UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

HONORS THESIS

THE ANTI-CYCLE IN BASEBALL

by

Peter Previte

May 2019

Thesis Advisor:	Dr. William Pulleyblank
Thesis Advisor:	Major Dusty Turner
Thesis Advisor:	Major Daniel Baller

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** May 2019	**3. REPORT TYPE AND DATES COVERED**

4. TITLE AND SUBTITLE The Anti-Cycle in Baseball

5. FUNDING NUMBERS

6. AUTHOR(S) Peter Previte

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
United States Military Academy
West Point, NY 10996

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
NA

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

In baseball, a batter hits for the cycle when he hits a single, double, triple, and home-run all in a single game. Since the inception of Major League Baseball (MLB) in 1876, this has only occurred 324 times, less than three times per season on average. We propose a new statistic which replaces the successes of the cycle with failures. We define the anti-cycle to occur when a single player records an out at first base, second base, third base, and home plate all in a single game. Our fundamental question is how often does the anti-cycle occur?

A first step in answering this question was to define a logical and consistent way of associating any out made by a player with a base to which we attribute the out. In some cases, this is obvious (player thrown out at first base), but in other cases it can be subtler (a player caught in a rundown right in the middle of two bases).

The database called PITCHf/x served as a source of data. It contains a text description of every plate appearance in every MLB game from 2008-2016. We developed code that analyzed this data over 131,000 baseball outs in total and determined when players hit for the anti-cycle. The anti-cycle occurred 176 times between the 2008 and 2016 seasons. Compared to the cycle, which occurred only 34 times during that time span, the anti-cycle occurred almost six times more frequently, on average.

After determining the number of anti-cycles that occurred, we analyzed the occurrences of recording outs at each base to determine if they are independent events. We did this through both theoretical and simulated approaches. The goal in determining if recording outs at each base are independent events was to gain insight as to if there is "interestingness" that exists within baseball games. Ultimately, we conclude that recording outs at each base are not independent events and there may be something "interesting" taking place.

14. SUBJECT TERMS			**15. NUMBER OF PAGES**
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT OF** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

THE ANTI-CYCLE IN BASEBALL

Peter Previte
Cadet, United States Army
B.S., United States Military Academy, 2019

Submitted in partial fulfillment of the
requirements for the degree of
BACHELOR OF SCIENCE
in **MATHEMATICAL SCIENCES**
with Honors
from the
UNITED STATES MILITARY ACADEMY
May 2019

Author: Peter Previte

Advisory Team: Dr. William Pulleyblank
Thesis Advisor

Major Dusty Turner
Thesis Advisor

Major Daniel Baller
Thesis Advisor

Colonel Tina Hartley
Chairman, Department of Mathematical Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In baseball, a batter hits for the cycle when he hits a single, double, triple, and home-run all in a single game. Since the inception of Major League Baseball (MLB) in 1876, this has only occurred 324 times, less than three times per season on average. We propose a new statistic which replaces the successes of the cycle with failures. We define the anti-cycle to occur when a single player records an out at first base, second base, third base, and home plate all in a single game. Our fundamental question is how often does the anti-cycle occur?

A first step in answering this question was to define a logical and consistent way of associating any out made by a player with a base to which we attribute the out. In some cases, this is obvious (player thrown out at first base), but in other cases it can be subtler (a player caught in a rundown right in the middle of two bases).

The database called PITCHf/x served as a source of data. It contains a text description of every plate appearance in every MLB game from 2008-2016. We developed code that analyzed this data over 131,000 baseball outs in total and determined when players hit for the anti-cycle. The anti-cycle occurred 176 times between the 2008 and 2016 seasons. Compared to the cycle, which occurred only 34 times during that time span, the anti-cycle occurred almost six times more frequently, on average.

After determining the number of anti-cycles that occurred, we analyzed the occurrences of recording outs at each base to determine if they are independent events. We did this through both theoretical and simulated approaches. The goal in determining if recording outs at each base are independent events was to gain insight as to if there is "interestingness" that exists within baseball games. Ultimately, we conclude that recording outs at each base are not independent events and there may be something "interesting" taking place.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This thesis would not have been possible without the mentorship and guidance of each of my advisers, Dr. William Pulleyblank, MAJ Dusty Turner, and MAJ Daniel Baller. I would also like to thank the Department of Mathematics at USMA for their incredible support during my four year experience at West Point. Specifically, I would like to thank LTC Kevin Cummiskey for helping with the creation of our theoretical simulation. Finally, I would like to thank CDT Jinwon Seo, Class of 2019, for his assistance in creating Python code to determine anti-cycle occurrences in our data.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

In baseball, a batter hits for the cycle when he hits a single, double, triple, and home-run all in a single game. Since the inception of Major League Baseball (MLB) in 1876, this has only occurred 324 times, less than three times per season on average. We are interested in an event that is opposite of the cycle, which we will call the anti-cycle. Where the cycle is focused hits, the anti-cycle is focused on recording outs. We define the anti-cycle as the event where a single player records an out at each base in a single game. That is, a player records an anti-cycle if he records an out at 1st base, 2nd base, 3rd base, and home plate in one game. With the anti-cycle define, our fundamental question is how often does the anti-cycle occur?

A first step in answering this question was to associate outs with bases. Specifically, whenever an out is recorded, we need to have a consistent and logical way of attributing that out to a base. In order to do this, a comprehensive list was made of every single possible way a player can record an out. This was made primarily through analysis of the official MLB rule book. Once this list was made, we created a method for attributing every single type of out to a base. Some are obvious; for example, every time a batter hits a fly ball that is caught, the batter's out is attributed to home plate. However, others are not so obvious; when a player is tagged out in a rundown, for example, we attribute that out to the base in which the player started the rundown.

We also define categories of the anti-cycle. First, a strong anti-cycle is one in which a player's out at home plate occurs after he has made his way around the bases. That is, we divide outs at home into two types: outs made before a player has made any progress around the bases, and outs made after a player has made it all the way around the bases. Therefore, in order to record a strong anti-cycle, a player must be called out at home after making his way around the bases. Next, an ordered anti-cycle is one in which a player's outs are recorded in increasing order.

