Project 36: MLB

ECE 356

Group 36

Matthew Lee, kh37lee

Jeffrey Yuen, jwy2yuen

Martin Hsu, y5hsu

Table Of Contents

Project Description and Scope **3**

Project Description 3

Project Scope 3

Client Application **4**

Client Experience 4

Viewing Data 5

Game Data 5

Team Data 8

Player Data 9

Modifying Data 12

Updating Data 13

Inserting Data 14

Deleting Data 15

Work Plan and Responsibilities 16

Project Assumptions 18

ER design **18**

Relational Schema **22**

Test Plan **24**

Client Application Testing 24

Testing Viewing Data 25

Testing Altering Data 31

SQL Testing 38

Project Description and Scope

Project Description

The project that we are working on will be using MLB data. As we are using a dataset involving sports statistics, we will be making a client for users to easily look up stats and information through a command-line client application. The command-line interface will act as a menu for the user to navigate through to choose what specific type of information they want to view, whether it be the plays that took place during a specific game, or the stats on a specific player, etc. The user can also modify some data that won't impact a game's outcome including a player's name, ejection data, or the starting time of a game. This is to make our application as simple as possible without making the queries overly complicated. As for game statistics, much of the data is linked together, specifically between game scores and the play-by-play events that occur during a game. Altering these data sets although may not break our database, it will make the data illogical without taking a series of complicated steps to update multiple tables and records based on a single change on a specific table.

Project Scope

The scope of our project will mainly be focusing on the MLB data from 2015 to 2018. Initially, we were given two data sets: play by play from 2010 to 2018 and pitch data from 2015 to 2019. The play-by-play data has nearly 90 columns in a CSV file, and it will be a lot of work to break down these fields since they may need to be normalized. Furthermore, the data lack other basic information that users may want to know such as the players and teams specifically whereas the pitch data from 2015 to 2019 is more organized and normalized as it has data for common information including players, teams, ejections, etc. The first dataset

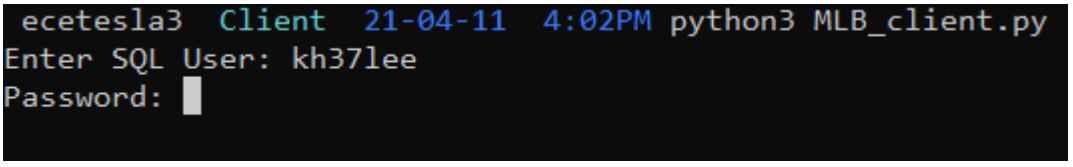
will not be used, and we will only focus on the latter dataset. However, much of the 2019 information is missing, so we have decided to remove that year from our dataset so our information is complete and consistent when users are looking up data. The link to the data set is as follows: https://www.kaggle.com/pschale/mlb-pitch-data-20152018.

Client Application

Client Experience

As mentioned in the project description, the application will focus on three main aspects: game data, team data, and player data. As our dataset does not have much information on other aspects such as trades, salary, fielding, we will not include these statistics in our client application.

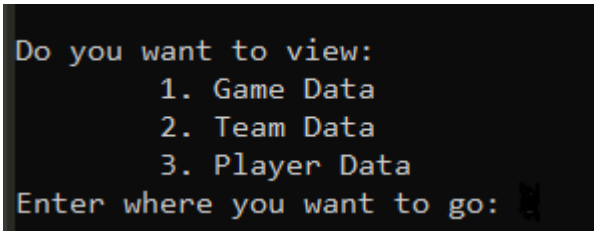
Our application is written in Python as our group found it easiest to create the application. It is a powerful high-level language and is commonly used for data analysis and fast application development. It also includes back-end development which we will use to connect to the marmoset database to query and update our tables. In addition to the default libraries in Python, we have used the '*mysql.connector'* library to connect to our database and perform SQL queries, the '*prettytable'* library to easily format and display information using their built in table functions, '*datetime'*, and '*math'* to format our data, and '*getpass'* to allow the user to enter their login credentials to connect to the marmoset database.



When the application starts, the program will ask the user to enter their login credentials to connect to the database. We have a prompt menu for the user to determine

where they want to navigate to. They have 2 options: they can either view data or modify it. Through all menus, we also allow the user to navigate to the previous page by entering *'back'* and go back to this home page by entering *'home'*.

Viewing Data



When the user chooses to view data by entering *'1'*, they can then choose which area they want to view between 3 options: Game, Player, and Team data.

Game Data

If the user selects game data, they will have to select a game, by following the format and they will be offered information to work with as they will be able to view. If they are not sure which game they want to view, they can filter the results by entering *'year team'* to enter a sub search to display games from a specific year for a specific team. For example, if I entered *'2015, Blue Jays'*, after entering *'year team'*, I would see a list of games that the Blue Jays played in 2015. This helps if you are not sure exactly what game you want to look for, or if you do not remember the exact date of a game, but you know the team that played and the year. The displayed output will look like something below.

```
Enter the number of the games you want to see from the desired team for a given year (Ex: 2015, Blue Jays): 2018, Blue Jays
+---------------------------------+---------------------------------+---------------------------------+
|                                 |                                 |                                 |
+---------------------------------+---------------------------------+---------------------------------+
| Blue Jays, Yankees, 2018-03-29  | Mets, Blue Jays, 2018-05-15     | Blue Jays, Tigers, 2018-07-01   |
| Blue Jays, Yankees, 2018-03-30  | Mets, Blue Jays, 2018-05-16     | Blue Jays, Tigers, 2018-07-02   |
| Blue Jays, Yankees, 2018-03-31  | Blue Jays, Athletics, 2018-05-17| Blue Jays, Mets, 2018-07-03     |
| Blue Jays, Yankees, 2018-04-01  | Blue Jays, Athletics, 2018-05-18| Blue Jays, Mets, 2018-07-04     |
| Blue Jays, White Sox, 2018-04-02| Blue Jays, Athletics, 2018-05-19| Blue Jays, Yankees, 2018-07-06  |
| Blue Jays, White Sox, 2018-04-03| Blue Jays, Athletics, 2018-05-20| Blue Jays, Yankees, 2018-07-07  |
| Blue Jays, White Sox, 2018-04-04| Blue Jays, Angels, 2018-05-22   | Blue Jays, Yankees, 2018-07-08  |
| Rangers, Blue Jays, 2018-04-06  | Blue Jays, Angels, 2018-05-23   | Braves, Blue Jays, 2018-07-10   |
| Rangers, Blue Jays, 2018-04-07  | Blue Jays, Angels, 2018-05-24   | Braves, Blue Jays, 2018-07-11   |
| Rangers, Blue Jays, 2018-04-08  | Phillies, Blue Jays, 2018-05-25 | Red Sox, Blue Jays, 2018-07-12  |
| Orioles, Blue Jays, 2018-04-09  | Phillies, Blue Jays, 2018-05-26 | Red Sox, Blue Jays, 2018-07-13  |
| Orioles, Blue Jays, 2018-04-10  | Phillies, Blue Jays, 2018-05-27 | Red Sox, Blue Jays, 2018-07-14  |
| Orioles, Blue Jays, 2018-04-11  | Red Sox, Blue Jays, 2018-05-28  | Red Sox, Blue Jays, 2018-07-15  |
| Indians, Blue Jays, 2018-04-13  | Red Sox, Blue Jays, 2018-05-29  | Blue Jays, Orioles, 2018-07-20  |
| Blue Jays, Royals, 2018-04-17   | Red Sox, Blue Jays, 2018-05-30  | Blue Jays, Orioles, 2018-07-21  |
| Blue Jays, Royals, 2018-04-17   | Tigers, Blue Jays, 2018-06-01   | Blue Jays, Orioles, 2018-07-22  |
| Blue Jays, Royals, 2018-04-18   | Tigers, Blue Jays, 2018-06-02   | Blue Jays, Twins, 2018-07-23    |
| Yankees, Blue Jays, 2018-04-19  | Tigers, Blue Jays, 2018-06-03   | Blue Jays, Twins, 2018-07-24    |
| Yankees, Blue Jays, 2018-04-20  | Blue Jays, Yankees, 2018-06-05  | Blue Jays, Twins, 2018-07-25    |
| Yankees, Blue Jays, 2018-04-21  | Blue Jays, Yankees, 2018-06-06  | White Sox, Blue Jays, 2018-07-27|
| Yankees, Blue Jays, 2018-04-22  | Blue Jays, Orioles, 2018-06-07  | White Sox, Blue Jays, 2018-07-28|
| Blue Jays, Red Sox, 2018-04-24  | Blue Jays, Orioles, 2018-06-08  | White Sox, Blue Jays, 2018-07-29|
| Blue Jays, Red Sox, 2018-04-25  | Blue Jays, Orioles, 2018-06-09  | Athletics, Blue Jays, 2018-07-30|
| Blue Jays, Red Sox, 2018-04-26  | Blue Jays, Orioles, 2018-06-10  | Athletics, Blue Jays, 2018-07-31|
| Blue Jays, Rangers, 2018-04-27  | Rays, Blue Jays, 2018-06-11     | Athletics, Blue Jays, 2018-08-01|
| Blue Jays, Rangers, 2018-04-28  | Rays, Blue Jays, 2018-06-12     | Mariners, Blue Jays, 2018-08-02 |
| Blue Jays, Rangers, 2018-04-29  | Rays, Blue Jays, 2018-06-13     | Mariners, Blue Jays, 2018-08-03 |
| Twins, Blue Jays, 2018-04-30    | Blue Jays, Nationals, 2018-06-15| Mariners, Blue Jays, 2018-08-04 |
| Twins, Blue Jays, 2018-05-01    | Blue Jays, Nationals, 2018-06-16| Mariners, Blue Jays, 2018-08-05 |
| Twins, Blue Jays, 2018-05-02    | Blue Jays, Nationals, 2018-06-17| Blue Jays, Red Sox, 2018-08-07  |
| Indians, Blue Jays, 2018-05-03  | Blue Jays, Braves, 2018-06-19   | Blue Jays, Red Sox, 2018-08-08  |
| Indians, Blue Jays, 2018-05-03  | Blue Jays, Braves, 2018-06-20   | Blue Jays, Red Sox, 2018-08-09  |
| Rays, Blue Jays, 2018-05-04     | Angels, Blue Jays, 2018-06-21   | Blue Jays, Rays, 2018-08-10     |
| Rays, Blue Jays, 2018-05-05     | Angels, Blue Jays, 2018-06-22   | Blue Jays, Rays, 2018-08-11     |
| Rays, Blue Jays, 2018-05-06     | Angels, Blue Jays, 2018-06-23   | Blue Jays, Rays, 2018-08-12     |
| Blue Jays, Mariners, 2018-05-08 | Angels, Blue Jays, 2018-06-24   | Royals, Blue Jays, 2018-08-13   |
| Blue Jays, Mariners, 2018-05-09 | Astros, Blue Jays, 2018-06-25   | Royals, Blue Jays, 2018-08-14   |
| Blue Jays, Mariners, 2018-05-10 | Astros, Blue Jays, 2018-06-26   | Royals, Blue Jays, 2018-08-15   |
| Blue Jays, Red Sox, 2018-05-11  | Astros, Blue Jays, 2018-06-27   | Royals, Blue Jays, 2018-08-16   |
| Blue Jays, Red Sox, 2018-05-12  | Blue Jays, Tigers, 2018-06-29   | Yankees, Blue Jays, 2018-08-17  |
| Blue Jays, Red Sox, 2018-05-13  | Blue Jays, Tigers, 2018-06-30   | Yankees, Blue Jays, 2018-08-18  |
| Mets, Blue Jays, 2018-05-15     | Blue Jays, Tigers, 2018-07-01   | Yankees, Blue Jays, 2018-08-19  |
+---------------------------------+---------------------------------+---------------------------------+
```

