

Baseball Databank is a compilation of historical baseball data in a convenient, tidy format, distributed under Open Data terms.

This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License. For details see:  
<http://creativecommons.org/licenses/by-sa/3.0/>

Person identification and demographics data are provided by Chadwick Baseball Bureau (<http://www.chadwick-bureau.com>), from its Register of baseball personnel.

Player performance data for 1871 through 2014 is based on the Lahman Baseball Database, version 2015-01-24, which is Copyright (C) 1996-2015 by Sean Lahman.

The tables Parks.csv and HomeGames.csv are based on the game logs and park code table published by Retrosheet. This information is available free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at <http://www.retrosheet.org>.

## 1. Introduction

This database contains pitching, hitting, and fielding statistics for Major League Baseball from 1871 through 2014. It includes data from the two current leagues (American and National), the four other "major" leagues (American Association, Union Association, Players League, and Federal League), and the National Association of 1871-1875.

This database was created by Sean Lahman, who pioneered the effort to make baseball statistics freely available to the general public. What started as a one man effort in 1994 has grown tremendously, and now a team of researchers have collected their efforts to make this the largest and most accurate source for baseball statistics available anywhere.

This version of the database is available in SQL files or in a generic, comma delimited format.

## 2. Data Tables

The design follows these general principles. Each player is assigned a unique number (playerID). All of the information relating to that player is tagged with his playerID. The playerIDs are linked to names and birthdates in the PLAYER table.

The database is comprised of the following main tables:

- Player - Player names, DOB, and biographical info
- Batting - batting statistics
- Pitching - pitching statistics

Fielding - fielding statistics

It is supplemented by these tables:

AllStarFull - All-Star appearances  
HallofFame - Hall of Fame voting data  
Managers - managerial statistics  
Teams - yearly stats and standings  
BattingPost - post-season batting statistics  
PitchingPost - post-season pitching statistics  
TeamFranchises - franchise information  
FieldingOutfield - outfield position data  
FieldingPostseason- post-season fielding data  
ManagersHalf - split season data for managers  
TeamsHalf - split season data for teams  
Salary - player salary data  
PostSeason - post-season series information  
ManagerAward - awards won by managers  
PlayerAward - awards won by players  
ManagerAwardVote - award voting for manager awards  
PlayerAwardVote - award voting for player awards  
Appearances - details on the positions a player appeared at  
College - list of colleges that players attended  
PlayerCollege - list of players and the colleges they attended

The tables parks.csv and home\_game.csv are based on the game logs and park code table published by Retrosheet. This information is available free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at <http://www.retrosheet.org>.

## 2.1 Player table

playerID	A unique code assigned to each player. The playerID links the data in this file with records in the other files.
birthYear	Year player was born
birthMonth	Month player was born
birthDay	Day player was born
birthCountry	Country where player was born
birthState	State where player was born
birthCity	City where player was born
deathYear	Year player died
deathMonth	Month player died
deathDay	Day player died
deathCountry	Country where player died
deathState	State where player died
deathCity	City where player died
nameFirst	Player's first name
nameLast	Player's last name
nameGiven	Player's given name (typically first and middle)

weight	Player's weight in pounds
height	Player's height in inches
bats	Player's batting hand (left, right, or both)
throws	Player's throwing hand (left or right)
debut	Date that player made first major league appearance
finalGame	Date that player made first major league appearance (blank if still active)
retroID	ID used by retrosheet
bbrefID	ID used by Baseball Reference website

## 2.2 Batting Table

playerID	Player ID code
yearID	Year
stint	player's stint (order of appearances within a season)
teamID	Team
lgID	League
G	Games
AB	At Bats
R	Runs
H	Hits
2B	Doubles
3B	Triples
HR	Homeruns
RBI	Runs Batted In
SB	Stolen Bases
CS	Caught Stealing
BB	Base on Balls
SO	Strikeouts
IBB	Intentional walks
HBP	Hit by pitch
SH	Sacrifice hits
SF	Sacrifice flies
GIDP	Grounded into double plays

## 2.3 Pitching table

playerID	Player ID code
yearID	Year
stint	player's stint (order of appearances within a season)
teamID	Team
lgID	League
W	Wins
L	Losses
G	Games
GS	Games Started
CG	Complete Games
SHO	Shutouts