In 2006, MLBAM and Sportvision installed pitch-tracking systems in every MLB stadium. The pitch-tracking systems recorded data on every single player, pitch, plate-appearance, and game from 2008 to 2016 and stored the data in a database called PITCHf/x. One piece of data stored in PITCHf/x is a text description of the result of every single plate appearance during games. These text descriptions follow distinct patterns; patterns that carry through the entire database. Because of their consistency, the text descriptions

provided us a way of finding outs in the data using character and string matching. Using a combination of R and Python, we were able to develop code that identifies where outs occur, attributes each out to the base at which it occurred, and identifies if a single player records an out at each base in a game.

After running our data through our code, we determine that the anti-cycle occurs 176 times between the years 2008 and 2016, just under 20 times per season. Compared to the cycle, which occurred only 34 times during that same nine year time span, the anti-cycle occurred almost six times more frequently, on average. In addition, one of the 176 anti-cycles were strong, and two were ordered.

Once we determined the number of anti-cycle occurrences, a further question of interest was whether the events of recording an out at each base are independent. That is, if we multiplied the probabilities of recording an out at each base together, would that equal the total probability of recording an anti-cycle? In order to answer this question, we first found the frequencies with which players record outs at each base. We then applied both theoretical and simulated approaches to determine the probability of recording an anti-cycle. With our results, we applied a chi-square test for independence to the observed frequencies in the data and the predicted results of the theoretical solution and the simulations. Ultimately, we conclude that the events of recording outs at each base are NOT independent and that there may be something "interesting" going on in within the game of baseball.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

June 28th, 2008. The Toronto Blue Jays lead the Atlanta Braves 7 to 5 in the bottom of the 8th inning. Vernon Wells comes to bat with men on 1st and 2nd. Having already been to bat three other times in the game, each ultimately resulting in being put out, Wells knew he could put a dagger into the hearts of the Braves with a clutch hit. With a 1-2 count, Wells struck a hard-hit line drive to left field, which would score the men on 1st and 2nd, extending the Blue Jays' lead to 9-5. In an attempt to make his contribution even greater, Wells was then thrown out advancing to 3rd. And in doing so, Wells accomplished a feat he probably never even thought about. Wells recorded an anti-cycle.

A. OVERVIEW

Initial research of any mention of the "anti-cycle" returned little results. The only mention of any type of anti-cycle that hours of research turned up was by Jon Marthaler, who discussed his idea of an anti-cycle. [2] However, Marthaler's idea of the anti-cycle is different than ours. Marthaler's definition of an anti-cycle has to do with a player recording four different mishaps in a single baseball game. For example, a player could strike out, hit into a double play, commit an error in the field, and get thrown out stealing. There are no four specific events that must happen, so the event of the anti-cycle according to this persons definition is up for interpretation. After hours of research, this is the only mention of any type of anti-cycle that others have looked into, but it is ultimately not the same as ours.

B. DEFINING THE ANTI-CYCLE

We first must clarify our definition of the anti-cycle. We define the anti-cycle in baseball as the event where a player records an out at each of the four bases in a single game. For example, that may mean the player strikes out (home plate), grounds out to the shortstop (1st base), gets caught stealing 2nd (2nd base), and gets picked off at 3rd (3rd base).

Our next task is to define every single way a player can record an out in baseball. The official MLB rule book includes a list of all the different ways an out can be recorded in baseball, so this was the primary source for that information.[3] Specifically, Part (b) of

Section 5.09 lists the various ways a player can record an out. In addition, a blog which discusses the different ways a player can record an out in baseball also provided a different explanation for some of the occurrences.[4]

Next we must define, for every possible circumstance where a player can record an out, the base in which the player recorded the out. In most cases, this will be fairly obvious. For example, if a player hits a ground ball to the shortstop and is thrown out at first, the player records an out at first base. Some cases are not so simple, however. For example, if a runner is tagged out in a rundown, is the runner considered out at the base in which they started, or the base to which they are running? This is one example of a tricky scenario which must be defined. Below is a list of all possible ways a player can record an out, and the base at which the out is recorded. The list is split into four categories: outs that can occur at any base, outs that can occur at any base except home plate, outs that can only occur at home plate, and miscellaneous outs.

1. Outs Which Can Occur at Any Base

- When a player is thrown out at a base which he was forced to run to, he is out at that base
- When a player is hit by a ground ball while running the bases before it passes an infielder, he is out at the base to which he was running
- When a player hits into a fielder's choice, the batter does not record an out, but the runner records an out at the base to which he was running*
- When a player is thrown out attempting to advance to the next base, whether tagging on a fly ball or taking an extra base on a base hit, he records an out at the base to which he was running or the base at which he was thrown out
- When a player runs out of the baseline to avoid a tag, he records an out at the base to which he is running
- When a player runs out of the baseline to interfere with a throw being made, he records an out at the base to which he was running

2. Outs Which Can Occur at Any Base Except Home Plate

- When a player is picked off, he records an out at the base from which he is picked off
- When a player is caught stealing, he records an out at the base to which he is running

- When a player is tagged out in a rundown, he records an out at the base in which he began the rundown
- When a player misses a base and touches the next base before correcting himself by going back to touch the missed base, he records an out at the missed base
- When a player passes a different runner on the base paths, the runner who made the pass records an out at the base from which he is coming

3. Outs Which Can Only Occur at Home Plate

- When a player strikes out, he records an out at home plate
- When a batted ball is caught by a fielder, whether it be a fly ball, pop out, line out, sacrifice fly, or foul ball, the batter records an out at home plate
- When the catcher drops a called third strike, the batter is out at home plate, regardless of if they are tagged or if the catcher throws to first in time
- When a player is hit by a ground ball while touching a base, the batter records an out at home plate
- When a player misses home plate as he is scoring, once he reaches the dugout or the fielding team touches home plate with the ball, the player records an out at home plate
- When a player is put out due to batting out-of-order, the batter is put out at home plate
- When a player uses an illegal bat, they are out at home plate
- When a player makes contact with a pitch or a foul ball while outside the batter's box or standing on home plate, the batter is out at home plate

4. Miscellaneous

- When a player is physically assisted by a base coach while on the base paths, the runner records an out at the base at which the coach is located (can only occur at first or third base)

C. ANTI-CYCLE SPECIFICS AND CATEGORIZATIONS

Since we have defined the anti-cycle as the event where a single player records an out at each base in a single game, an example of an anti-cycle occurrence could then look

like this: a player strikes out, grounds out to the shortstop, gets caught stealing 2nd base, and gets picked off at 3rd base, all in a single game. Here, striking out is the out at home plate, grounding out is the out at 1st base, getting caught stealing is the out at 2nd base, and getting picked off is the out at 3rd base. Every out is attributed to exactly one base, and the values for each base are: 0 = home plate, 1 = 1st base, 2 = 2nd base, 3 = 3rd base, 4 = home plate.