Once a game is correctly inputted, we will display the basic Game data that we grabbed from the Games table and then we let the user see the batting and pitching stats for the home or away team. Below is an example of the Chicago Cubs Batting and Pitching Stats on their game against the St. Louis Cardinals on April 5th, 2015.

```
+-------------------------------------------------+
| St. Louis Cardinals at Chicago Cubs on 2015-04-05  |
+----------------+--------------------------------+
| Label          | Value                          |
+----------------+--------------------------------+
| Final Score    | (3) Cubs vs. (0) Cardinals     |
| Venue          | Wrigley Field                  |
| Attendance     | 35055                          |
| Start Time     | 19:17:00                       |
| Delay          | 0 mins                         |
| Game Length    | 184 mins                       |
| Weather        | 44 degrees                     |
| Wind           | 7 mph, In from CF              |
+----------------+--------------------------------+

1. (Home) Cubs
2. (Away) Cardinals
Enter the number of the team you want to view the Batter and Pitcher Stats for:1

+-----------------------------------------------------------------------------------+
|                          Chicago Cubs Batting Game Stats                          |
+--------------------+-----+-----+-----+-----+-----+-------+-------+-------+-------+
| Batting            | AB  | H   | BB  | RBI | SO  | BA    | OBP   | SLG   | OPS   |
+--------------------+-----+-----+-----+-----+-----+-------+-------+-------+-------+
| Dexter Fowler      | 4   | 1   | 0   | 0   | 2   | 0.250 | 0.250 | 0.500 | 0.750 |
| Jorge Soler        | 4   | 0   | 0   | 0   | 1   | 0.000 | 0.000 | 0.000 | 0.000 |
| Anthony Rizzo      | 3   | 0   | 1   | 0   | 2   | 0.000 | 0.250 | 0.000 | 0.250 |
| Starlin Castro     | 4   | 1   | 0   | 0   | 1   | 0.250 | 0.250 | 0.250 | 0.500 |
| Chris Coghlan      | 4   | 1   | 0   | 0   | 1   | 0.250 | 0.250 | 0.500 | 0.750 |
| Mike Olt           | 4   | 0   | 0   | 0   | 2   | 0.000 | 0.000 | 0.000 | 0.000 |
| David Ross         | 2   | 1   | 0   | 0   | 0   | 0.500 | 0.500 | 1.000 | 1.500 |
| Jon Lester         | 1   | 0   | 0   | 0   | 1   | 0.000 | 0.000 | 0.000 | 0.000 |
| Tommy La Stella    | 3   | 1   | 0   | 0   | 0   | 0.333 | 0.333 | 0.333 | 0.666 |
| Miguel Montero     | 2   | 0   | 0   | 0   | 1   | 0.000 | 0.000 | 0.000 | 0.000 |
| Arismendy Alcantara| 1   | 0   | 1   | 0   | 1   | 0.000 | 0.500 | 0.000 | 0.500 |
| ---------------    | --- | --- | --- | --- | --- | ----- | ----- | ----- | ----- |
| Team Totals        | 32  | 5   | 2   | 0   | 12  | 0.144 | 0.212 | 0.235 | 0.447 |
+--------------------+-----+-----+-----+-----+-----+-------+-------+-------+-------+

+-----------------------------------------------------------+
|              Chicago Cubs Pitching Game Stats              |
+-----------------+-----+-----+-----+-----+-----+-----+-----+
| Pitching        | IP  | H   | ER  | BB  | SO  | HR  | ERA |
+-----------------+-----+-----+-----+-----+-----+-----+-----+
| Jon Lester      | 4.1 | 8   | 3   | 2   | 6   | 0   | 6.59 |
| Phil Coke       | 0.2 | 0   | 0   | 1   | 2   | 0   | 0.00 |
| Jason Motte     | 1.0 | 0   | 0   | 0   | 0   | 0   | 0.00 |
| Neil Ramirez    | 1.0 | 1   | 0   | 0   | 1   | 0   | 0.00 |
| Pedro Strop     | 1.0 | 1   | 0   | 0   | 2   | 0   | 0.00 |
| Hector Rondon   | 1.0 | 0   | 0   | 1   | 0   | 0   | 0.00 |
| --------------- | --- | --- | --- | --- | --- | --- | ---- |
| Team Totals     | 9   | 10  | 3   | 4   | 11  | 0   | 0.39 |
+-----------------+-----+-----+-----+-----+-----+-----+-----+
```

Initially, most of the data that will be displayed on the application will be computed statistics that are collected through joins between the *Games*, *AtBats*, and *Pitches* Tables. These can include results in displaying global stats like total hits from each team, total runs, as well as individual stats for each batter and pitcher that played in the specific game. We can

take all this data and compute interesting information that will be presented to the user and allow them to see a player's performance for that game.

Team Data

For the team data, the user can select a team and view the team's performance throughout the season including its records and average allowed runs per game, etc. Firstly, they will have to enter the name of the team that they want to view. The format of the input is just the team name, not including the city. An example of this would be *'Blue Jays'*. If they want to see all of the active teams in the MLB, they can enter *'show teams'*. This command will display all the teams in the MLB.

```
Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for: show teams
Below is all MLB Teams
Angels              White Sox           Tigers              Brewers             Phillies            Cardinals
Diamondbacks        Cubs                Astros              Twins               Pirates             Rays
Braves              Reds                Royals              Yankees             Padres              Rangers
Orioles             Indians             Dodgers             Mets                Mariners            Maple Leafs
Red Sox             Rockies             Marlins             Athletics           Giants              Nationals
```

They can enter a team name, and it will be considered valid if it exists as a key in our *team_abbrev* dictionary that uses the team name as the key and the value will be the abbreviation of that team. We use that abbreviation in our query to grab the data. Here, we mainly use the *COUNT* and *SUM* aggregate functions in SQL to compute most of the stats that we will display for that team and we group the results by year so that we can display each row as a year from 2015 - 2018. We run a separate query for the ejections as that is grabbed from the *Ejections* table rather than the *Games* table. The ejections for the team over the years are also shown and provided to the user. The table shows all records of players of the selected team that have been ejected and some information regarding that ejection including the inning they were ejected, the game, the player, and what that ejection was

regarding. Below is an example of the Team Data for the Toronto Blue Jays, which can appear when the user inputs *'Blue Jays'*.

```
Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for: Blue Jays

+-----------------------------------------------------------------------------------------------------------+
|                                      Toronto Blue Jays 2015-2018                                          |
+--------+--------+------------+--------------+--------------------+----------------------+-----------------+
| Season | Record | Total Runs | Avg Runs/Game | Total Runs Allowed | Avg Runs Allowed/Game | Total Attendance |
+--------+--------+------------+--------------+--------------------+----------------------+-----------------+
| 2015   | 94-69  | 891        | 9.48          | 736                | 7.83                 | 2794891         |
| 2016   | 89-73  | 758        | 8.52          | 721                | 8.1                  | 3392099         |
| 2017   | 76-86  | 693        | 9.12          | 747                | 9.83                 | 3203886         |
| 2018   | 73-89  | 709        | 9.71          | 741                | 10.15                | 2325282         |
+--------+--------+------------+--------------+--------------------+----------------------+-----------------+

+-----------------------------------------------------------------------------------------------------+
|                                    Team Ejections 2015-2018                                         |
+-------------------------------------------------+--------+------------------+----------+---------+
| Game                                            | Inning | Player           | Argue BS | Correct |
+-------------------------------------------------+--------+------------------+----------+---------+
| Baltimore Orioles at Toronto Blue Jays, 2015-06-19  | 8      | Roberto Osuna    | No       | N/A     |
| Toronto Blue Jays at Kansas City Royals, 2015-07-12 | 8      | Jose Bautista    | No       | N/A     |
| Kansas City Royals at Toronto Blue Jays, 2015-08-02 | 8      | Aaron Sanchez    | No       | N/A     |
| Baltimore Orioles at Toronto Blue Jays, 2015-09-05  | 7      | Cliff Pennington | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jose Bautista    | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Josh Donaldson   | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jesse Chavez     | No       | N/A     |
| Toronto Blue Jays at Minnesota Twins, 2016-05-21    | 1      | Josh Donaldson   | Yes      | Yes     |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 2      | Edwin Encarnacion | Yes     | No      |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 14     | Russell Martin   | Yes      | Yes     |
| Toronto Blue Jays at Seattle Mariners, 2016-09-19   | 7      | Josh Donaldson   | Yes      | No      |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Russell Martin   | Yes      | Yes     |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Marcus Stroman   | Yes      | Yes     |
| Toronto Blue Jays at Tampa Bay Rays, 2017-08-22     | 3      | Kevin Pillar     | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-05     | 18     | Josh Donaldson   | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-25     | 8      | Jose Bautista    | Yes      | Yes     |
| Boston Red Sox at Toronto Blue Jays, 2018-08-07     | 8      | Ryan Tepera      | Yes      | N/A     |
| Toronto Blue Jays at New York Yankees, 2018-09-15   | 2      | Luke Maile       | Yes      | Yes     |
+-------------------------------------------------+--------+------------------+----------+---------+
```

Player Data

The functionality for the player data will follow a similar route to the game data where they will be able to view an individual player's performance throughout the several seasons of games that exist in the database. If the player is a pitcher or a batter, the information displayed will be different, but ultimately, the information will be computed using the tables that exist in the database similar to the stats for the game and team data. The main difference between game data and player data is that game data will have stats for the performance of the players in the game, whereas player data will be the performance of the player over the games in the database that they have participated in.

On the first instance of the menu for Player Data, users will be allowed to choose between pitchers and batters. If they choose pitchers, they will see pitching stats for that player and if they choose batters, they will see batting info for that player. There will be some overlap between the players since many pitchers also bat for their team, but not all batters are pitchers. That is why we decided to separate the data to keep the displayed format consistent.

For batters, they can enter the *'show batters'* command to show all players who have batted for their team in our database. We query this by simply joining the *PlayerNames* table with the *AtBats* table using *PlayerNames.id* and *AtBats.batterID*. We only grab distinct ids since we are only focusing on who has batted and selected the first and last names. For pitchers, we do the same thing except we join the tables using *AtBats.pticherID* instead.

As for the actual data that is computed, the common data between the pitchers and batters is the hand that they play with (batting direction and pitching direction), the past and current teams that they played for, and their ejections. The way we retrieve the player ejections is similar to our team and game data, but we focus on the *playerID* attribute and only select those whose id matches the player we are focusing on.

For pitchers, when we are looking for their data, we have two tables; one table is for their pitching stats such as their earned runs, pitching average, etc. through the years whereas the second table focuses more on their pitch types. In MLB a pitcher can pitch a ball in a variety of ways and our client and dataset has allowed us to see the different types of pitches a player throws with details including the start speed, end speed, and spin rate of the baseball. We use the aggregate *AVG* and *COUNT* functions to get the results that we add directly to the table. An example of a pitcher's stats on the client application is seen below.

