (i.e. the umpire tends to call strikes on pitches that are too low, but not those that are too high).

Not surprisingly, this led to much support for some form of technology in calling balls and strikes. Of the 96 respondents, 28% were in favor of either a fully or a semi-automated zone, 8% were only in favor of a fully automated zone, 39% were only in favor of a semi-automated zone, and 25% were against a fully or semi-automated strike zone. Despite effectively being a sample size of one, Coach Callahan’s responses were consistent with these findings. Those surveyed were generally in favor of some sort of technology to assist umpires, but they were not willing to abandon the current system to let PITCHf/x make the calls. Coach Callahan seemed open to the idea of using technology to improve umpire accuracy, but admitted he was not sure how he would feel about making big changes to the foundation of the game, citing the “human element” in particular.

It seems there are a few folks who just can’t be pleased, though. Of the 25% who were against any type of automated strike zone, more than a quarter of them think that any number of bad calls that can’t be counted on one hand is unacceptable. It is foolish to think that umpires today represent the pinnacle of human strike-calling ability, but they are likely very close. Umpires have been around for over 100 years and have only had to undergo more rigorous training and scrutiny as time passes. It seems extremely unlikely that they will ever overcome the physical limitations that involve tracking 100 MPH pitches, so expecting their error rate to drop to the 3% these respondents require is just not very feasible.

Breaking the survey respondents into two groups based on how many missed calls they find acceptable provides some fascinating results. One group is those who think more than 5 missed calls is unacceptable, and the other group is everyone else. There were 43 people in the first group, and they exhibit some distinct response patterns that go against the results of the survey as a whole. They actually supported a fully automated zone, 23 to 20 (Figure 5.5), and even more so if it were guaranteed to bring more offense, 29 to 14 (Figure 5.6). These patterns make sense, but are strikingly different from those more lenient about missed calls. They were against an automated zone, 44 to 13 (Figure

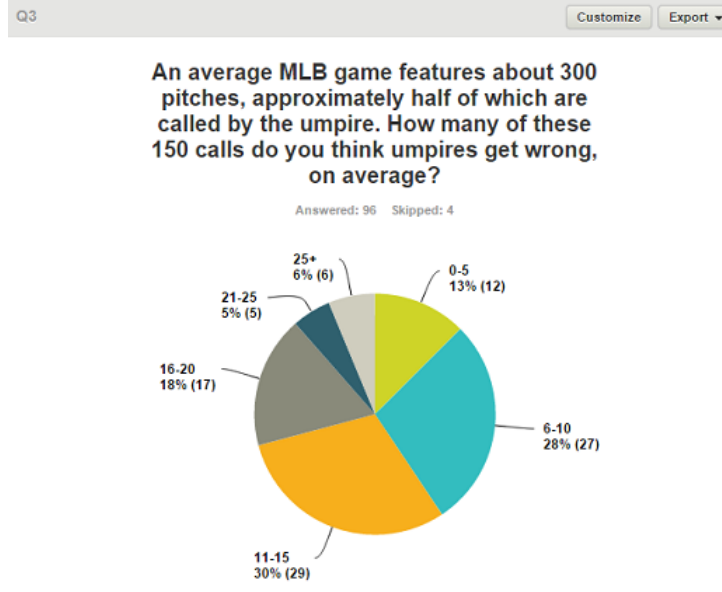


Figure 5.3: Survey respondents think umpires today average 10-15 missed calls per game