Notice how an out at home plate can have a value of either 0 or 4. The distinction is due to the progress the runner has made and the type of the out at home plate. If a batter strikes out, their out will have a value of 0 because they have made no progress around the bases at all. However, if a runner is thrown out trying to steal home, his out will have a value of 4 because he has made it all the way around the bases.

In addition, we categorize each out based on if it is "earned" or "unearned". An "unearned" out is an out in which the circumstance behind a player's recording of an out is not that player's fault. For example, if there is a runner on 1st base and the batter hits into a standard ground-ball double play, the runner's out is considered unearned, because he was required to run to 2nd base regardless of his own actions. In other words, there was nothing the runner could have done to prevent his being out. The batter's out at 1st base, on the other hand, is earned, because his being thrown out is a direct result of his actions at the plate. With this in mind, we will consider an "earned" anti-cycle to be an anti-cycle where all of the outs recorded by the player are earned outs.

Further, we also divide the anti-cycle into two more categories: strong vs. weak, and ordered vs. unordered. A strong anti-cycle is one in which a players recorded outs contain the values 1, 2, 3, and 4. A weak anti-cycle is one in which a players recorded outs contain the values 0, 1, 2, and 3. Next, an ordered anti-cycle is one in which a players outs are recorded in order from smallest to largest, while an unordered anti-cycle is one in which a players outs are recorded in no order.

II. THE DATA

A. SOURCE

In 2006, Major League Baseball Advanced Media (MLBAM) and Sportvision worked together to install a pitch tracking system in every MLB stadium.[5] The system, known as PITCHf/x, would soon revolutionize the field of sabermetrics as well as the way people watch baseball. From 2008 to 2016, PITCHf/x maintained a database on every pitch, at-bat, and player for every MLB game (these companies continue to collect the same data now but the current database is called ”Statcast” and collects even more data for every game than PITCHf/x used to). Particularly, I was interested in the at-bat table because it includes the most relevant information for trying to determine the results of players plate appearances for each game. For reference, the database with only at-bat information for the years 2008 to 2016, in an SQL database, is 6.7 GB.

As mentioned earlier, the most useful data table in the PITCHf/x database is the at-bat database. For the remainder of this paper, when not referring directly to the at-bat table in PITCHf/x, I will use the term plate appearance as opposed to ”at-bat,” because we care about plate appearances that result in things like walks and sacrifices. For example, an at-bat where a player walks is still an entry in the at-bat table and we care about the result of that walk because we want to know if that player eventually recorded an out. However, in MLB record-keeping, a walk is not counted on a players scorecard as an at-bat.

B. PATTERNS

One of the data types in the at-bat table is a written description of the outcome of each plate appearance. For example, one at-bat description reads, “Lyle Overbay grounds into double play, second baseman Andy Cannizaro to shortstop Elliot Johnson to first baseman Carlos Pena. Shannon Stewart out at 2nd.” Every single out has a distinct pattern for the way it is recorded in the database. Every time a ground ball occurs and the batter is thrown out at 1st Base, for example, the description will read ”(Batter Name) grounds out, (Name of Fielder who threw the ball) to first basemen (Name of first basemen). These patterns remain consistent throughout the entire database. This means that we can use string and character matching to identify when an out occurred and to which base we would attribute each out.

THIS PAGE INTENTIONALLY LEFT BLANK

III. USING RSTUDIO

A. ACCESSING THE DATA

RStudio was the primary tool used to analyze the data. Using an article by Carson Sievert[6], we were able to begin the process of accessing the PITCHf/x data and downloading it for use in R. Sievert created a package within R that can scrape the PITCHf/x data into an SQL database.[7] Then, his package "pitchRx" allows us to do the majority of our data manipulations and analysis directly in R. After scraping each seasons worth of data into R from 2008-2016, the data was exported into a Microsoft Excel file so the contents of the data could be visualized. For reference, one seasons worth of data in an Excel file is 6,543,402 entries, and each entry contains 33 column's worth of information.

Before we began manipulating the data in R, we needed to ensure we understood exactly what the desired output from the data would look like. We determined that, from the unparsed data, the important information we need to collect about each out is the inning in which it occurred, the side (home or away) and name of the player's team, the players name, the base at which the player recorded the out, the game ID for that game, and whether the out was earned or unearned. With this framework in mind, we created an example of what one game would look like by hand. To do this, each at-bat for one game was visually checked and a table was created with the information we need for each out. This example is shown below.

Once we knew what type of information we wanted to collect about each out, we began to create our code that would collect all this information and identify where anti-cycles occur.

B. METHODOLOGY

As mentioned earlier, the primary method of analyzing the data we used is character and string matching. In R, this is done primarily using the tidyverse package, created by Hadley Wickam[8]. The steps taken in order to get the data in the desired output are:

1. Separate each sentence of the at-bat description into its own entry.
2. Identify the entries which describe an out occurring.
3. Isolate the entries identified in Step 2 so we only have entries where an out occurs.