Justin Verlander
Pitching: Right-Handed

Past and Current Teams:
Detroit Tigers
Houston Astros

Yearly Pitching Stats

Year	G	GS	W-L	ERA	IP	H	ER	HR	NP	HB	BB	IBB	SO	AVG	WHIP
2015	20	20	7-13	1.55	279.0	113	48	13	2150	3	32	1	113	0.229	0.52
2016	34	34	20-14	1.47	423.0	171	69	30	3668	8	57	1	254	0.206	0.54
2017	33	33	18-15	1.48	396.0	170	65	27	3546	4	72	4	218	0.221	0.61
2018	34	34	21-13	1.59	346.0	156	61	28	3427	8	37	0	290	0.199	0.56
----------	----	----	-------	------	-------	-----	----	----	------	----	----	-----	-----	------	------
Player Totals	121	121	66-55	1.52	1444.0	610	243	98	12791	23	198	6	875	0.214	0.56

Pitch Breakdown

Pitch Type	Total Thrown	Average Start Speed (MPH)	Average End Speed (MPH)	Average Spin Rate
Four-seam Fastball	7506	94.26	86.02	2654.60
Changeup	705	85.63	79.19	2087.13
Curveball	1968	79.51	73.79	1322.43
Slider	2537	87.25	80.95	813.06
Intentional ball	6	77.68	71.78	1768.58
Pitchout	2	85.65	79.75	2254.64
Two-seam Fastball	12	95.64	86.81	3042.89
Cutter	34	91.76	84.38	1568.40
-------------------	-------------	--------------------------	------------------------	------------------+
Player Totals	12770	87.17	80.33	1938.97

Player Ejections

Game	Team	Inning	Argue BS	Correct
Seattle Mariners at Houston Astros, 2018-08-09	Houston Astros	2	No	N/A

Moving on to batting players, most of the procedures are the same, although it is much simpler to analyze the results to compute the data that is displayed on the table. Most of the columns like 1B, 2B, 3B, HR are grabbed based on the type of event from the record. The more complex stats like BA is the player's batting average, which is the number of bases they were able to make based on their hit over the number of times they were at bat (AB). OBP is their on-base percentage and SLG is their slugging percentage. The final column is the sum of their on-base percentage and slugging percentage. Below, you can see an example of what the batting stats look like for a batting player.

```
Enter the name of the batter you are looking for: Pillar, Kevin

Kevin Pillar
Batting: Right-Handed

Past and Current Teams:
Toronto Blue Jays
```

Yearly Batting Stats												
Year	GP	W-L	AB	H	1B	2B	3B	HR	BA	OBP	SLG	OPS
2015	157	92-65	587	163	118	31	2	12	0.278	0.332	0.399	0.731
2016	143	78-65	548	146	102	35	2	7	0.266	0.321	0.376	0.697
2017	151	72-79	588	150	96	37	1	16	0.255	0.321	0.403	0.724
2018	139	63-76	513	129	72	40	2	15	0.251	0.298	0.425	0.723
Player Totals	590	305-285	2236	588	388	143	7	50	0.263	0.318	0.401	0.719

Player Ejections				
Game	Team	Inning	Argue BS	Correct
Toronto Blue Jays at Tampa Bay Rays, 2017-08-22	Toronto Blue Jays	3	Yes	Yes

Modifying Data

```
Do you want to modify it by:
        1. Updating Data
        2. Inserting Data
        3. Deleting Data
Enter where you want to go:
```

In these three aspects, the users can also modify the data including updating existing records, adding new records, and removing old records. The application will not allow the user to modify the data that may affect a game's outcome including at-bat or pitch details.

```
Do you want to update:
        1. Teams Data
        2. Players Data
        3. Ejections Data
        4. Games Data
Enter where you want to go:
```

The Team Data, Player Data, Ejection Data, and Game Data are opened to the user to modify. However, there are some limitations for the users to keep the database consistent. For example, the user cannot change or delete the primary key of a data because it is referenced

from other data by the foreign key. The user cannot also insert a duplicate primary key because the primary key has to be unique. The application will attempt to perform modifications on the server-side and catch any error from the server. The user will receive a message about whether the operation is successful.

The following operations will use Team Data as an example. The image below shows the currently active teams in MLB. The user can type *'show teams'* in the command to show all teams in the league. The list can help the user to identify whether their modifications are successful or just to check the currently active teams in the league so far.

```
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: show teams
Below is all MLB Teams
Angels              White Sox           Tigers              Brewers             Phillies            Cardinals
Diamondbacks        Cubs                Astros              Twins               Pirates             Rays
Braves              Reds                Royals              Yankees             Padres              Rangers
Orioles             Indians             Dodgers             Mets                Mariners            Blue Jays
Red Sox             Rockies             Marlins             Athletics           Giants              Nationals
```

Updating Data

When the user wants to update the data, the application will ask the user to input a team's full information including the ones the user does not wish to update. For instance, if the user would like to change the name *'Blue Jays'* to *'Maple Leafs'* without changing the *city*, the user still needs to input *'TOR, Toronto, Maple Leafs'*. Although the city stays unchanged, the application will still ask for full information from the user for the simplicity of SQL query and code algorithm.

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: TOR, Toronto, Maple Leafs
Success!
```

If the SQL query is done successfully, the application will return a *'Success'* message. From the *'show teams'* command, we can see that the team Toronto *'Blue Jays'* has changed its name to *'Maple Leafs'*. It is in the bottom right of the list.

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: show teams
Below is all MLB Teams
Angels              White Sox           Tigers              Brewers             Phillies            Cardinals
Diamondbacks        Cubs                Astros              Twins               Pirates             Rays
Braves              Reds                Royals              Yankees             Padres              Rangers
Orioles             Indians             Dodgers             Mets                Mariners            Maple Leafs
Red Sox             Rockies             Marlins             Athletics           Giants              Nationals
```

Since the database uses *abbreviation* as a primary key, the user cannot update an *abbreviation* value that does not exist in the database. The application initially attempts to execute the SQL query and catches any error and returns it if the server reports unsuccessful attempts. For instance, if the user wants to modify the team *'ABC'*, the application will return an error stating that the input is not valid.

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: ABC, DEF, GHI
Not a valid team name!
```

Because the modification is constrained by the primary key and foreign key, the user cannot modify the *abbreviation.* This will stay unchanged for the consistency and correctness of the database. For example, if the user would like to change the *abbreviation* of Toronto Maple Leafs from *'TOR'* to *'ABC'*, the application will throw an error indicating that the input is invalid.

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: ABC, Toronto, Maple Leafs
Not a valid team name!
```

Inserting Data

Inserting data works the same way as updating the data. The input requires the user to provide full information. Due to its primary key constraint, the user cannot enter an existing *abbreviation*. The application will show an error.

```
Insert New Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Example: TOR, Toronto, Blue Jays
Enter the team's info: TOR, Toronto, Blue Jays
Invalid Input
```

If the *abbreviation* is unique and does not exist in the database yet, the application will perform insertion if all the integrity tests of the input pass. Take the following, for example, the user would like to add a new team that has the *abbreviation 'WTL'* in *'Waterloo'* that has the name *'Warriors'*.

```
Insert New Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Example: TOR, Toronto, Blue Jays
Enter the team's info: WTL, Waterloo, Warriors
Success!
```

The user can double-check whether the team is successfully added by checking the list of teams. From *'show teams'*, a team named *'Warriors'* is added to the current active list of teams in MLB.

```
Insert New Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Example: TOR, Toronto, Blue Jays
Enter the team's info: show teams
Below is all MLB Teams
Angels              Cubs                Royals              Mets                Giants              Warriors
Diamondbacks        Reds                Dodgers             Athletics           Cardinals
Braves              Indians             Marlins             Phillies            Rays
Orioles             Rockies             Brewers             Pirates             Rangers
Red Sox             Tigers              Twins               Padres              Maple Leafs
White Sox           Astros              Yankees             Mariners            Nationals
```

Deleting Data

Deleting data is relatively easy compared to the other two. The input only requires the unique identification of the information a user wishes to delete, which is the primary key of the data. In the following example, the user should only input the *abbreviation* of the team to delete it. The user is also offered a chance to make sure that the input is correct before deletion. The application will return a *'Success'* message if the query executes without an error and return an error message otherwise.

```
Delete Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>". Must be a valid abbreviation
Example: TOR
Enter the team's info: WTL
Are you sure you want to delete this data?
        1. Yes
        2. No
Your decision: 1
Success!
```

From *'show teams'*, a team named *'Warriors'* is deleted from the current active list of teams in MLB.

```
Delete Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>". Must be a valid abbreviation
Example: TOR
Enter the team's info: show teams
Below is all MLB Teams
Angels              White Sox           Tigers              Brewers             Phillies            Cardinals
Diamondbacks        Cubs                Astros              Twins               Pirates             Rays
Braves              Reds                Royals              Yankees             Padres              Rangers
Orioles             Indians             Dodgers             Mets                Mariners            Maple Leafs
Red Sox             Rockies             Marlins             Athletics           Giants              Nationals
```

Due to its foreign key constraints, the operation may not be successful if the user wishes to delete a team that has been referenced in other data. The user will receive an error message if the operation fails.

Work Plan and Responsibilities

This project is quite large and will be completed over several weeks. The first few weeks will primarily be spent on discussing our application outline and the goals of our project, including what features and things we want to implement into our application and their priority. Not everything we wish to put into our application will be in the final product, so it is important to get the priority and the outline done early to save time down the line. We will also be using this time to go over the dataset and choose which columns and tables are the most useful and what type of information we can create with the data in these tables. There are many tables and much of the data is overlapping so there is the possibility that we do not need to use all of the data that exists in the datasets to eventually create our entity-relationship design.

The next phase of the project will be the actual design and implementation. Here we will be actually coding and slowly developing the application. We will focus on the server-side of the application by developing the SQL code to create all of the tables, views, defining constraints and keys, etc. While this happens, we can also be working on the client-side at the same time as there are many aspects of the front end and back end that do not overlap, so by working on these components that are individual to their side, we will be able to get more work done efficiently by developing the core skeleton of the client and server-side of the application at the same time. For the client application, that includes setting up the command-line interface and developing the possible paths that the user can take through their inputs.

After that will be the testing and debugging phase. This is mainly just developing test cases and ensuring our application is displaying correct information as well as making sure our server-side database is configured correctly. There may be some errors or improper data from the CSV files, so we have to make sure that we aren't duplicating records or omitting records that aren't supposed to be removed.

The project is composed of three components, so it is sensible that each person is responsible for one of these components. These people will be ensuring that the other members are meeting the deadlines and working on their particular section of the project. This project is a collaborative effort, so it is important to keep the project schedule on track by making sure we are meeting the goals of the project scope.

1. SQL Server-Side Development – Matthew Lee
2. Client-Side Application Development – Martin Hsu
3. Testing (Client and Server-Side) and Documentation – Jeffrey Yuen

Project Assumptions

- The data is for the MLB regular season only. Baseball rules are applied
- At bat ids and game, ids are not related. They are related to the year instead.
- Players can bat/pitch left or right-handed or both. They can also switch their stand in a game. Therefore, there's no relation between a batter/pitcher and whether they are going to bat/pitch left or right-handed at each at-bat.
- Umpires are not that important for general fans. They are removed from the table because we don't have sufficient information to create another table for umpires. It lacks a primary key and we don't have the information for 2019 as well.
- A team can play in different venues when they are the home team, and we don't have the relations for that. Therefore, there's no specific home field for each team. For instance, TOR's home field is Rogers Centre in Toronto, but they usually play a couple of games in Montreal each year.
- Most stats of each pitch are too detailed that most of the general fans may not care or understand. We will focus on the stats that are generally available and visible for the audiences in-game or on TV, mainly focusing on the speed of the ball
- Some of the ejections in the ejections.csv file are related to coaches and managers. Since these people are not in the *player_names* table and it's better to enforce a foreign key on the playerID of the ejection table, we will remove these records and only focus our dataset on player ejections

