# HW07 Baseball Statistics

## Introduction

The Lahman Baseball Database is a comprehensive database of Major League baseball statistics. The journalist Sean Lahman provides all of this data freely to the public. We will make use of some of his data in this assignment. If you would like to learn more about the database, you can visit his website.

We have provided you with a CSV file named batting.csv that contains the annual batting performance data for all Major League Baseball players dating back to the year 1871. The first row in the file is a header indicating what data is stored in each column of the file. For example, column 12 is labeled "HR" and contains the number of home runs the player hit that year. Each of the next 99,846 lines contains a comma separated list of the data for that player and year. For example, the fifth line in the file indicates that a player with the id allisdo01 hit 2 home runs in 1871.

You should download batting.csv and place it in the same directory as your Python code. Your job will be to write a Python program that finds the player id of the player with the highest total career RBIs of all time. *Caution!* Your program should work with any similarly formatted CSV file.

## Data Input: Opening the file

First, you will need to read in the data file. You can do this by opening the file using the open function and iterating through each line. To parse the data contained in each line, you will need to use the split method. We are interested in two columns, 'playerID' and 'RBI'. (Your program should skip the header in the file and completely ignore any lines where the RBI column does not contain a digit.)

You should create an accumulator dictionary called "career_rbis" that maps each player id string to an integer representing the total number of RBIs for that player. As you iterate through input the file, you should update the "career_rbis" dictionary.

## Data Processing: Finding the most RBIs

After reading in the data and generating the "career_rbis" dictionary, you can now iterate through the dictionary to find the player with the most career RBIs. *Hint:* You will need two accumulator variables to track both the most RBIs you've seen so far AND the player having that many RBIs.

Store an integer representing the highest number of RBIs in a variable

named max_rbis and the corresponding player id string in a variable named max_player.

## Data Output: Submitting your solution

After you finish writing your code, you must submit it on http://others.zlcnup.com/cs101 before Dec 4th, 6pm. If you finish your code with Jupyter Notebook, you should create a hw07.py file, then paste your code into it. That's to say, you should upload your *.py file finally, other format is NOT allowed. You needn't rename your file, the system will check your information and rename your file. You'd better read the help information on website, we will firstly update on website if we have anything changed.

Note: During Wensday 8:00—18:00(LAB TIME), you can't upload your homework except lab's materials. The system will check it.

## Using files and directories

In your final submission, you should use **open('batting.csv')** with no directory path in your code, on your own machine things may behave differently. Briefly, the *best* thing to do is to figure out where Python is running and move your file batting.csv there. Otherwise you can try to find out where your file is located and refer to it directly using the code shared at the end of lecture #13.