SV	Saves
IPOuts	Outs Pitched (innings pitched x 3)
H	Hits
ER	Earned Runs
HR	Homeruns
BB	Walks
SO	Strikeouts
BAOpp	Opponent's Batting Average
ERA	Earned Run Average
IBB	Intentional Walks
WP	Wild Pitches
HBP	Batters Hit By Pitch
BK	Balks
BFP	Batters faced by Pitcher
GF	Games Finished
R	Runs Allowed
SH	Sacrifices by opposing batters
SF	Sacrifice flies by opposing batters
GIDP	Grounded into double plays by opposing batter

-----  
-----

## 2.4 Fielding Table

playerID	Player ID code
yearID	Year
stint	player's stint (order of appearances within a season)
teamID	Team
lgID	League
Pos	Position
G	Games
GS	Games Started
InnOuts	Time played in the field expressed as outs
PO	Putouts
A	Assists
E	Errors
DP	Double Plays
PB	Passed Balls (by catchers)
WP	Wild Pitches (by catchers)
SB	Opponent Stolen Bases (by catchers)
CS	Opponents Caught Stealing (by catchers)
ZR	Zone Rating

-----  
-----

## 2.5 AllstarFull table

playerID	Player ID code
YearID	Year
gameNum season)	Game number (zero if only one All-Star game played that season)
gameID	Retrosheet ID for the game idea
teamID	Team

lgID	League
GP	1 if Played in the game
startingPos	If player was game starter, the position played

-----  
2.6 HallOfFame table

playerID	Player ID code
yearID	Year of ballot
votedBy	Method by which player was voted upon
ballots	Total ballots cast in that year
needed	Number of votes needed for selection in that year
votes	Total votes received
inducted	Whether player was inducted by that vote or not (Y or N)
category	Category in which candidate was honored
needed_note	Explanation of qualifiers for special elections

-----  
2.7 Managers table

playerID	Player ID Number
yearID	Year
teamID	Team
lgID	League
inseason	Managerial order. Zero if the individual managed the
team	the entire year. Otherwise denotes where the manager
appeared	in the managerial order (1 for first manager, 2 for
second, etc.)	
G	Games managed
W	Wins
L	Losses
rank	Team's final position in standings that year
plyrMgr	Player Manager (denoted by 'Y')

-----  
2.8 Teams table

yearID	Year
lgID	League
teamID	Team
franchID	Franchise (links to TeamsFranchise table)
divID	Team's division
Rank	Position in final standings
G	Games played
GHome	Games played at home
W	Wins
L	Losses
DivWin	Division Winner (Y or N)

WCWin	Wild Card Winner (Y or N)
LgWin	League Champion(Y or N)
WSWin	World Series Winner (Y or N)
R	Runs scored
AB	At bats
H	Hits by batters
2B	Doubles
3B	Triples
HR	Homeruns by batters
BB	Walks by batters
SO	Strikeouts by batters
SB	Stolen bases
CS	Caught stealing
HBP	Batters hit by pitch
SF	Sacrifice flies
RA	Opponents runs scored
ER	Earned runs allowed
ERA	Earned run average
CG	Complete games
SHO	Shutouts
SV	Saves
IPOuts	Outs Pitched (innings pitched x 3)
HA	Hits allowed
HRA	Homeruns allowed
BBA	Walks allowed
SOA	Strikeouts by pitchers
E	Errors
DP	Double Plays
FP	Fielding percentage
name	Team's full name
park	Name of team's home ballpark
attendance	Home attendance total
BPF	Three-year park factor for batters
PPF	Three-year park factor for pitchers
teamIDBR	Team ID used by Baseball Reference website
teamIDlahman45	Team ID used in Lahman database version 4.5
teamIDretro	Team ID used by Retrosheet

-----  
-----  
2.9 BattingPost table

yearID	Year
round	Level of playoffs
playerID	Player ID code
teamID	Team
lgID	League
G	Games
AB	At Bats
R	Runs
H	Hits
2B	Doubles

3B	Triples
HR	Homeruns
RBI	Runs Batted In
SB	Stolen Bases
CS	Caught stealing
BB	Base on Balls
SO	Strikeouts
IBB	Intentional walks
HBP	Hit by pitch
SH	Sacrifices
SF	Sacrifice flies
GIDP	Grounded into double plays