inning	team_location	team_ID	player_name	base_out	gameday_link	earned_out
1	Away	BOS	Kevin Youkilis	1	gid_2008_03_25_bosmlb_oakmlb_1	0
1	Away	BOS	David Ortiz	0	gid_2008_03_25_bosmlb_oakmlb_1	0
1	Away	BOS	Manny Ramirez	0	gid_2008_03_25_bosmlb_oakmlb_1	0
1	Home	OAK	Travis Buck	1	gid_2008_03_25_bosmlb_oakmlb_1	0
1	Home	OAK	Bobby Crosby	1	gid_2008_03_25_bosmlb_oakmlb_1	0
1	Home	OAK	Jack Hannahan	0	gid_2008_03_25_bosmlb_oakmlb_1	0
2	Away	BOS	Mike Lowell	2	gid_2008_03_25_bosmlb_oakmlb_1	1
2	Away	BOS	Jason Varitek	1	gid_2008_03_25_bosmlb_oakmlb_1	0
2	Away	BOS	Brandon Moss	2	gid_2008_03_25_bosmlb_oakmlb_1	1
2	Home	OAK	Ryan Sweeney	0	gid_2008_03_25_bosmlb_oakmlb_1	0
2	Home	OAK	Travis Buck	0	gid_2008_03_25_bosmlb_oakmlb_1	0
2	Home	OAK	Jack Cust	0	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Away	BOS	Jacoby Ellsbury	0	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Away	BOS	Dustin Pedroia	0	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Away	BOS	David Ortiz	1	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Home	OAK	Emil Brown	0	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Home	OAK	Bobby Crosby	1	gid_2008_03_25_bosmlb_oakmlb_1	0
3	Home	OAK	Kurt Suzuki	0	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Away	BOS	Manny Ramirez	0	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Away	BOS	Mike Lowell	0	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Away	BOS	Brandon Moss	1	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Home	OAK	Ryan Sweeney	0	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Home	OAK	Travis Buck	0	gid_2008_03_25_bosmlb_oakmlb_1	0
4	Home	OAK	Mark Ellis	1	gid_2008_03_25_bosmlb_oakmlb_1	0

Figure 1. Example Output of First Game Created by Hand

4. Attribute each out to the base where it occurs.

5. Determine if each entry is an earned out.

6. Create a table that displays the desired information about each out (desired output is the inning, team location, team ID, player name, base at which the player was put out, game ID, and earned out classification.

IV. RESULTS

With the table of output for an entire season's worth of data containing the desired information about each out, we used a python script to analyze the data and return the times the anti-cycle occurred. For each occurrence, the script returns the player who recorded the anti-cycle, the game it was recorded in, and a list of the outs the player recorded. Within the list of outs, for each out the script returns the inning the out occurred in, the base attributed to the out, and if the out was blameless (0 = not blameless, 1 = blameless).

The following table shows the number of anti-cycles recorded between the years of 2008 and 2016. For each year, the table denotes the number of anti-cycles recorded in that year, the number of earned anti-cycles, the number of strong anti-cycles, and the number of ordered anti-cycles.

Year	Occurrences	Earned	Strong	Ordered
2008	24	1	0	1
2009	18	1	0	0
2010	13	0	0	0
2011	17	1	0	0
2012	17	2	0	0
2013	16	3	0	1
2014	25	2	1	0
2015	21	1	5	0
2016	25	1	0	0
Total	176	12	1	2

Table IV.1 Anti-Cycle Occurrences Between 2008 and 2016

From Table IV.1, we can see that 176 anti-cycles were recorded from 2008 to 2016, an average of just under 20 per year. Compared to the average number of times players hit for the cycle each season, which is about 1, the anti-cycle occurs much more frequently. Perhaps a better comparison to hitting for the cycle would be our classification of the earned anti-cycle. From 2008 to 2016, only 12 earned anti-cycles were recorded. This tells us that it is unlikely a player will make 4 distinct outs in a single game where each of those outs is that player's own fault. The average number of earned anti-cycles each season is just

over 1, which is much closer to the average number of times players hit for the cycle in an average season.

V. "INTERESTINGNESS" AND THE ANTI-CYCLE

After conducting our initial analysis and finding the number of anti-cycles that occurred between 2008 and 2016, we considered the question of whether the events of recording an out at each base are independent events. Theoretically, if the events of recording an out at each base were independent of each other, we could multiply the probabilities of recording an out at each base together to get the theoretical probability of an anti-cycle occurring. Then, if the theoretical probability of an anti-cycle occurrence is in line with the actual frequency with which it occurs, we could say that the events of recording outs at each base are independent of each other. If those values aer significantly different, however, then there is likely something "interesting" occurring.

A. "INTERESTINGNESS"

Inderpal Bhandari et al [9] introduced the idea of observing "interesting" occurrences in a basketball game using data mining. For example, suppose a basketball team's average field goal percentage hovers around 50 percent. Yet, when two specific players are on the floor at the same time, one of the players' field goal percentage is over 95 percent. Observing this would tell us there is likely something else going on in the game which is causing that result. This example, which Bhandari discusses in his paper as an actual result of the data mining, was a result of the plays the two players would run when they were in the game together. They would run a play that would ultimately set one player up with a lay-up every time, which is why that player's shooting percentage was so high (in basketball, a lay-up is a shot right under the hoop, considered to be an automatic basket if uncontested by the opposing team).

If the theoretical probability of recording an anti-cycle does not line up with the actual frequencies with which the anti-cycle occurs, there may be an interesting event occurring in baseball causing that to be the case. In order to perform this analysis, we will need some important pieces of information:

- The probability of recording an out at each base,
- the number of player-games for n plate appearances in a game (we define a "player-game" as any occurrence where a unique player has at least one plate appearance in a unique game) for $n = 4, 5, ...,$ the max number of plate appearances observed in a game (which is 10),

- the total number of player-games from 2008-2016,
- the number of recorded anti-cycles from 2008-2016 that occurred for each number of plate appearances (that is, of the recorded anti-cycles, how many occurred with 4 plate appearances, 5 plate appearances, etc.),
- the percentage of player-games that occur with each number of plate appearances for $n = 4, 5, ..., 10$.

B. OBSERVED FREQUENCIES FOR OUTS AT EACH BASE

One of the first goals in determining if the recording outs at each base is independent was to determine the percentage of plate appearances that result in outs at each base. That is, if a player comes to the plate, what is the probability he records an out at home, 1st, etc. The way we captured these probabilities was to observe the frequencies at which they occur throughout all of our data. Because our code was set up to analyze every single plate appearance and then determine when and where outs occur, we were able to determine the percentage of plate appeareances that result in the following: an out at home plate (0), and out at 1st base (1), and out at 2nd base (2), and out at 3rd base (3), and out at home plate after moving around the bases (4), or other (5 - here "other" could be left on base or scores; the code does not differentiate between those results). Table V.1 denoting those probabilities is below.

Base	Percentage
Home Plate (0)	0.4245395
1st Base (1)	0.2220522
2nd Base (2)	0.04155815
3rd Base (3)	0.00358202
Home Plate (4)	0.00355162
Other (6)	0.30471651
Sum	1

Table V.1 Percentages for Out Occurrences at Each Base

The sum of percentages for each base equals 1. Every time a player comes up to the plate, his plate appearance will result in either an out at home, 1st, 2nd, 3rd, home, or other (left on base or scores a run). We use these observed frequencies in our theoretical calcu-

lation of the anti-cycle to determine if anti-cycles occur as often as we would theoretically predict them to.