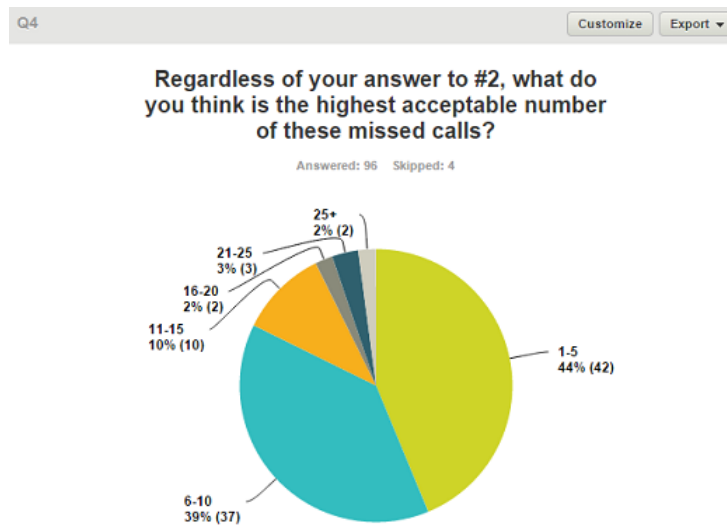


Figure 5.4: Generally, though, they believe that 5-10 missed calls is an acceptable number

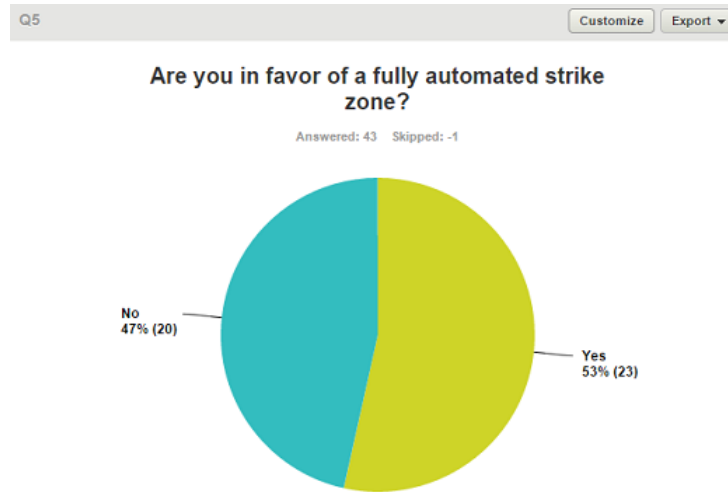


Figure 5.5: Those who demanded very few missed calls were the only group to support automating the strike zone

5.7), and did not change their tune much if it would bring more offense, being against that 38 to 19 (Figure 5.8).

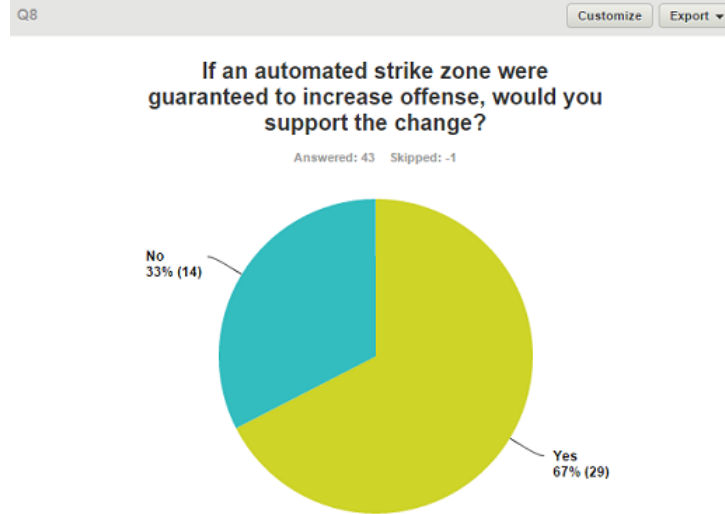


Figure 5.6: They were also the only group to support the change if it meant more offense