-----  
 -----  
 2.10 PitchingPost table

playerID	Player ID code
yearID	Year
round	Level of playoffs
teamID	Team
lgID	League
W	Wins
L	Losses
G	Games
GS	Games Started
CG	Complete Games
SHO	Shutouts
SV	Saves
IPOuts	Outs Pitched (innings pitched x 3)
H	Hits
ER	Earned Runs
HR	Homeruns
BB	Walks
SO	Strikeouts
BAOpp	Opponents' batting average
ERA	Earned Run Average
IBB	Intentional Walks
WP	Wild Pitches
HBP	Batters Hit By Pitch
BK	Balks
BFP	Batters faced by Pitcher
GF	Games Finished
R	Runs Allowed
SH	Sacrifice Hits allowed
SF	Sacrifice Flies allowed
GIDP	Grounded into Double Plays

-----  
 -----  
 2.11 TeamFranchises table

franchID	Franchise ID
franchName	Franchise name
active	Whether team is currently active (Y or N)
NAassoc	ID of National Association team franchise played as

-----  
-----

#### 2.12 FieldingOutfield table

playerID	Player ID code
yearID	Year
stint	player's stint (order of appearances within a season)
GlF	Games played in left field
GcF	Games played in center field
GrF	Games played in right field

-----  
-----

#### 2.13 FieldingPost table

playerID	Player ID code
yearID	Year
teamID	Team
lgID	League
round	Level of playoffs
Pos	Position
G	Games
GS	Games Started
InnOuts	Time played in the field expressed as outs
PO	Putouts
A	Assists
E	Errors
DP	Double Plays
TP	Triple Plays
PB	Passed Balls
SB	Stolen Bases allowed (by catcher)
CS	Caught Stealing (by catcher)

-----  
-----

#### 2.14 ManagersHalf table

playerID	Manager ID code
yearID	Year
teamID	Team
lgID	League
inseason	Managerial order. One if the individual managed the
team	the entire year. Otherwise denotes where the manager
appeared	in the managerial order (1 for first manager, 2 for
second, etc.)	
half	First or second half of season



G	Games managed
W	Wins
L	Losses
rank	Team's position in standings for the half

-----  
-----

#### 2.15 TeamsHalf table

yearID	Year
lgID	League
teamID	Team
half	First or second half of season
divID	Division
DivWin	Won Division (Y or N)
rank	Team's position in standings for the half
G	Games played
W	Wins
L	Losses

-----  
-----

#### 2.16 Salary table

yearID	Year
teamID	Team
lgID	League
playerID	Player ID code
salary	Salary

-----  
-----

#### 2.17 PostSeason table

yearID	Year
round	Level of playoffs
teamIDwinner	Team ID of the team that won the series
lgIDwinner	League ID of the team that won the series
teamIDloser	Team ID of the team that lost the series
lgIDloser	League ID of the team that lost the series
wins	Wins by team that won the series
losses	Losses by team that won the series
ties	Tie games

-----  
-----

#### 2.18 ManagerAward table

playerID	Manager ID code
awardID	Name of award won
yearID	Year
lgID	League
tie	Award was a tie (Y or N)

notes	Notes about the award
-------	-----------------------

-----  
-----  
2.19 PlayerAward table

playerID	Player ID code
awardID	Name of award won
yearID	Year
lgID	League
tie	Award was a tie (Y or N)
notes	Notes about the award

-----  
-----  
2.20 ManagerAwardVote table

awardID	name of award votes were received for
yearID	Year
lgID	League
playerID	Manager ID code
pointsWon	Number of points received
pointsMax	Maximum numner of points possible
votesFirst	Number of first place votes

-----  
-----  
2.21 PlayerAwardVote table

awardID	name of award votes were received for
yearID	Year
lgID	League
playerID	Player ID code
pointsWon	Number of points received
pointsMax	Maximum numner of points possible
votesFirst	Number of first place votes

-----  
-----  
2.22 Appearances table

yearID	Year
teamID	Team
lgID	League
playerID	Player ID code
G_all	Total games played
GS	Games started
G_batting	Games in which player batted
G_defense	Games in which player appeared on defense
G_p	Games as pitcher
G_c	Games as catcher

G_1b	Games as firstbaseman
G_2b	Games as secondbaseman
G_3b	Games as thirdbaseman
G_ss	Games as shortstop
G_lf	Games as leftfielder
G_cf	Games as centerfielder
G_rf	Games as right fielder
G_of	Games as outfielder
G_dh	Games as designated hitter
G_ph	Games as pinch hitter
G_pr	Games as pinch runner

---

-----

2.23 College table

schoolID	school ID code
schoolName	school name
schoolCity	city where school is located
schoolState	state where school's city is located
schoolNick	nickname for school's baseball team

---

-----

2.24 PlayerCollege table

playerid	Player ID code
schoolID	school ID code
year	year

<end of file>