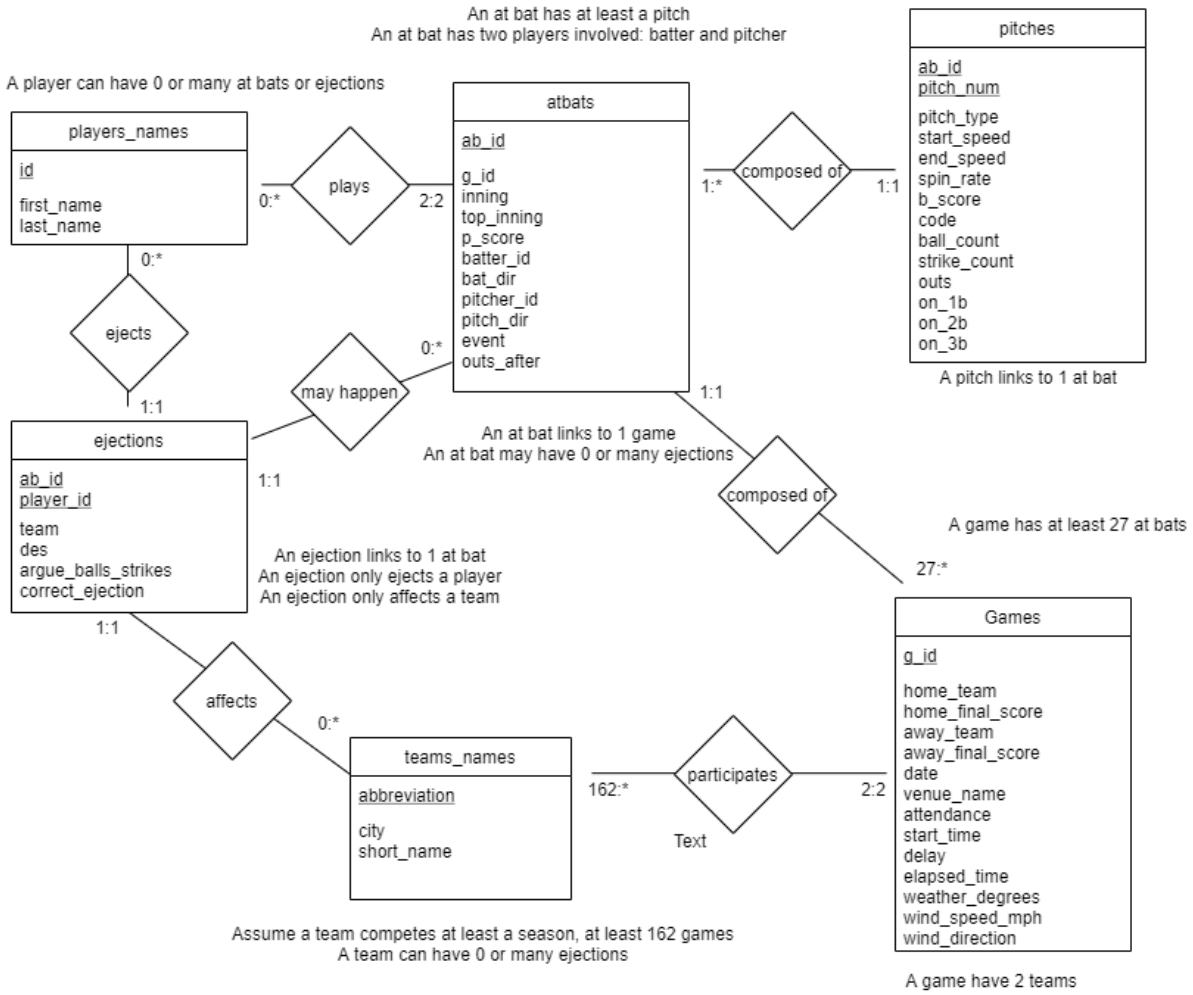
ER design

In our MLB dataset, there are several issues. First of all, the data from 2015 to 2018 are grouped while 2019 stands alone. For instance, we will have 2 sets of game data, and one

is from 2015 to 2018 and another one is 2019. We can load the data into the corresponding table without an issue but data from 2019 lacks a lot of information. The columns are the same but there are a significant number of NULL values. Data from 2019 simply is not compatible with the rest of the data so the team decides to discard the data from 2019.

Secondly, some data is less important to the general fans and we would like to construct a database that is for general baseball fans to use. We decide to drop specific data of the pitch, including the locations and break angle or length of the pitch as these columns seem too complex for general fans and would simply be a waste of space in our database. We also do not include officials in the game since it does not have any information for the officials except their names. There is a lack of information and we cannot ensure whether the dataset remains correct if we move the officials out, create a table for it and assign IDs for them. Furthermore, there is a column called event_num which is not explained by the author who created this dataset. We decide to remove this data as it does not mean anything to the dataset and we cannot find any meaningful correlation with this attribute to the other attributes in the table.

Third of all, the data is not normalized. We perform normalizations and decompose the data to meet the requirements. Most of the data satisfy 2NF but not 3NF and BCNF. After normalizations are completed, ER diagram is designed, which includes all the relations and can be seen below.



Our design is composed of 6 entities: *player_names*, *ejections*, *atbats*, *teams_names*, *pitches*, and *games*. In each, the underlined attributes are the primary key of that entity and will be the unique identifier for each row. Thankfully, most of the CSV files already have primary keys, so there was not much effort in combining rows or creating new rows to create primary keys. The *ejections* entity and *pitches* entity are the only entities that have a primary key as a set of attributes.

Considering the cardinality of each entity, we can see that the entities are related to one another in some sort of way. Starting with *player_names*, each player can appear in the *atbats* entity many times, or none at all. These can be players that we just created, so they will not have participated in any games, so they will have no at-bat appearances. Some players participate in many games and go to bat multiple times in a single game, so they will

have multiple instances in the *atbats* entity. The *player_names* entity also has a correlation to the *ejections* entity as a player can never be ejected, or they can be ejected several times through the years.

The next entity we will focus on is the *atbats* entity. Each record in the entity will have a pitcher and a batter ID. That means each row in the entity can be only two players that are tied to the *player_names* entity. Furthermore, an ejection can happen in a game, but it can also not happen. There are many instances in the dataset where there have been multiple ejections in a game, so the cardinality between the *ejections* to the *atbats* entities is defined as zero to many. Next, an *atbat* cardinality to the *pitches* entity is one-to-many. Each batter must receive at the very least one pitch, but it can also receive many pitches (fouls, balls, strikes, etc.). Finally, the last cardinality for this entity is with *games*. This cardinality is one-to-one since a batter's at-bat can only correlate to an instance in one game. He can bat multiple times in a game, but not for multiple games.

Moving to the next entity, the *pitches* entity is only related directly to the *atbats* entity. A pitcher can only pitch to one batter at a time, so the cardinality for this is one-to-one.

The *games* entity has two cardinalities that we will discuss. The first cardinality is with *atbats*. A game is composed of at the minimum nine innings and each inning will have three outs. That means if the pitching team has a complete shutout game where the batting team doesn't get any hits, there will be at the very least twenty-seven instances. On the other hand, there can also be many hits if the game goes on for a long time. The next cardinality is with the *team_names* entity. Only two teams can participate in a game: the home team and the away team, so the cardinality is two to two.

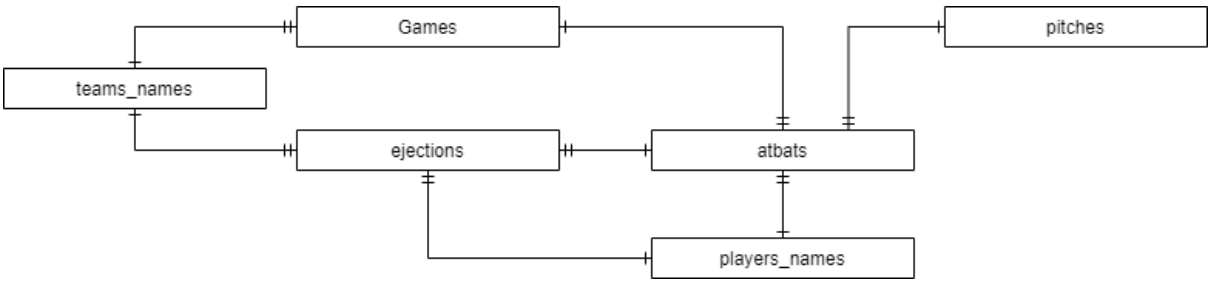
The penultimate entity is the *team_names* entity, which is composed of all the active teams in the MLB. If we assume that every team completes a full season, then they will play at the very least 162 games and can play more games than that if they play through the

playoffs. The relationship with the *ejections* entity is that when a player gets ejected, they belong to a team, so a team can have many ejections, but at the same time, it can also have no ejections if none of their players have been ejected from a game.

Finally, the last entity is the *ejections* entity. As mentioned previously, it has a direct relationship with the *atbats, player_names,* and *team_names* entities. During an ejection, we eject one player at a time and each player belongs to only one team. Also, we only consider ejections that occur during live play, so when there is someone at-bat. That is why all of the cardinalities for the *ejections* entity are one to one. In a real game, multiple ejections can happen, but since they are to multiple people, we consider these as separate instances.

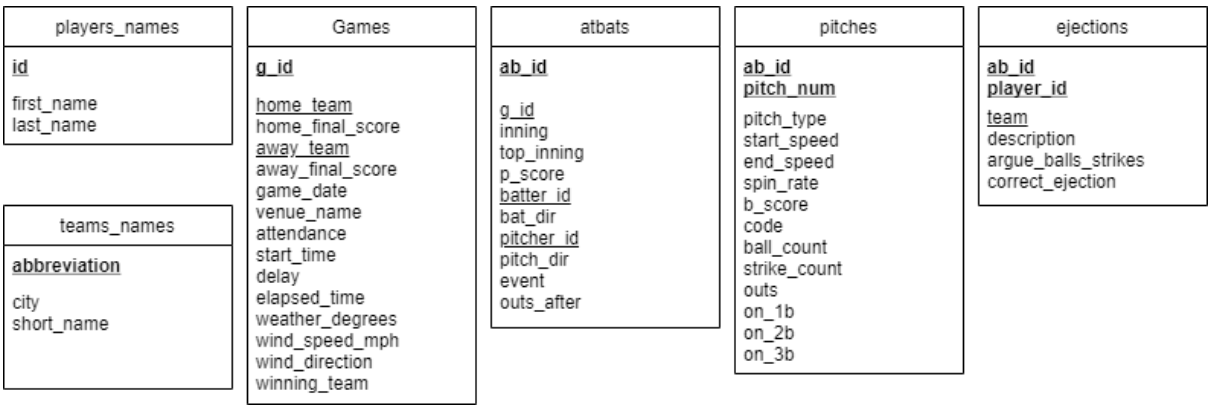
Relational Schema

Similar to our ER Design, we essentially created our relational schema following the same assumptions and ideas.



The diagram above shows the six relations and the basic cardinalities it has with the other relations. This follows the same principle as the cardinalities but is reduced to just uni-directional and will either be one to one or one to many. A simple example is the *ejections* relation and the *player_names* relation. In the ER Design, we had two cardinalities for the two entities to describe the cardinality from both directions. In this schema, we simplified it to say the *ejections* and *player_names* relation is many to one. In other words, many ejections can belong to one player, or one player can have many ejections. This may

not be as detailed to explain the cardinality between the relations as it is shown in the ER Design, but it shows the basic idea of how they are related.



The image above shows more specifically each relation and its defined attributes. The attributes in bold represent the relation's primary keys and the underlined attributes that are not bold represent foreign keys to attributes of other relations. Other aspects of the schema that will be explained but are not mentioned are the constraints on the relations. In the *Games* relation, the *home_team* and *away_team* attributes are foreign keys to the *Team_Names* relation's *abbreviation* attribute. In *ejections*, *the team attribute* is also a foreign key referencing the same attribute. In the *AtBats* relation, *g_id* is a foreign key that references *Games.g_id*, and the *batter_id* and *pitcher_id* attributes are foreign keys that reference *player_names.id*. The last set of foreign keys is the *ab_id* attribute in *pitches* and *ejections* that are foreign keys to *atbats.ab_id.*

To make sure that the data from the dataset, as well as the data we as users insert or update using the client application, is valid, we have applied several constraints and check statements in addition to these keys. In the *Games* relation, we added check statements to both *home_final_score* and *away_final_score* to make sure that they are both non-negative integers. We also add a check to the *g_id* attribute to make sure that the first four digits of the id corresponding to the year that the game took place, which is from the *gameDate* attribute. There's another check statement to the *winning_team* attribute to make sure that the attribute