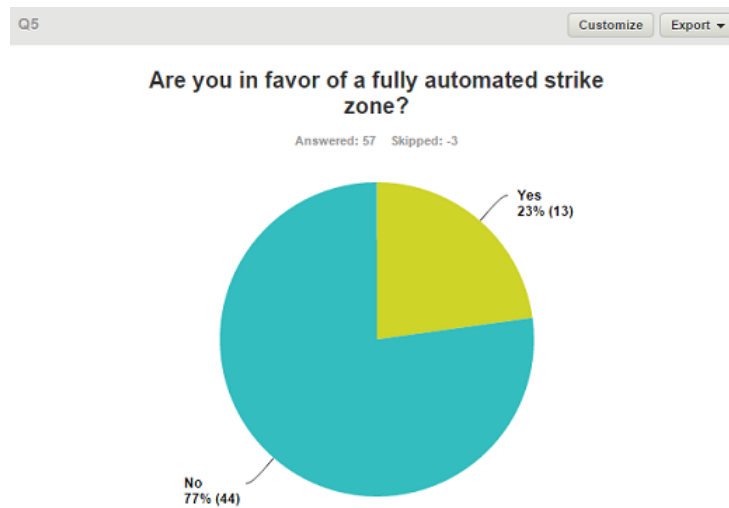


Figure 5.7: Those who are willing to accept more bad calls were strongly opposed to a fully automated zone

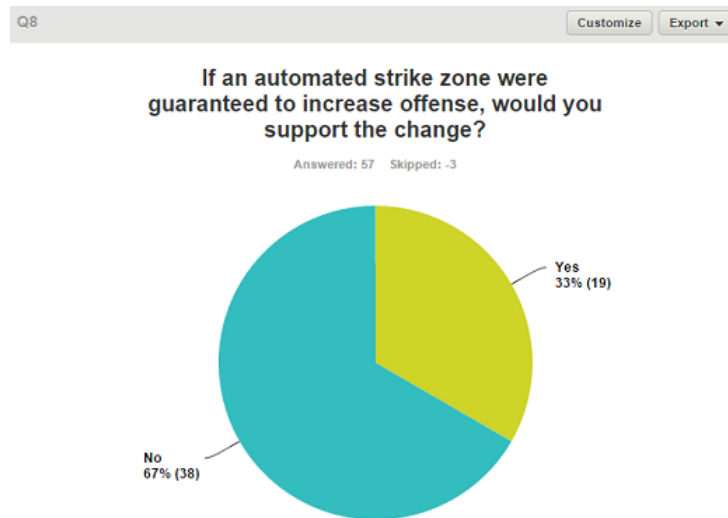


Figure 5.8: In general, their responses were more in line with the general survey results

Chapter 6

Discussion

Are these figures what we should actually expect to see, though? The league strikeout rate has not been that low in 15 years, and the league walk rate has never reached 11%, topping out at 10.5% in 1949. I strongly suspect that pitchers would begin to adapt almost immediately to counteract the huge increase in walks. In doing so, they would also probably lose a few more percentage points from their strikeout rates, since the path to allowing fewer walks is through throwing more hittable pitches. So the increase in walks may not be quite as large as predicted by the Markov model, but the decrease in strikeouts may very well be even larger.

In addition to the fact that players will adjust their tendencies with an automated strike zone, there are a few other caveats to consider about this project. While the adjustments made for zone and swing rates are important and have a significant effect, there are still more possible factors to be accounted for, most importantly the handedness of the batter and pitcher. The preceding pitch is also likely to have some influence, as pitchers are generally less willing to throw the same pitch twice in a row than random chance would suggest. If the prior pitch was the pitcher's weakest pitch type, we may expect a higher chance of a strike than usual in a certain count. However, this would make the process technically not a Markov chain, as its transition probabilities would not be independent of past states. It would also be incredibly difficult to implement for what seems like a marginal improvement.

The future implications of this study are nearly endless, though. With the ability to predict strikeout and walk rates from more granular data that stabilizes more quickly, a Markov analysis like this could be used to determine if an overperforming pitcher could keep up his performance or if an aging vet is actually done for or could make a comeback. Markov chains could also be applied at the plate appearance level and use base-out states rather than states of the count to better estimate a pitcher's "deserved" ERA.

