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

This version of the Baseball databank was downloaded from [Sean Lahman's website](http://seanlahman.com/baseball-archive/statistics/).

Note that as of v1, this dataset is missing a few tables because of a restriction on the number of individual files that can be added. This is in the process of being fixed. The missing tables are Parks, HomeGames, CollegePlaying, Schools, Appearances, and FieldingPost.

The Data

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 MASTER table.

The database is comprised of the following main tables:

* MASTER - 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
* FieldingOF - outfield position data
* FieldingPost- post-season fielding data
* ManagersHalf - split season data for managers
* TeamsHalf - split season data for teams
* Salaries - player salary data
* SeriesPost - post-season series information
* AwardsManagers - awards won by managers
* AwardsPlayers - awards won by players
* AwardsShareManagers - award voting for manager awards
* AwardsSharePlayers - award voting for player awards
* Appearances - details on the positions a player appeared at
* Schools - list of colleges that players attended
* CollegePlaying - list of players and the colleges they attended

Descriptions of each of these tables can be found attached to their associated files, below.

Acknowledgments

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](http://www.retrosheet.org/).