C. OBSERVED FREQUENCIES FOR PERCENTAGE OF PLAYER-GAMES WITH N PLATE APPEARANCES

We next determine the number of player-games that occur with n plate-appearances. Again, we define a player-game as the combination of a unique player and a unique game. For example, if Derek Jeter comes up to the plate for the first time in a game, he just logged a player-game. Regardless of how many plate appearances he has in this game, it only counts as 1 player-game. The number of plate appearances he has in the game is still important, however. This is because, in our calculation we will go through later, we are mainly interested in player-games where players had at least four plate appearances in that game. This is because a player cannot record an anti-cycle if he does not have at least four plate appearances.

From our code we were able to determine the total number of player-games that occurred between the years of 2008 and 2016 was 414,833. We then dissected that number to create a table of the player-games that occur with each number of plate appearances from n = 1...10 (10 because in all of our data, the largest number of plate appearances any one player had in a single game was 10). Table V.2 shows the number of player-games and their respective percentage frequencies for n = 1...10.

n	Player-Games	Percentage
1	56684	0.136643
2	25380	0.061181
3	47347	0.1141351
4	191939	0.46268905
5	82229	0.198221935
6	9223	0.022233043
7	1620	0.0039052
8	298	0.00071836
9	93	0.0002242
10	20	0.0000482122
Sum	414813	1

Table V.2 Amounts and Percentages for Player-Games with n Plate Appearances

D. THEORETICAL CALCULATION: STARS AND BARS[1]

This section will discuss a method for calculating the theoretical number of anti-cycles that will occur given our observed frequencies just discussed, using a method known as "stars and bars."

1. Basic Stars and Bars

Suppose we have k labeled buckets $B_1, B_2, \ldots, B_k$ and n indistinguishable balls. Basic Stars and Bars gives the number of different ways that the balls can be distributed in the buckets.

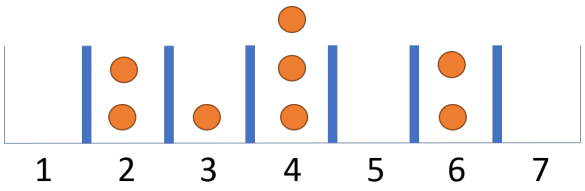

Figure 2. Basic Stars and Bars

In the above figure we have $n = 8$ and $k = 7$.

Two distributions of balls are considered identical if they have the same number of (identical) balls in each of the k buckets.

A basic problem is to compute the number of different ways to put the eight indistinguishable balls into the seven labeled buckets. In Figure 2, the buckets are drawn in a line with vertical dividers (thick blue lines) separating the buckets. These lines are called bars. There are six bars in the figure.

We can represent a distribution by a $0-1$ vector having fourteen elements, one for each bar, denoted by a 1, and one for each ball, denoted by a 0.

The figure above corresponds to the 14 digit sequence

$$1\,0\,0\,1\,0\,1\,0\,0\,0\,1\,1\,0\,0\,1.$$

Recall that the binomial coefficient $\binom{p}{q} = \frac{p!}{(p-q)!q!}$ and equals the number of ways that we can choose q items from a set of p elements.

Here is a basic theorem in the combinatorics of counting:

Theorem (Stars and Bars): The number of different distributions of n identical balls into k distinguishable buckets, denoted by $\delta(n,k)$, equals $\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$.

This is because the number of bars is $k-1$, the number of balls is n and so the total number of objects is $n+k-1$. Each distribution is determined by choosing n of the $n+k-1$ objects for the balls or, equivalently, $k-1$ of the $n+k-1$ objects for the bars. See Figure 1.

So, in the above example, the number of different distributions is $\binom{14}{8} = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{6!}$ =3,003.

A binary sequence such as 0 0 1 0 1 1 0 0 0 1 would represent two balls in bucket 1, one in bucket 2, none in bucket 3 or 5 and three in bucket 4.

2. How many ways can a given distribution be generated?

Here is a different way we can represent the outcome of tossing n balls into k buckets: Create a sequence of n values $(l_1, l_2, \ldots, l_n)$ where $l_i \in \{1, 2, \ldots, k\}$ is the number of the bucket into which the i^{th} ball goes. The example in Figure 1 could be generated by the sequence of outcomes $(2, 3, 4, 6, 2, 4, 6, 4)$ but there are many other sequences that would generate the same distribution.

In general, we may have empty buckets and some buckets containing more than one ball as in the example of Figure 1. We handle this as follows. Initially, we make all numbers in the sequence be distinct. We do this by adding subscripts to any numbers that occur more than once. In this example, three balls went into bucket 4. We use $4_1, 4_2, 4_3$ to represent these three events. A sequence of distinct elements representing the distribution of Figure 1 is $(2_1, 3_1, 4_1, 6_1, 2_2, 4_2, 6_2, 4_3)$. The number of different permutations of this set is $n!$. However this over-counts the number of distributions.

Suppose that there is more than one ball going into bucket j. Each of these events has a different subscript and so corresponds to a different permutation of the sequence with subscripts, but which results in the same distribution. For example, $(2_2, 3_1, 4_2, 6_1, 2_1, 4_3, 6_2, 4_1)$ is a different permutation of the same set of labels which results in the original distribution if the subscripts are removed.

Let $\sigma(D)$ denote the number of different sequences that yield the distribution $D = (d_1, d_2, \ldots . d_k)$.

We are interested in counting distributions, not sequences. We can avoid the over-counting by dividing $n!$ by $d_i!$, for every bucket B_i. This yields the following.

$$\sigma(D) = n!/(d_1! \cdot d_2! \cdots d_k!). \tag{V-1}$$

(Recall that $1! = 0! = 1$.)

For the distribution D of Figure 1, there are

$$\sigma(D) = 8!/(0! \cdot 2! \cdots 1! \cdot 3! \cdot 0! \cdot 2! \cdot 1!) = 8!/(2 \cdot 6 \cdot 2) = 1,680$$

different sequences that result in this distribution.

3. Adding probabilities

Now suppose we toss each of the n balls into one of the k buckets and each ball has probability p_i of going into bucket B_i. Assuming every ball goes into a bucket, that is, no misses, $\sum_{i=1}^{k} p_i = 1$. However, some distributions of balls are more likely than others.