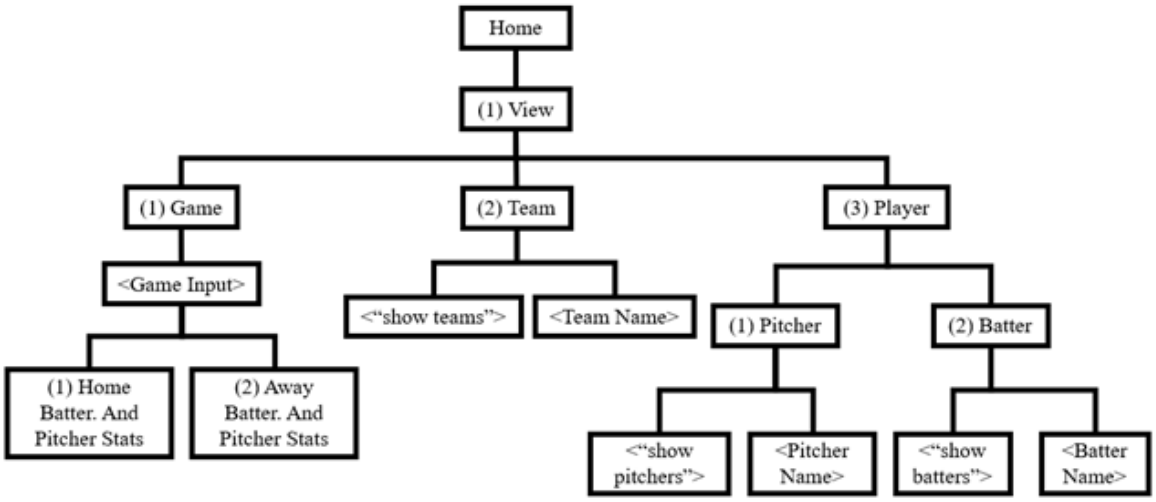
matches either the *home_team* or *away_team* attribute since those should be the only possibilities. For the *AtBats* relation, we have a check constraint on the *pScore* that it has to be non-negative. For the *inning* attribute, it has to be greater than 0, since games start in the first inning. Additionally, we have check statements on both *batDir* and *pitchDir* that they can only be 'L' or 'R'. The *outsAfter* attribute can only be an int from 0 to 3 since that's the most number of outs in an inning. A sanity check we added as a constraint is to make sure that the first four digits of *abID* and *gID* are the same as they should both correspond to the year of the game date. Finally, *topInning* can only be 'TRUE' or 'FALSE'. For the *pitches* relation, we have check statements on *ballCount, outs,* and *strikeCount* to make sure they are between 0-4, 0-3. and 0-3 respectively. The *bScore* attribute must be non-negative, and *pitchNum*, *startSpeed,* and *endSpeed* all must be greater than zero. The three attributes *on1B, on2B, on3B*, represent whether or not someone is on that base and they can only be values 'TRUE' and 'FALSE'. Lastly, we have two attributes *pitchType* and *code* and they can only be specific values from a predefined list, so we added check constraints for those 2 attributes. Finally, the last check constraints are on the *ejections* relation and we make sure that *argueBallsStrikes* and *correctEjection* can only be 'TRUE' or 'FALSE'. An important note is that we did not add any additional indexes to our database since after testing, and developing our application, most of our queries are filtering using the primary and foreign keys that have been provided and translated by the user. These keys are indexes by default, so the need to add additional keys is not necessary.

Test Plan

Client Application Testing

To test our application, we created test cases for all the features we implemented and manually tested them with a visual inspection. As we have described, these features can be split up into two distinct categories: Viewing data and Altering data.

Testing Viewing Data



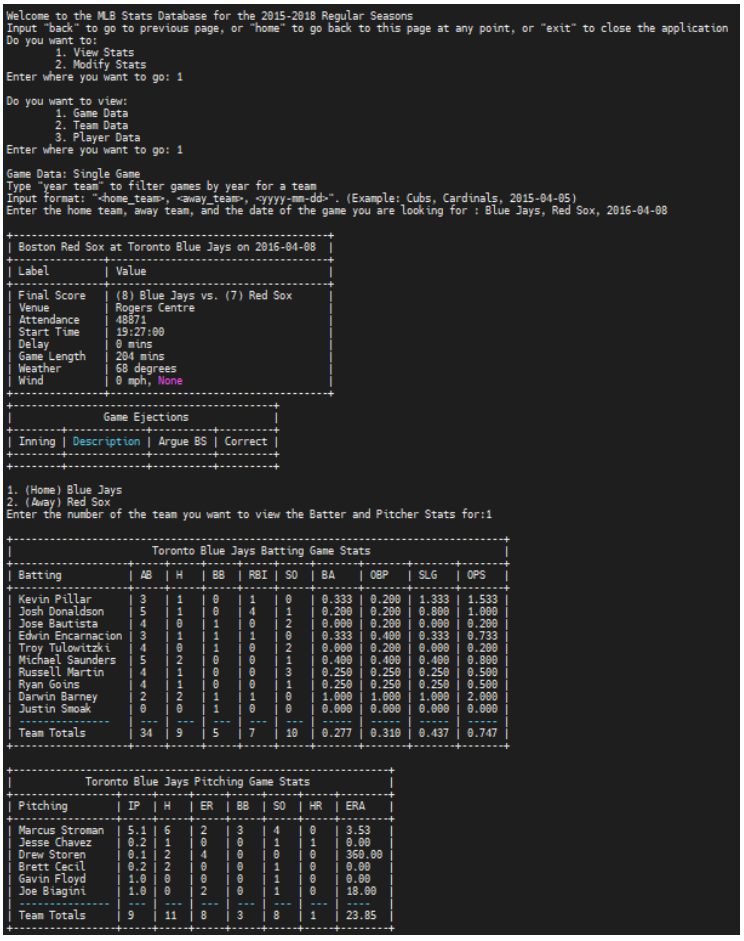
This figure shows all the possible pathways the user can take for viewing data. As such, we have the following test cases and inputs for those test cases (in the table, strings within parentheses are the input strings which will be followed by the return key, and each input is separated by a comma):

Test Case	Inputs
1: Game and Home Stats	(1), (1), (Blue Jays, Red Sox, 2016-04-08), (1)
2: Game and Away Stats	(1), (1), (Blue Jays, Red Sox, 2016-04-08), (2)

3: Show all teams	(1), (2), (show teams)
4: Single Team Stats	(1), (2), (Blue Jays)
5: Show all pitchers	(1), (3), (1), (show pitchers)
6: Single Pitcher Stats	(1), (3), (1), (Kershaw, Clayton)
7: Show all batters	(1), (3), (2), (show batters)
8: Single Batter Stats	(1), (3), (2), (Blackmon, Charlie)

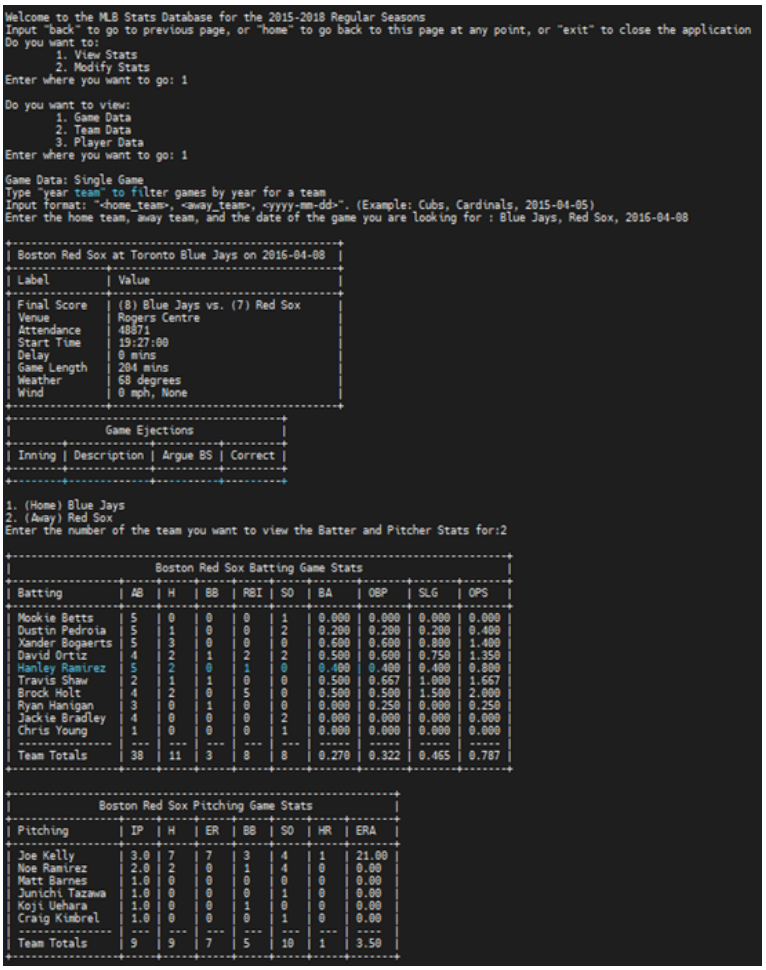
Test Case 1: Game and Home Stats

The data for the inputted game is displayed properly and requesting the home teams data gives us all the batting and pitching data for the home team players for this game.



Test Case 2: Game and Away Stats

The data for the inputted game is displayed properly and requesting the away teams data gives us all the batting and pitching data for the away team players for this game.



```
Welcome to the MLB Stats Database for the 2015-2018 Regular Seasons
Input "back" to go to previous page, or "home" to go back to this page at any point, or "exit" to close the application
Do you want to:
        1. View Stats
        2. Modify Stats
Enter where you want to go: 1

Do you want to view:
        1. Game Data
        2. Team Data
        3. Player Data
Enter where you want to go: 1

Game Data: Single Game
Type "year team" to filter games by year for a team
Input format: "<home_team>, <away_team>, <yyyy-mm-dd>". (Example: Cubs, Cardinals, 2015-04-05)
Enter the home team, away team, and the date of the game you are looking for : Blue Jays, Red Sox, 2016-04-08

+-----------------------------------------+
| Boston Red Sox at Toronto Blue Jays on 2016-04-08  |
+--------------+--------------------------+
| Label        | Value                    |
+--------------+--------------------------+
| Final Score  | (8) Blue Jays vs. (7) Red Sox      |
| Venue        | Rogers Centre            |
| Attendance   | 48871                    |
| Start Time   | 19:27:00                 |
| Delay        | 0 mins                   |
| Game Length  | 204 mins                 |
| Weather      | 68 degrees               |
| Wind         | 0 mph, None              |
+--------------+--------------------------+
+-----------------------------------+
|          Game Ejections           |
+--------+-------------+----------+---------+
| Inning | Description | Argue BS | Correct |
+--------+-------------+----------+---------+
+--------+-------------+----------+---------+

1. (Home) Blue Jays
2. (Away) Red Sox
Enter the number of the team you want to view the Batter and Pitcher Stats for:2

+-----------------------------------------------------------------------+
|                    Boston Red Sox Batting Game Stats                  |
+----------------+----+---+----+-----+----+-------+-------+-------+-------+
| Batting        | AB | H | BB | RBI | SO | BA    | OBP   | SLG   | OPS   |
+----------------+----+---+----+-----+----+-------+-------+-------+-------+
| Mookie Betts   | 5  | 0 | 0  | 0   | 1  | 0.000 | 0.000 | 0.000 | 0.000 |
| Dustin Pedroia | 5  | 1 | 0  | 0   | 2  | 0.200 | 0.200 | 0.200 | 0.400 |
| Xander Bogaerts| 5  | 3 | 0  | 0   | 0  | 0.600 | 0.600 | 0.800 | 1.400 |
| David Ortiz    | 4  | 2 | 1  | 2   | 2  | 0.500 | 0.600 | 0.750 | 1.350 |
| Hanley Ramirez | 5  | 2 | 0  | 1   | 0  | 0.400 | 0.400 | 0.400 | 0.800 |
| Travis Shaw    | 2  | 1 | 1  | 0   | 0  | 0.500 | 0.667 | 1.000 | 1.667 |
| Brock Holt     | 4  | 2 | 0  | 5   | 0  | 0.500 | 0.500 | 1.500 | 2.000 |
| Ryan Hanigan   | 3  | 0 | 1  | 0   | 0  | 0.000 | 0.250 | 0.000 | 0.250 |
| Jackie Bradley | 4  | 0 | 0  | 0   | 2  | 0.000 | 0.000 | 0.000 | 0.000 |
| Chris Young    | 1  | 0 | 0  | 0   | 1  | 0.000 | 0.000 | 0.000 | 0.000 |
| -------------- | --- | --- | --- | --- | --- | ----- | ----- | ----- | ----- |
| Team Totals    | 38 | 11 | 3  | 8   | 8  | 0.270 | 0.322 | 0.465 | 0.787 |
+----------------+----+---+----+-----+----+-------+-------+-------+-------+

+-----------------------------------------------------------+
|              Boston Red Sox Pitching Game Stats            |
+----------------+-----+---+----+----+----+----+-----------+
| Pitching       | IP  | H | ER | BB | SO | HR | ERA       |
+----------------+-----+---+----+----+----+----+-----------+
| Joe Kelly      | 3.0 | 7 | 7  | 3  | 4  | 1  | 21.00     |
| Noe Ramirez    | 2.0 | 2 | 0  | 1  | 4  | 0  | 0.00      |
| Matt Barnes    | 1.0 | 0 | 0  | 0  | 0  | 0  | 0.00      |
| Junichi Tazawa | 1.0 | 0 | 0  | 0  | 1  | 0  | 0.00      |
| Koji Uehara    | 1.0 | 0 | 0  | 1  | 0  | 0  | 0.00      |
| Craig Kimbrel  | 1.0 | 0 | 0  | 0  | 1  | 0  | 0.00      |
| -------------- | --- | --- | --- | --- | --- | --- | ----      |
| Team Totals    | 9   | 9 | 7  | 5  | 10 | 1  | 3.50      |
+----------------+-----+---+----+----+----+----+-----------+
```