Unfortunately, however, it seems that we may have to wait some time before the baseball world at large is ready for automated strike zones. The responses from my survey indicated that baseball fans are not yet ready to take pitch calling out of the hands of umpires. And perhaps this is for the better, as the expected strikeout and walk rates I found are dramatically different from the rates we see today. Such a large and sudden change could produce unintended effects, and it is difficult to blame those fans for being wary of a major change.

They are willing, though, to give the umpire some help in the form of a semi-automated strike zone. This could be achieved by simply having someone relay the PITCHf/x call to the home plate umpire via headset. This would eliminate the most egregious calls and still allow some of the "human element" that has always been such a big part of the game. As the umpires grow more comfortable with this system, they may begin to defer to the headset call more and more often, effectively introducing a fully automated strike zone over a longer period of time.

As noted in chapter 5, I also interviewed Mike Callahan, WPI's head baseball coach and former minor league umpire (transcribed in full in section 7.2). I did not ask the same questions used in the survey, but our conversation revealed that he was against a fully automated strike zone, while being open to something perhaps less extreme that umpires could use for confirmation. This is consistent with the views expressed by the survey respondents, and perhaps lends further weight to the opinions of those who may not be as knowledgeable about baseball as he is.

Chapter 7

Appendix

7.1 MATLAB code

```
fileroot = 'IQP';
xlRange = 'A1:Q12';
MATLABquirk='.xlsx';
C=cell(9,5);
C(1,1)=cellstr('Year');
C(1,2)=cellstr('K');
C(1,3)=cellstr('BB');
C(1,4)=cellstr('K(pfx)');
C(1,5)=cellstr('BB(pfx)');
for k=7:14
    nst=int2str(200+k);
    st=(200+k);
    filename=[fileroot nst MATLABquirk];
M = xlsread(filename, xlRange);
[ndata, text, alldata] = xlsread(filename);

Z=zeros(15);
for i=1:11
    Z(i,i+1)=M(i,1);
end

Z(3,4)=0;
Z(6,7)=0;
Z(9,10)=0;

for i=1:9
    Z(i,i+3)=M(i,2);
end

for i=10:12
    Z(i,14)=M(i,2);
```

```

end

for i=1:4
    Z(3*i,13)=M(3*i,3);
    Z(3*i,3*i)=M(3*i,4);
end

for i=1:12
    Z(i,15)=M(i,5);
end

for i=13:15
    Z(i,i)=1;
end

A=Z^15;
C(k-5,2)=num2cell(A(1,13));
C(k-5,3)=num2cell(A(1,14));
C(k-5,1)=num2cell(2000+k);

for i=1:11
    Z(i,i+1)=M(i,13);
end

Z(3,4)=0;
Z(6,7)=0;
Z(9,10)=0;

for i=1:9
    Z(i,i+3)=M(i,14);
end

for i=10:12
    Z(i,14)=M(i,14);
end

for i=1:4
    Z(3*i,13)=M(3*i,15);
    Z(3*i,3*i)=M(3*i,16);
end

for i=1:12
    Z(i,15)=M(i,17);
end

```

```

for i=13:15
    Z(i,i)=1;
end

B=Z^15;
C(k-5,4)=num2cell(B(1,13));
C(k-5,5)=num2cell(B(1,14));
end

C

```

7.2 Interview transcript

I managed to get WPI head coach and former college umpire Mike Callahan to sit down for a chat about the possibility of automating MLB's strike zone. Here is our conversation.

Joe: So, are you for automating the strike zone?

Coach Callahan: No, I don't think I would be. I think human error is kind of part of baseball and automating the strike zone - the game might just become too robotic I guess is the way I would put it.

J: I'm looking into full automation, where a system would make the call no matter what, or perhaps one where the umpire still makes the initial call, but then has a system-generated call relayed to him for confirmation.

C: I guess that's a little bit like the replay that they have now, but I wouldn't want to see the entire game fully automated. Like I said, the game has always been partly human error, but I don't know if it wouldn't take away from it a little. It would make the game 100 percent correct I guess, but it might make it a little boring. I don't really know how I would feel about doing the whole thing.