Let $D = (d_1, d_2, \ldots, d_k)$ be a desired distribution of balls. What is the probability that if we randomly toss n balls into the buckets, we will get the distribution D?

We assume that the results of each ball toss are statistically independent. That is, the outcome of each toss is not affected by the result of the toss of any other balls. In this case, the probability of having t balls end up in bucket B_i is p_i^t. More generally, the probability $p(D)$ of getting the distribution D is given by

$$p(D) = \sigma(D) \cdot \prod_{i=1}^{k} p_i^{d_i}. \tag{V-2}$$

This is called the multinomial distribution.

We return to the example of Figure 1. Suppose $p = (0.05, 0.1, 0.15, 0.2, 0.3, 0.15, 0.05)$ is the vector of probabilities of a tossed ball going into each bucket. The probability of obtaining the distribution $(0, 2, 1, 3, 0, 2, 0)$ is

$$1,680 \cdot 0.1^2 \cdot 0.15 \cdot 0.2^3 \cdot 0.15^2 = 0.0004536.$$

Table V.3 below gives a simpler example and analyzes the 15 possible different distributions of four balls in three buckets as well as the probability $P(D)$ of each distribution D occurring if the probabilities of balls landing in each bucket are $0.5, 0.3$ and 0.2. The sixth column is the product of the values in the fourth and fifth columns. Note that the probabilities do sum to one.

The sum of the number of sequences for all distributions equals $81 = 3^4$ – each of four balls can go into any of three buckets.

Bucket 1 $p_1 = 0.5$	Bucket 2 $p_2 = 0.3$	Bucket3 $p_3 = 0.2$	No. of Sequences giving distribution D: $\sigma(D) = 4!/(d_1!d_2!d_3!)$	$\prod_{i=1}^{3} p_i^{d_i} = p_1^{d_1} p_2^{d_2} p_3^{d_3}$	$p(D) = \sigma(D) \cdot \prod_{i=1}^{3} p_i^{d_i}$
4	0	0	1	$.5^4 = .0625$	.0625
0	4	0	1	$.3^4 = .0081$	.0081
0	0	4	1	$.2^4 = .0016$	.0016
3	0	1	4	$.5^3 \cdot .2 = .0250$	.1000
3	1	0	4	$.5^3 \cdot .3 = .0375$	.1500
0	3	1	4	$.3^3 \cdot .2 = .0054$	.0216
1	3	0	4	$.5 \cdot .3^3 = .0135$	.0540
0	1	3	4	$.3 \cdot .2^3 = .0024$	.0096
1	0	3	4	$.5 \cdot .2^3 = .0040$	.0160
2	2	0	6	$.5^2 \cdot .3^2 = .0225$	.1350
2	0	2	6	$.5^2 \cdot .2^2 = .0100$	.0600
0	2	2	6	$.3^2 \cdot .2^2 = .0036$	.0216
2	1	1	12	$.5^2 \cdot .3 \cdot .2 = .0150$	.1800
1	2	1	12	$.5 \cdot .3^2 \cdot .2 = .0090$	.1080
1	1	2	12	$.5 \cdot .3 \cdot .2^2 = .0060$	.0720
			Total: 81		Total : 1.000

Table V.3 Distributions of four balls in three buckets with probabilities

4. Requiring some buckets be nonempty

Suppose we add the requirement that there be at least one ball in some or all of the buckets. How many ways can this be done? We start with the simpler case of distributions of four balls in three buckets, with the added requirement that every bucket be nonempty.

Table V.3 shows that there are only three distributions of four balls in three buckets that result in every bucket being nonempty, the last three in the table. The probability of there being no empty bucket is 0.3600, the sum of the last three entries in the $p(D)$ column.

Now, suppose we have seven balls instead of four to toss into the three buckets with probabilities $0.5, 0.3, 0.2$ which we analyzed in the previous section. What would be the probability of getting a distribution with at least one ball in every bucket? We can compute this as follows.

Begin by "pre-loading" one ball in every bucket. Now we are left with four balls to randomly distribute. Table V.3 gives all the possible distributions of the four balls. If we add 1 to all the entries in the three columns, we get all possible distributions of seven identical balls in three distinguishable buckets, subject to the restriction that there must be at least one ball in each bucket.

The number of different such distributions equals the number of different distributions of the remaining four balls in three buckets, which we computed in Table V.3. However, the number of sequences of seven balls giving each distribution and the resulting probabilities of these distributions does change.

The number of different sequences giving the distribution $D = (d_1{+}1, d_2{+}1, d_3{+}1)$ is given by

$$\delta(D) = (7!/((d_1{+}1)!(d_2{+}1)!(d_3{+}1)! = [(7{\cdot}6{\cdot}5)/((d_1{+}1){\cdot}(d_2{+}1){\cdot}(d_3{+}1))]{\cdot}4!/(d_1!d_2!d_3!).$$

So we can update Column 4 of Table V.3 by multiplying each entry (d_1, d_2, d_3) by $(7 \cdot 6 \cdot 5)/((d_1 + 1) \cdot (d_2 + 1) \cdot (d_3 + 1))$. We can update Column 5 of Table V.3 by multiplying each entry (d_1, d_2, d_3) by $p_1p_2p_3 = 0.03$.

So the probability of having all three buckets nonempty when seven identical balls are distributed is 0.7082. Applying the same calculations to the distributions which leave 1 or 2 buckets empty, we get their probability equal to 0.2918, reassuring since these sums to add to 1.