Test Case 3: Show all teams

All of the teams' names are successfully displayed.

```
Welcome to the MLB Stats Database for the 2015-2018 Regular Seasons
Input "back" to go to previous page, or "home" to go back to this page at any point, or "exit" to close the application
Do you want to:
        1. View Stats
        2. Modify Stats
Enter where you want to go: 1

Do you want to view:
        1. Game Data
        2. Team Data
        3. Player Data
Enter where you want to go: 2

Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for: show teams
Below is all MLB Teams
Angels              White Sox           Tigers              Brewers             Phillies            Cardinals
Diamondbacks        Cubs                Astros              Twins               Pirates             Rays
Braves              Reds                Royals              Yankees             Padres              Rangers
Orioles             Indians             Dodgers             Mets                Mariners            Maple Leafs
Red Sox             Rockies             Marlins             Athletics           Giants              Nationals

Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for:
```

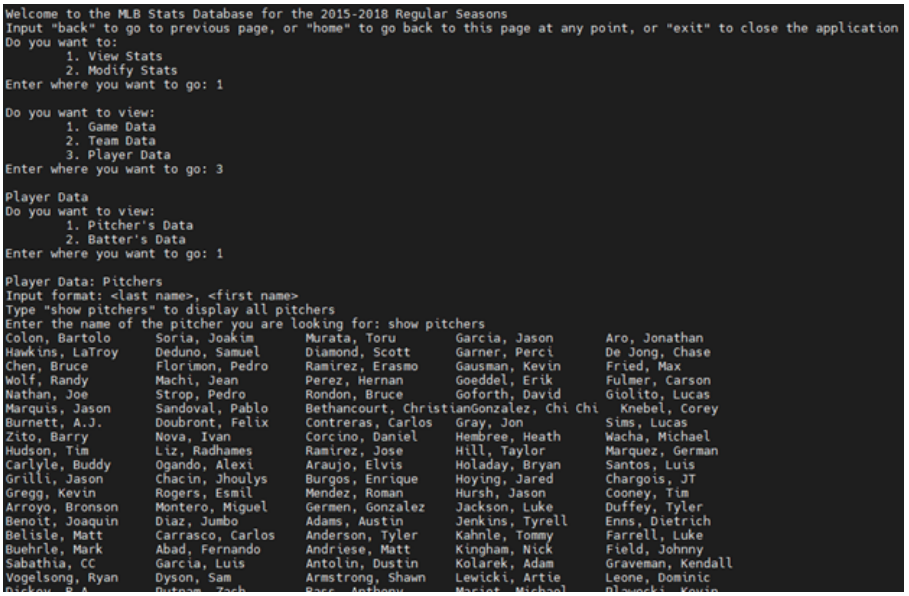
Test Case 4: Single Team Stats

All the data for the inputted team from 2015-2018 seasons are displayed correctly along with ejection data.

```
Welcome to the MLB Stats Database for the 2015-2018 Regular Seasons
Input "back" to go to previous page, or "home" to go back to this page at any point, or "exit" to close the application
Do you want to:
       1. View Stats
       2. Modify Stats
Enter where you want to go: 1

Do you want to view:
       1. Game Data
       2. Team Data
       3. Player Data
Enter where you want to go: 2

Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for: Blue Jays

+-------------------------------------------------------------------------------------------------------+
|                                    Toronto Blue Jays 2015-2018                                        |
+--------+--------+------------+---------------+--------------------+---------------------+-----------------+
| Season | Record | Total Runs | Avg Runs/Game | Total Runs Allowed | Avg Runs Allowed/Game | Total Attendance |
+--------+--------+------------+---------------+--------------------+---------------------+-----------------+
| 2015   | 94-69  | 891        | 9.48          | 736                | 7.83                | 2794891         |
| 2016   | 89-73  | 758        | 8.52          | 721                | 8.1                 | 3392099         |
| 2017   | 76-86  | 693        | 9.12          | 747                | 9.83                | 3203886         |
| 2018   | 73-89  | 709        | 9.71          | 741                | 10.15               | 2325282         |
+--------+--------+------------+---------------+--------------------+---------------------+-----------------+

+-----------------------------------------------------------------------------------------------+
|                                  Team Ejections 2015-2018                                     |
+-----------------------------------------------+--------+------------------+----------+---------+
| Game                                          | Inning | Player           | Argue BS | Correct |
+-----------------------------------------------+--------+------------------+----------+---------+
| Baltimore Orioles at Toronto Blue Jays, 2015-06-19  | 8      | Roberto Osuna    | No       | N/A     |
| Toronto Blue Jays at Kansas City Royals, 2015-07-12 | 8      | Jose Bautista    | No       | N/A     |
| Kansas City Royals at Toronto Blue Jays, 2015-08-02 | 8      | Aaron Sanchez    | No       | N/A     |
| Baltimore Orioles at Toronto Blue Jays, 2015-09-05  | 7      | Cliff Pennington | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jose Bautista    | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Josh Donaldson   | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jesse Chavez     | No       | N/A     |
| Toronto Blue Jays at Minnesota Twins, 2016-05-21    | 1      | Josh Donaldson   | Yes      | Yes     |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 2      | Edwin Encarnacion | Yes     | No      |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 14     | Russell Martin   | Yes      | Yes     |
| Toronto Blue Jays at Seattle Mariners, 2016-09-19   | 7      | Josh Donaldson   | Yes      | No      |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Russell Martin   | Yes      | Yes     |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Marcus Stroman   | Yes      | Yes     |
| Toronto Blue Jays at Tampa Bay Rays, 2017-08-22     | 3      | Kevin Pillar     | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-05     | 18     | Josh Donaldson   | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-25     | 8      | Jose Bautista    | Yes      | Yes     |
| Boston Red Sox at Toronto Blue Jays, 2018-08-07     | 8      | Ryan Tepera      | Yes      | N/A     |
| Toronto Blue Jays at New York Yankees, 2018-09-15   | 2      | Luke Maile       | Yes      | Yes     |
+-----------------------------------------------+--------+------------------+----------+---------+

Team Data
Type "show teams" to display all active teams
Input format: "<team_name>". Do not include the city. (Example: Blue Jays)
Enter the team's name of the team you are looking for:
```

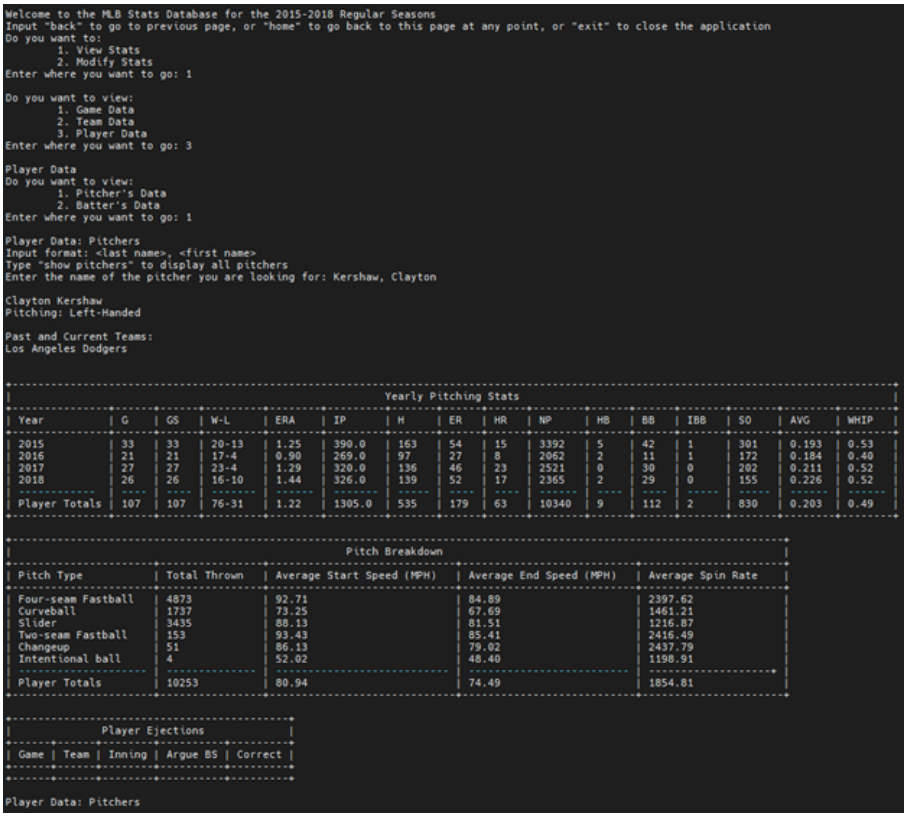
Test Case 5: Show all pitchers

All the pitchers in our database are displayed (there are far more rows that are not captured in this screenshot).



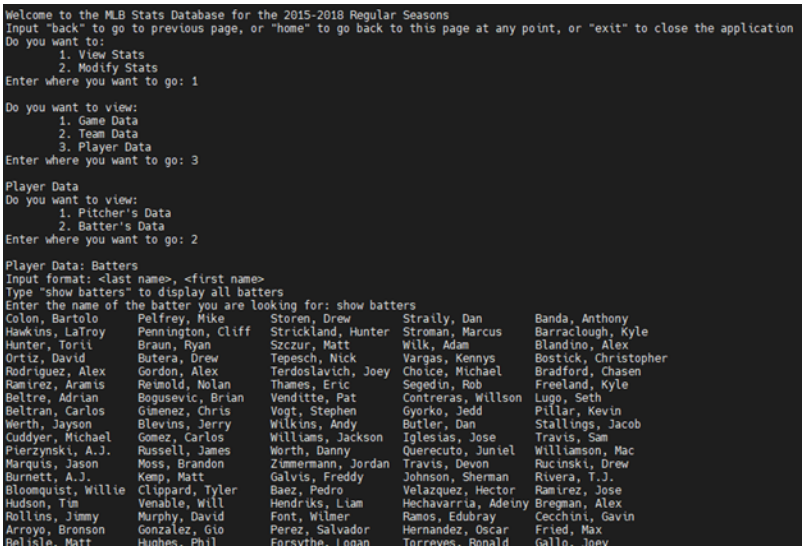
Test Case 6: Single Pitcher Stats

All the inputted pitcher's pitching stats and pitch breakdowns are displayed correctly.



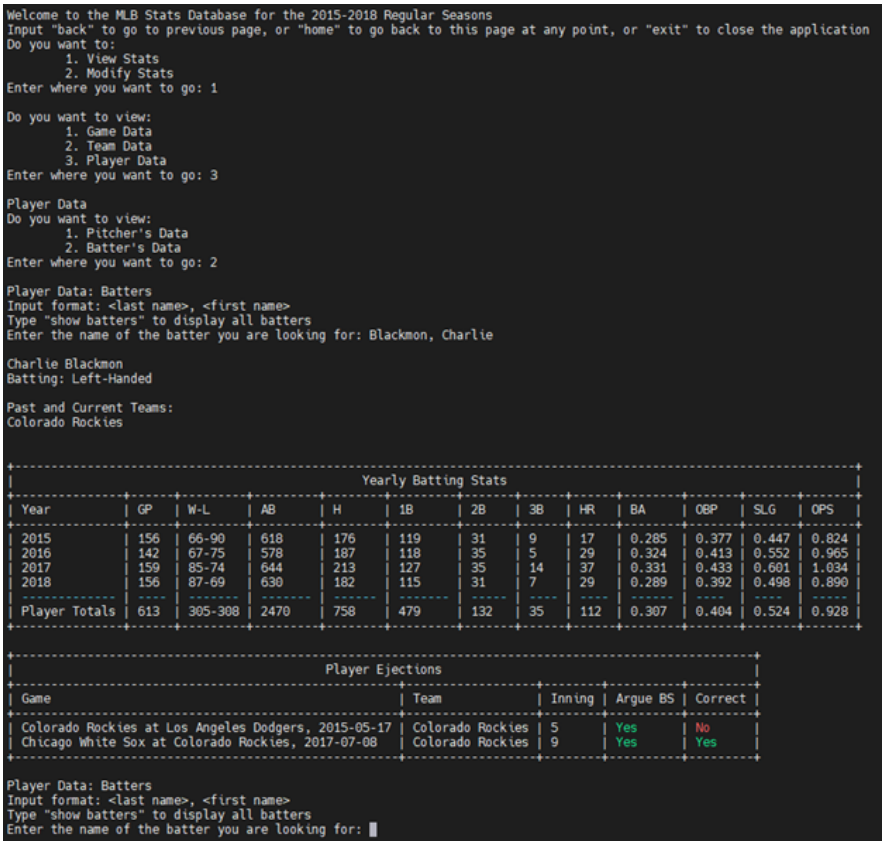
Test Case 7: Show all batters

All the batters in our database are displayed (there are far more rows that are not captured in this screenshot).