J: So, in regards to the human error that comes from umpires making the calls, how good of a job do you think they're doing now? Are they getting almost every call correct?

C: I mean, if you watch when they show the replays I bet you there were more calls that were upheld than were overturned when they went to review it, so I think they do a pretty good job. It is a pretty slow game.

J: What about the calls umpires make purely on balls and strikes, rather than plays in the field?

C: I think they get more right than they miss for sure, but there are some guys that have a bigger zone than others, so I think part of it is knowing what umpire you have and knowing if he's got a big strike zone, or he's got a small strike zone, or he's going to call the ball high vs. call the ball low.

J: So you're more concerned with consistency.

C: Yeah. And like I said, I'm for the replays on home runs and balls and strikes and diving catches. I'm okay with it all.

J: So you're for replay if it determines the finality of the play?

C: Yeah, pretty much, but even still if the umpire at first base calls him out, he's out. But if it's a diving play and it's not clear if he traps the ball or catches it, I'm fine with reviewing that, because it could be one of the best catches you've ever seen. But then you can go back to the play at first base that happened two years ago where he almost threw the perfect game, and that could be a good reason why you can argue both sides. Being a coach and playing my whole life though, I'm just not 100 percent for it or against it. I'm kind of somewhere in between.

J: In your experience, is integrating technology into sports generally good for the sport?

C: Like I said, I don't know if I'm for it or against it, but I guess any time you can make sure things are called correctly more often it's a pretty good argument for why you should have it. We don't have that kind of technology at the college level so everything is based on human error. I would be more apt to use technology at our level than I would be at a professional level because our umps aren't professional. MLB umpires are better than ours so it would probably benefit us more to make calls correct. We also only have two umpires in college vs. four or six for MLB so there are already a lot more eyes out there for them.

J: And if you don't have an automated strike zone at any of the lower levels, it's possible that players would have a big adjustment to make when they reach the major leagues and the strike zone is automated.

C: Exactly. I don't know to what degree I would want it; the entire game is pretty big on tradition.

J: It does have that charm. I'm a huge baseball fan, too, and I generally don't like change. There are a lot of articles being written now that go into detail about the criticisms of the umps, but for the most part, I think umpires do a great job.

C: Yeah, especially at a professional level. Like I said, if you come and watch some college games you'd shake your head and wonder how you can get paid and miss so many calls. That's why I think it would probably benefit us more because I bet 75-80 percent were probably the right call in the majors.

Bibliography

- [1] D. Boyle, B. Hicks, D. Kinney, T. Larwin, L. Leonard, A. McCue, F. Rodgers, and A. Strasberg. (2011) Observations of Umpires at Work. [Online]. Available: <http://sabr.org/research/observations-umpires-work>
- [2] B. Bukiet, E. Harold, and J. Palacios, “A markov chain approach to baseball,” *Operations Research*, vol. 45, pp. 14–23, 1997.
- [3] “Sportvision,” 2014. [Online]. Available: <http://www.sportvision.com/baseball/pitchfx>
- [4] M. Newman. (2007, Oct.) Pitch-f/x adds insight to watching games. [Online]. Available: <http://m.mlb.com/news/article/2245402/>
- [5] “New York Times,” 2011. [Online]. Available: <http://goal.blogs.nytimes.com/2011/08/07/vanishing-spray-and-the-future-of-technology>
- [6] D. Appelman, “Fangraphs,” 2014. [Online]. Available: <http://www.fangraphs.com>
- [7] D. Willman, “BaseballSavant,” 2014. [Online]. Available: <http://www.baseballsavant.com>
- [8] D. Lay, *Linear Algebra and its Applications*, 3rd ed. Pearson Education, 2006.
- [9] D. Goldberg, “SurveyMonkey,” 2014. [Online]. Available: <http://www.surveymonkey.com>
- [10] A. Fink, *How to ask survey questions*, 2nd ed. Sage Publications, 2003.