Bucket 1 $p_1 = 0.5$	Bucket 2 $p_2 = 0.3$	Bucket 3 $p_3 = 0.2$	Number $\sigma(D)$ of sequences giving D, all buckets nonempty $7!/((d_1+1)!(d_2+1)!(d_3+1)!)$	$\prod_{i=1}^{3} p_i^{d_i+1} = 0.03 \cdot \prod_{i=1}^{3} p_i^{d_i}$	$p(D)$
5	1	1	42	$0.03 \cdot 0.0625 = 0.001875$	.0788
1	5	1	42	$0.03 \cdot 0.0081 = 0.000243$	.0102
1	1	5	42	$0.03 \cdot 0.0016 = 0.000480$	.0020
4	1	2	105	$0.03 \cdot 0.0250 = 0.007500$	.0788
4	2	1	105	$0.03 \cdot 0.0375 = 0.001125$	.1181
1	4	2	105	$0.03 \cdot 0.0054 = 0.000162$	.0170
2	4	1	105	$0.03 \cdot 0.0135 = 0.004050$	.0425
1	2	4	105	$0.03 \cdot 0.0024 = 0.000720$	.0076
2	1	4	105	$0.03 \cdot 0.0040 = 0.000120$	.0126
3	3	1	140	$0.03 \cdot 0.0225 = 0.006750$	.0945
3	1	3	140	$0.03 \cdot 0.0100 = 0.003000$	.0420
1	3	3	140	$0.03 \cdot 0.0036 = 0.000108$	.0151
3	2	2	210	$0.03 \cdot 0.0150 = 0.004500$	.0945
2	3	2	210	$0.03 \cdot 0.0090 = 0.000270$	.0567
2	2	3	210	$0.03 \cdot 0.0060 = 0.000180$	.0378
				Total	0.7082

Table V.4 Distributions of seven balls in three buckets, each containing at least one ball, with probabilities

E. THEORETICAL SIMULATION BASED ESTIMATES

Based on the idea of stars and bars, we created a simulation that will use that process, along with the observed frequencies for outs recorded at each base as well as player-games with n plate appearances, to calculate the theoretical probability of an anti-cycle occurring. The simulation takes the following steps:

1. Assign for the number of plate appearances (n), the number of possible outcomes of each plate appearance (6), and the number of possible permutations for plate appearances with each result.
2. Create a data-frame with those permutations.
3. Assign the observed frequencies for outs at each base to a vector.
4. Find the probabilities of each entry in the data-frame and check they sum to 1.
5. Check to see if each entry has an anti-cycle occurrence.
6. Assign the observed frequencies for player-games with n plate appearances for 1 to 10 to a vector.

7. Create a data-frame with the number of plate appearances, theoretical probability an anti-cycle occurring given n plate appearances, and probability of both n plate appearances and an anti-cycle occurring.

Table V.5 below denotes the results of the simulation after iterating through 10 times.

# of Plate Appearances	Prob. of A-C Occurrence	Prob. of A-C With n Plate Appearances
1	0.00	0.00
2	0.00	0.00
3	0.00	0.00
4	0.00033961	0.00015713
5	0.00110775	0.00021958
6	0.00226742	0.00005041
7	0.00375421	0.00001466
8	0.00551988	0.00000395
9	0.00749318	0.00000167
10	0.00966762	0.00000046

Table V.5 Results of Theoretical Simulation

F. GENERAL SIMULATION

In addition to the theoretical simulation described, we also created a different simulation which attempted to "recreate" the 9 season's worth of data we analyzed. While the theoretical simulation focused on using the "stars and bars" approach, this simulation used the same observed frequencies from our data but instead focused more on recreating player-games based on our observed frequencies and using those player-games to look for anti-cycles in the same way our original code does. In other words, this general simulation created player-games based on the number of player-games we observed from 2008-2016, while the theoretical simulation iterated through all the possibilities of balls and buckets. The process of the general simulation is seen below.

1. Assign the observed frequencies for outs at each base to a vector.
2. Assign the observed frequencies for player-games with n plate appearances for 1 to 10.

3. Create a matrix that is 414833x10 (414833 is the total number of player-games observed; 10 is the most amount of plate appearances by a single player in a single game observed). Therefore, each row represents a player-game and each column represents a plate appearance.

4. Assign each entry as a plate-appearance or not (each entry in the matrix represents a potential plate-appearance - that is, each entry either does or does not represent a plate appearance, based on the observed frequencies for number of plate appearances in player-games).

5. For each assigned plate appearance, assign a result of that plate appearance based on observed frequencies for outs recorded (that is, each plate appearance will have a value of 0-5).

6. Once every player-game is assigned plate appearances and those plate appearances are assigned outcomes, search through the matrix to determine which player-games display an anti-cycle occurrence.

7. Repeat this process using a for-loop 10 times, and finds the average number of anti-cycles that occur.

This simulation recreated nine season's worth of data by creating 414833 player-games and randomly assigning them plate appearances based on probabilities with outcomes. It then searched through the created data to determine when anti-cycles occur. It repeated the process 10 times. In total, the observed number of anti-cycles from 2008-2016 was 176. Using this simulation, the average number of observed anti-cycles during over all 10 iterations was 186.2! This tells us that our simulation aligns closely with the actual observed number of anti-cycles.

G. CHI-SQUARE INDEPENDENCE TEST

Based on the observed number of anti-cycles from 2008 to 2016, and the results of both our theoretical and general simulations, we can determine whether the events of recording outs at each base are statistically independent and apply the concept of interestingness.

A chi-square test determines whether two events are statistically independent of each other.[10] The null hypothesis of the chi-square test is that Variable A and Variable B are independent. The alternate hypothesis then is that Variable A and Variable B are not independent. We will use the chi-square test to see whether the events of recording an out

at each base are independent events. That is, does recording an out at one base affect the probability of recording an out at another base?

1. Theoretical Simulation Results

Using the built-in chi-squared function in R, we tested both the simulations to the observed frequencies of anti-cycles to see if they are independent. For the theoretical simulation, we tested the anti-cycle occurrences for all values of n from 1 to 10 where n is the number of player-games with n plate appearances. After running the chi-square test with all values of n, we get a resulting p-value of 0.06. Based on an alpha level of 0.05, which represents a confidence level of 95%, we would declare that the data are too different to declare the events of recording an out at each base are independent. Next, we run the chi-square test only with values of n from 4 to 10, because an anti-cycle cannot occur with less than four plate appearances in a game. When running the chi-square test with only 4 through 10 as the values for n, our resulting p-value is 0.26, much larger than before. In this instance, we would definitely reject our null hypothesis and say our data are not independent.

2. General Simulation Results

Next we used the chi-square test for the general simulation to see if our results are similar. For the test with all values of n, we get a resulting p-value of 0.066 once again. Then, for the test with only value 4 to 10 as the values for n, we get a resulting p-value of 0.16. Therefore, after both of these tests, we would conclude again to reject the null hypothesis meaning the events are not independent.