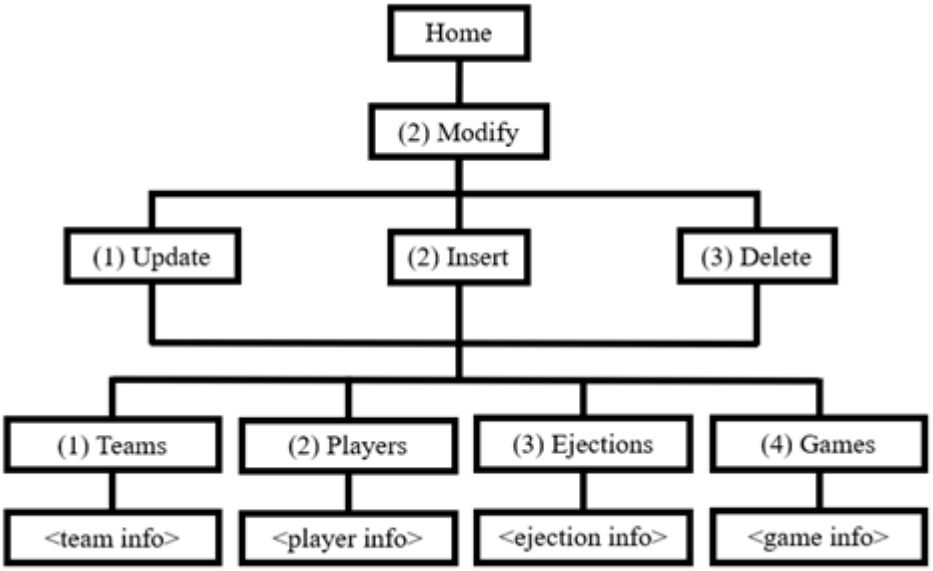


Test Case 8: Single Batter Stats

All the inputted batter's batting stats and batting breakdowns are displayed correctly.



Testing Altering Data



The above figure shows all the pathways a user can take to alter data in the database. For each of the three altering methods, 4 sets can be altered: Teams, Players, Ejections, and Games. As such, we have the following test cases and inputs for those test cases (same format as before) in the table below. For each test case, after the inputs are inputted, we will check if the data has been altered using our view functions (inputs will not be shown as view was shown extensively in the previous section, but results will be shown in screenshots). We did not test the error cases for trivial invalid inputs and SQL errors (i.e. inserting game where gameID for game already exists) as we expect the user to follow the input format and know the limitations of the SQL database.

Test Case	Inputs
1: Inserting Teams	(2), (2), (1), (ABC, abcdef, Abcs)
2: Updating Team	(2), (1), (1), (ABC, abcdef, defs)
3: Deleting Team	(2), (3), (1), (ABC), (1)
4: Inserting Players	(2), (2), (2), (999999, Jeffrey, Yuen)
5: Updating Player	(2), (1), (2), (999999, Jeff, Yuen)
6: Deleting Player	(2), (3), (2), (999999), (1)
7: Inserting Ejections	(2), (2), (3), (2015000009, 516770, TOR, Ejection, FALSE, TRUE)
8: Updating Ejection	(2), (1), (3), (2015000009, 516770, TOR, Ejection, TRUE, FALSE)
9: Deleting Ejection	(2), (3), (3), (2015000009, 516770), (1)
10: Inserting Games	(2), (2), (4), (201599999, TOR, 3, ANA, 0, 2015-04-16, Wrigley Field, 35000, 19:00:00, 0, 184, 45, 7, CF, ANA)
11: Updating Game	(2), (1), (4), (201599999, TOR, 3, ANA, 0, 2015-04-16, Wrigley Field, 10000, 15:00:00, 5, 190, 40, 5, RF, TOR)
12: Deleting Game	(2), (3), (4), (201599999)

Note: The following screenshots will not show the initial navigation using the number inputs to save space and as the functionality has been shown in the previous section's test cases.

Test Case 1: Inserting Teams

Input returns successfully:

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: ABC, abcdef, Abcs
Success!
```

When we show all teams, team has been inserted:

```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: show teams
Below is all MLB Teams
Abcs              White Sox         Astros            Yankees           Mariners          Nationals
Angels            Cubs              Royals            Mets              Giants
Diamondbacks      Reds              Dodgers           Athletics         Cardinals
Braves            Indians           Marlins           Phillies          Rays
Orioles           Rockies           Brewers           Pirates           Rangers
Red Sox           Tigers            Twins             Padres            Blue Jays
```

Test Case 2:

Input returns successfully:

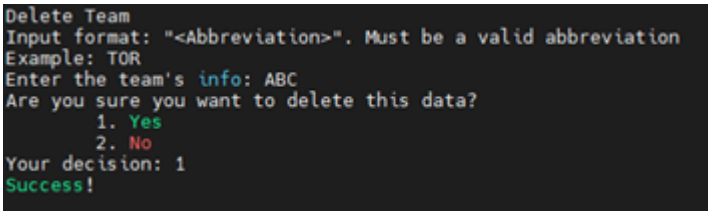
```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: ABC, abcdef, defs
Success!
```

When we show all teams, the “ABC” team’s short name has been changed to “defs”:

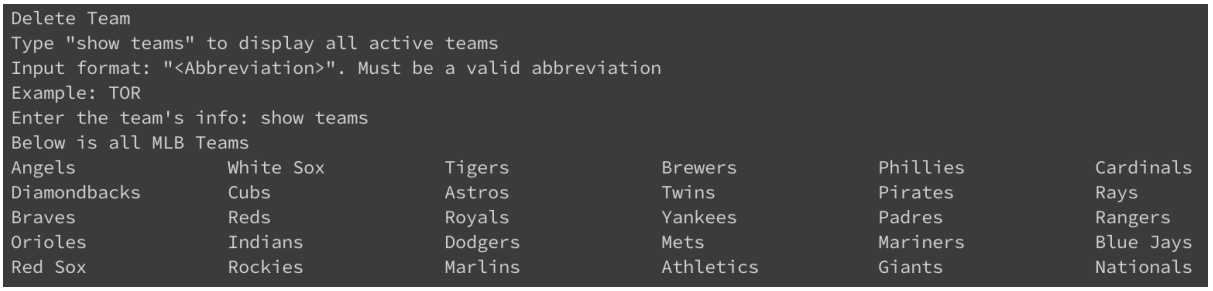
```
Update Team
Type "show teams" to display all active teams
Input format: "<Abbreviation>, <City>, <Short Name>". Must be a valid Abbreviation. Example: TOR, Toronto, Blue Jays
Enter the team's info: show teams
Below is all MLB Teams
defs              White Sox         Astros            Yankees           Mariners          Nationals
Angels            Cubs              Royals            Mets              Giants
Diamondbacks      Reds              Dodgers           Athletics         Cardinals
Braves            Indians           Marlins           Phillies          Rays
Orioles           Rockies           Brewers           Pirates           Rangers
Red Sox           Tigers            Twins             Padres            Blue Jays
```

Test Case 3: Deleting Team

Deleting team ABC return successfully:

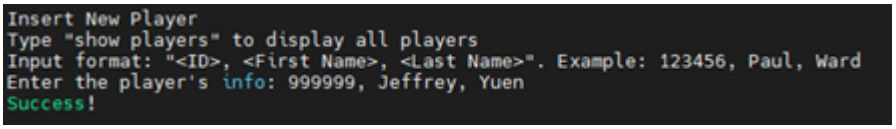


When we show all teams, the team no longer exists:

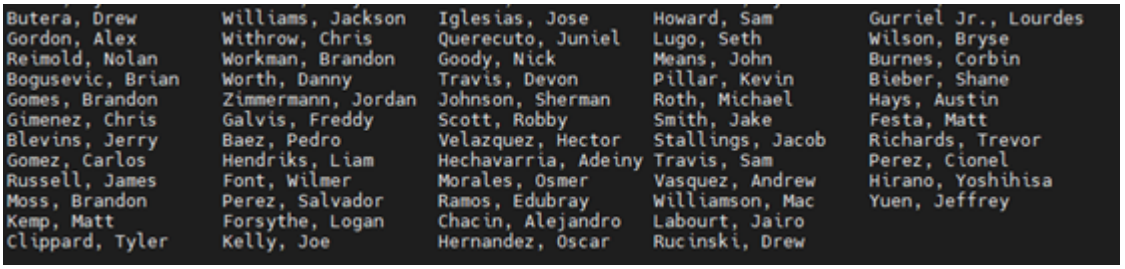


Test Case 4: Inserting Players

Input returns successfully:



Showing all players, we can see new player Yuen, Jeffrey has added (bottom right):



Test Case 5: Updating Player

Input returns success:

```
Update Player
Type "show players" to display all players
Input format: "<ID>, <First Name>, <Last Name>". Must be a valid ID. Example: 123456, Paul, Ward
Enter the player's info: 999999, Jeff, Yuen
Success!
```

Showing all players, we can see that the player's name has been updated to Jeff instead of Jeffrey:

```
Gurriel Jr., Lourdes
Wilson, Bryse
Burnes, Corbin
Bieber, Shane
Hays, Austin
Festa, Matt
Richards, Trevor
Perez, Cionel
Hirano, Yoshihisa
Yuen, Jeff
```

Test Case 6: Deleting Player

Deleting the newly created player with Player ID 999999 returns success:

```
Delete Player
Input format: "<Player ID>". Must be a valid ID
Example: 000001
Enter the player's info: 999999
Are you sure you want to delete this data?
        1. Yes
        2. No
Your decision: 1
Success!
```

Showing all players, we can see that player does not exist anymore in our database:

```
Garcia, Adolis
Gurriel Jr., Lourdes
Wilson, Bryse
Burnes, Corbin
Bieber, Shane
Hays, Austin
Festa, Matt
Richards, Trevor
Perez, Cionel
Hirano, Yoshihisa
```

Test Case 7: Inserting Ejections

Input returns success:

```
Insert Ejection
Input format: "<At Bat ID>, <Player ID>, <Team>, <Description>, <Arguing B/S>, <Correct Ejection>". Must be valid IDs and Team. Last two are T/F value
Example: 2015000001, 000001, TOR, Toronto first baseman Paul Ward ejected by HP Harder W, FALSE, TRUE
Enter the ejection's info: 2015000009, 516770, TOR, Ejection, FALSE, TRUE
Success!
```

When viewing the team data, we can see the new ejection shows up in the team ejections