3. Interestingness Conclusion

Because the events of recording an out at each base in a game are not independent of each other, we then conclude that there is something "interesting" going on in the data. Future work might involve further analysis to see what exactly that "something" might be.

VI. APPENDIX

The figure below shows the results of running our code that searches for anti-cycle occurrences. For each anti-cycle, the output consists of the game ID for the game in which the anti-cycle occurred, the player who recorded the anti-cycle, and a list of the plate-appearances which resulted in outs, each with three pieces of information. The first is the inning in which the out was recorded. The second is the base at which the out was recorded. Finally, the third is whether the out was earned or unearned; earned is denoted by a "0" and unearned is denoted by a "1."

```
gid_2008_04_06_chamlb_detmlb_1, Ramon Santiago, [(3, 3, 0), (5, 2, 1), (7, 0, 0), (9, 1, 0)]
gid_2008_04_20_pitmlb_chnmlb_1, Nate McLoth, [(1, 3, 1), (3, 0, 0), (7, 2, 1), (8, 1, 0)]
gid_2008_04_29_detmlb_nyamlb_1, Johnny Damon, [(1, 0, 0), (2, 2, 1), (4, 1, 0), (6, 0, 0), (8, 3, 1)]
gid_2008_05_06_sfnmlb_pitmlb_1, Emmanuel Burriss, [(1, 1, 0), (3, 2, 1), (5, 1, 0), (8, 3, 1), (9, 0, 0)]
gid_2008_05_09_chamlb_seamlb_1, Jermaine Dye, [(2, 3, 1), (3, 0, 0), (5, 1, 0), (8, 2, 1)]
gid_2008_05_12_tormlb_clemlb_1, Scott Rolen, [(2, 2, 1), (4, 3, 0), (6, 1, 0), (9, 0, 0)]
gid_2008_05_14_nyamlb_tbamlb_1, Cliff Floyd, [(2, 3, 0), (5, 0, 0), (7, 2, 1), (9, 1, 0)]
gid_2008_05_25_anamlb_chamlb_1, Orlando Cabrera, [(1, 1, 0), (3, 0, 0), (5, 3, 0), (8, 2, 1)]
gid_2008_05_29_colmlb_chnmlb_1, Seth Smith, [(1, 2, 1), (3, 3, 1), (5, 1, 0), (6, 0, 0), (8, 0, 0)]
gid_2008_05_29_tormlb_oakmlb_1, Jack Cust, [(1, 0, 0), (4, 2, 0), (6, 1, 0), (8, 3, 1)]
gid_2008_06_15_kcamlb_arimlb_1, John Buck, [(2, 0, 0), (6, 3, 0), (7, 1, 0), (9, 2, 1)]
gid_2008_06_18_wasmlb_minmlb_1, Aaron Boone, [(2, 1, 0), (4, 0, 0), (6, 2, 1), (8, 3, 1)]
gid_2008_06_21_flomlb_oakmlb_1, Matt Treanor, [(2, 3, 1), (4, 2, 1), (6, 1, 0), (8, 0, 0)]
gid_2008_06_22_slnmlb_bosmlb_1, Dustin Pedroia, [(1, 2, 1), (4, 0, 0), (10, 1, 0), (12, 3, 0)]
gid_2008_06_25_nyamlb_pitmlb_1, Robinson Cano, [(2, 0, 0), (3, 1, 0), (7, 2, 1), (9, 3, 0)]
gid_2008_06_28_atlmlb_tormlb_1, Vernon Wells, [(2, 1, 0), (4, 0, 0), (7, 2, 1), (8, 3, 0)]
gid_2008_07_09_minmlb_bosmlb_1, Nick Punto, [(2, 2, 1), (4, 0, 0), (6, 3, 0), (7, 1, 0), (9, 0, 0)]
gid_2008_07_13_slnmlb_pitmlb_1, Xavier Nady, [(1, 3, 1), (4, 1, 0), (5, 2, 1), (7, 1, 1), (9, 0, 0)]
gid_2008_07_27_chamlb_detmlb_1, Edgar Renteria, [(2, 0, 0), (4, 2, 1), (6, 3, 0), (8, 1, 0)]
gid_2008_08_08_minmlb_kcamlb_1, Justin Morneau, [(2, 3, 1), (3, 2, 0), (5, 0, 0), (7, 1, 0), (9, 2, 1)]
gid_2008_08_17_seamlb_minmlb_1, Nick Punto, [(2, 1, 0), (4, 3, 0), (5, 2, 1), (8, 0, 0)]
gid_2008_08_18_sfnmlb_atlmlb_1, Pablo Sandoval, [(1, 2, 1), (4, 0, 0), (7, 3, 0), (9, 1, 0)]
gid_2008_09_12_lanmlb_colmlb_1, Russell Martin, [(1, 1, 0), (3, 2, 1), (5, 3, 1), (7, 0, 0), (9, 0, 0)]
gid_2008_09_19_minmlb_tbamlb_1, Jason Bartlett, [(1, 1, 0), (2, 3, 0), (4, 0, 0), (6, 2, 0)]
```

Figure 3. Python Output of Anti-Cycles Recorded in 2008

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] William Pulleyblank. Stars and bars. Class Notes - United States Military Academy at West Point, April 2019. I used this in my thesis to explain how we would calculate our probabilities theoretically. The material in this section of the thesis is extracted directly from Dr. William Pulleyblank's work, "Stars and Bars.".

[2] Jon Marthaler. Discussion topic: What is the anti-cycle? Twinkie Town, 2009.

[3] Major League Baseball. Official mlb rulebook, 2018.

[4] Baseball Fever: A Baseball Community. How many ways can a batter get out?, December 2008.

[5] Brooks Baseball. Pitchf/x. Brooks Baseball provides a means of accessing PITCHf/x data one game at a time, but I simply used this site for a description of the database, not the data itself.

[6] Carson Sievert. Starting and updating a pitchf/x database with pitchrx and dplyr. Exploring Baseball Data with R, 2014.

[7] Josh Errickson. Sql in r. This thread is a collapsed version of information about using SQL in R from w3schools.com which has a more comprehensive description of its various uses and applications.

[8] Hadley Wickam. Tidyverse.

[9] Inderpal Bhandari. Advanced scout: Data mining and knowledge discovery in nba data. Kluwer Academic Publishers, 1997.

[10] Stat Trek: Teach Yourself Statistics. Chi-square test for independence, 2019.