```
                                    Team Ejections 2015-2018
+-----------------------------------------------+--------+------------------+----------+---------+
| Game                                          | Inning | Player           | Argue BS | Correct |
+-----------------------------------------------+--------+------------------+----------+---------+
| St. Louis Cardinals at Chicago Cubs, 2015-04-05     | 1      | Starlin Castro   | No       | Yes     |
| Baltimore Orioles at Toronto Blue Jays, 2015-06-19  | 8      | Roberto Osuna    | No       | N/A     |
| Toronto Blue Jays at Kansas City Royals, 2015-07-12 | 8      | Jose Bautista    | No       | N/A     |
| Kansas City Royals at Toronto Blue Jays, 2015-08-02 | 8      | Aaron Sanchez    | No       | N/A     |
| Baltimore Orioles at Toronto Blue Jays, 2015-09-05  | 7      | Cliff Pennington | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jose Bautista    | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Josh Donaldson   | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jesse Chavez     | No       | N/A     |
| Toronto Blue Jays at Minnesota Twins, 2016-05-21    | 1      | Josh Donaldson   | Yes      | Yes     |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 2      | Edwin Encarnacion | Yes     | No      |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 14     | Russell Martin   | Yes      | Yes     |
| Toronto Blue Jays at Seattle Mariners, 2016-09-19   | 7      | Josh Donaldson   | Yes      | No      |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Russell Martin   | Yes      | Yes     |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Marcus Stroman   | Yes      | Yes     |
| Toronto Blue Jays at Tampa Bay Rays, 2017-08-22     | 3      | Kevin Pillar     | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-05     | 18     | Josh Donaldson   | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-25     | 8      | Jose Bautista    | Yes      | Yes     |
| Boston Red Sox at Toronto Blue Jays, 2018-08-07     | 8      | Ryan Tepera      | Yes      | N/A     |
| Toronto Blue Jays at New York Yankees, 2018-09-15   | 2      | Luke Maile       | Yes      | Yes     |
+-----------------------------------------------+--------+------------------+----------+---------+
```

Test Case 8: Updating Ejection

Input returns success. We changed the Argue BS to TRUE and Correct to FALSE:

```
Update Ejection
Input format: "<At Bat ID>, <Player ID>, <Team>, <Description>, <Arguing B/S>, <Correct Ejection>". Must be valid IDs and Team. Last two are T/F value
Example: 2015000001, 000001, TOR, Toronto first baseman Paul Ward ejected by HP Harder W, FALSE, TRUE
Enter the ejection's info: 2015000009, 516771, TOR, Ejection, TRUE, FALSE
Success!
```

When viewing the team data for the Blue Jays, we can see that the data is updated:

```
                                    Team Ejections 2015-2018
+-----------------------------------------------+--------+------------------+----------+---------+
| Game                                          | Inning | Player           | Argue BS | Correct |
+-----------------------------------------------+--------+------------------+----------+---------+
| St. Louis Cardinals at Chicago Cubs, 2015-04-05     | 1      | Starlin Castro   | Yes      | No      |
| Baltimore Orioles at Toronto Blue Jays, 2015-06-19  | 8      | Roberto Osuna    | No       | N/A     |
| Toronto Blue Jays at Kansas City Royals, 2015-07-12 | 8      | Jose Bautista    | No       | N/A     |
| Kansas City Royals at Toronto Blue Jays, 2015-08-02 | 8      | Aaron Sanchez    | No       | N/A     |
| Baltimore Orioles at Toronto Blue Jays, 2015-09-05  | 7      | Cliff Pennington | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jose Bautista    | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Josh Donaldson   | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jesse Chavez     | No       | N/A     |
| Toronto Blue Jays at Minnesota Twins, 2016-05-21    | 1      | Josh Donaldson   | Yes      | Yes     |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 2      | Edwin Encarnacion | Yes     | No      |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 14     | Russell Martin   | Yes      | Yes     |
| Toronto Blue Jays at Seattle Mariners, 2016-09-19   | 7      | Josh Donaldson   | Yes      | No      |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Russell Martin   | Yes      | Yes     |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Marcus Stroman   | Yes      | Yes     |
| Toronto Blue Jays at Tampa Bay Rays, 2017-08-22     | 3      | Kevin Pillar     | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-05     | 18     | Josh Donaldson   | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-25     | 8      | Jose Bautista    | Yes      | Yes     |
| Boston Red Sox at Toronto Blue Jays, 2018-08-07     | 8      | Ryan Tepera      | Yes      | N/A     |
| Toronto Blue Jays at New York Yankees, 2018-09-15   | 2      | Luke Maile       | Yes      | Yes     |
+-----------------------------------------------+--------+------------------+----------+---------+
```

Test Case 9: Deleting Ejection

Deleting the newly created ejection returns success:

```
Delete Ejection
Input format: "<At Bat ID>, <Player ID>". Must be a valid IDs
Example: 2015000000, 000001
Enter the ejection's info: 2015000009, 516771
Are you sure you want to delete this data?
        1. Yes
        2. No
Your decision: 1
Success!
```

When viewing the team data for the Blue Jays, we can see that the newly created ejection no longer exists:

```
                                    Team Ejections 2015-2018
+-----------------------------------------------------+--------+--------------------+----------+---------+
| Game                                                | Inning | Player             | Argue BS | Correct |
+-----------------------------------------------------+--------+--------------------+----------+---------+
| Baltimore Orioles at Toronto Blue Jays, 2015-06-19  | 8      | Roberto Osuna      | No       | N/A     |
| Toronto Blue Jays at Kansas City Royals, 2015-07-12 | 8      | Jose Bautista      | No       | N/A     |
| Kansas City Royals at Toronto Blue Jays, 2015-08-02 | 8      | Aaron Sanchez      | No       | N/A     |
| Baltimore Orioles at Toronto Blue Jays, 2015-09-05  | 7      | Cliff Pennington   | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jose Bautista      | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Josh Donaldson     | No       | N/A     |
| Toronto Blue Jays at Texas Rangers, 2016-05-15      | 8      | Jesse Chavez       | No       | N/A     |
| Toronto Blue Jays at Minnesota Twins, 2016-05-21    | 1      | Josh Donaldson     | Yes      | Yes     |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 2      | Edwin Encarnacion  | Yes      | No      |
| Cleveland Indians at Toronto Blue Jays, 2016-07-01  | 14     | Russell Martin     | Yes      | Yes     |
| Toronto Blue Jays at Seattle Mariners, 2016-09-19   | 7      | Josh Donaldson     | Yes      | No      |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Russell Martin     | Yes      | Yes     |
| Oakland Athletics at Toronto Blue Jays, 2017-07-27  | 5      | Marcus Stroman     | Yes      | Yes     |
| Toronto Blue Jays at Tampa Bay Rays, 2017-08-22     | 3      | Kevin Pillar       | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-05     | 18     | Josh Donaldson     | Yes      | Yes     |
| Toronto Blue Jays at Boston Red Sox, 2017-09-25     | 8      | Jose Bautista      | Yes      | Yes     |
| Boston Red Sox at Toronto Blue Jays, 2018-08-07     | 8      | Ryan Tepera        | Yes      | N/A     |
| Toronto Blue Jays at New York Yankees, 2018-09-15   | 2      | Luke Maile         | Yes      | Yes     |
+-----------------------------------------------------+--------+--------------------+----------+---------+
```

Test Case 10: Inserting Games

Input returns success:

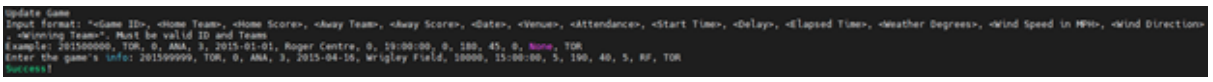
```
Insert Game
Input format: "<Game ID>, <Home Team>, <Home Score>, <Away Team>, <Away Score>, <Date>, <Venue>, <Attendance>, <Start Time>, <Delay>, <Elapsed Time>, <Weather Degrees>, <Wind Speed in MPH>, <Wind Direction>
, <Winning Team>". Must be valid ID and Teams
Example: 201500000, TOR, 0, ANA, 3, 2015-01-01, Roger Centre, 0, 19:00:00, 0, 180, 45, 0, None, ANA
Enter the game's info: 201599999, TOR, 0, ANA, 3, 2015-04-16, Wrigley Field, 35000, 19:00:00, 0, 184, 45, 7, CF, ANA
Success!
```

When we view the game, we can see that all of the data we inputted is in the displayed:

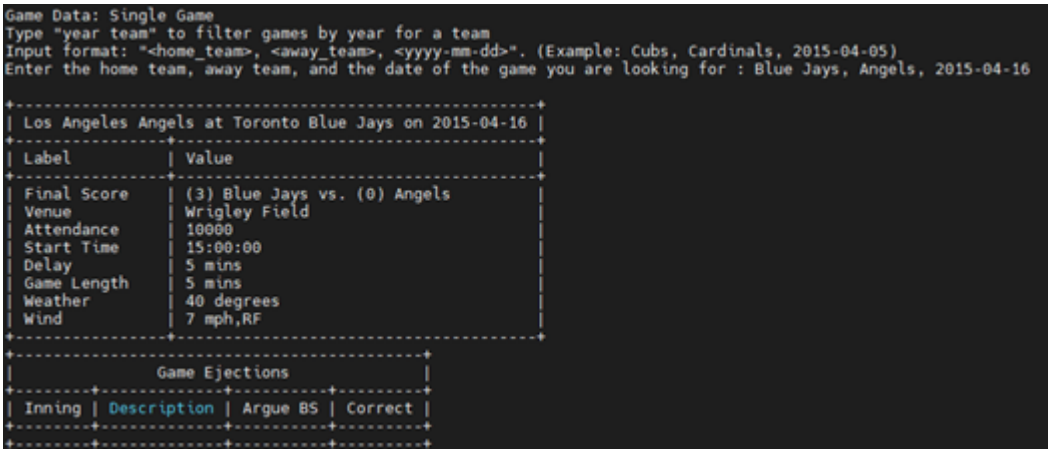
```
+-----------------------------------------------+
| Los Angeles Angels at Toronto Blue Jays on 2015-04-16 |
+----------------+------------------------------+
| Label          | Value                        |
+----------------+------------------------------+
| Final Score    | (3) Blue Jays vs. (0) Angels |
| Venue          | Wrigley Field                |
| Attendance     | 35000                        |
| Start Time     | 19:00:00                     |
| Delay          | 0 mins                       |
| Game Length    | 184 mins                     |
| Weather        | 45 degrees                   |
| Wind           | 7 mph,CF                     |
+----------------+------------------------------+
+-----------------------------------------+
|             Game Ejections              |
+--------+-------------+----------+---------+
| Inning | Description | Argue BS | Correct |
+--------+-------------+----------+---------+
+--------+-------------+----------+---------+
```

Test Case 11: Updating Game

Input returns success:

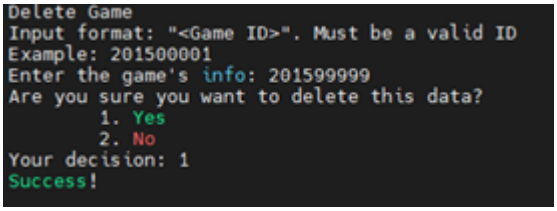


When we view the game, we can see that all the updates we made to the game are changed and displayed correctly:

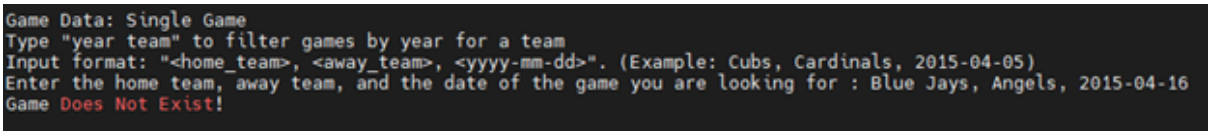


Test Case 12: Deleting Game

Deleting the newly created game returns success:



When we try to look for the game it tells us that “Game Does Not Exist!” so it was deleted correctly:



SQL Testing

In SQL testing, we create another set of data that does not have integrity constraints and any check statement to check the correctness of each data value. The idea is similar to assignment 2: we would like to see if the dataset loses data once the constraints are added.

We use the same code from assignment 2 and instead of showing the table that lost data, we would like to see the total count of the table that lost data. The result should be an empty set for the table that lost data which will result in a total count of 0.

```
+--------------------------+
| TotalCountTablesLostData |
+--------------------------+
|                        0 |
+--------------------------+
1 row in set (34.92 sec)
```

From the final result, it is indeed 0. Therefore, we can conclude that we do not lose any data after adding constraints to the data set. Our data set should pass all integrity checks.