

README CONTENTS 0.1 Copyright Notice 0.2 Contact Information

1.0 Release Contents 1.1 Introduction 1.2 What's New 1.3 Acknowledgements 1.4 Using this Database 1.5 Revision History

2.0 Data Tables 2.1 MASTER table 2.2 Batting Table 2.3 Pitching table 2.4 Fielding Table 2.5 All-Star table 2.6 Hall of Fame table 2.7 Managers table 2.8 Teams table 2.9 BattingPost table 2.10 PitchingPost table 2.11 TeamFranchises table 2.12 FieldingOF table 2.13 ManagersHalf table 2.14 TeamsHalf table 2.15 Salaries table 2.16 SeriesPost table 2.17 AwardsManagers table 2.18 AwardsPlayers table 2.19 AwardsShareManagers table 2.20 AwardsSharePlayers table 2.21 FieldingPost table 2.22 Appearances table 2.23 Schools table 2.24 SchoolsPlayers table

0.1 Copyright Notice & Limited Use License

This database is copyright 1996-2014 by Sean Lahman.

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/>

For licensing information or further information, contact Sean Lahman at: seanlahman@gmail.com

0.2 Contact Information

Web site: <http://www.baseball1.com> E-Mail : seanlahman@gmail.com

If you're interested in contributing to the maintenance of this database or making suggestions for improvement, please consider joining our mailinglist at:

<http://groups.yahoo.com/group/baseball-databank/>

If you are interested in similar databases for other sports, please vist the Open Source Sports website at <http://OpenSourceSports.com>

1.0 Release Contents

This release of the database can be downloaded in several formats. The contents of each version are listed below.

MS Access Versions: lahman2013.mdb 2013readme.txt

SQL version lahman2013.sql lahman2013_tables.sql 2013readme.txt

Comma Delimited Version: 2013readme.txt

AllStarFull.csv Appearances.csv AwardsManagers.csv AwardsPlayers.csv AwardsShareManagers.csv AwardsSharePlayers.csv Batting.csv BattingPost.csv Fielding.csv FieldingOF.csv FieldingPost.csv HallOfFame.csv Managers.csv ManagersHalf.csv Master.csv Pitching.csv PitchingPost.csv Salaries.csv Schools.csv SchoolsPlayers.csv SeriesPost.csv Teams.csv TeamsFranchises.csv TeamsHalf.csv

1.1 Introduction

This database contains pitching, hitting, and fielding statistics for Major League Baseball from 1871 through 2013. 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. (See Acknowledgements below for a list of the key contributors to this project.)

None of what we have done would have been possible without the pioneering work of Hy Turkin, S.C. Thompson, David Neft, and Pete Palmer (among others). All baseball fans owe a debt of gratitude to the people who have worked so hard to build the tremendous set of data that we have today. Our thanks also to the many members of the Society for American Baseball Research who have helped us over the years. We strongly urge you to support and join their efforts. Please vist their website (www.sabr.org).

This database can never take the place of a good reference book like The Baseball Encyclopedia. But it will enable people do to the kind of queries and analysis that those traditional sources don't allow.

If you have any problems or find any errors, please let us know. Any feedback is appreciated

1.2 What's New in 2013

Player stats have been updated through 2013 season.

There has been significant cleanup in the master file. Death dates have been updated for over 600 players who died or whose death dates became known in the last few years.

Date of last game has been added for all players who did not play during the 2013 season.

Some deprecated fields in the master table have been removed.

1.3 Acknowledgements

Much of the raw data contained in this database comes from the work of Pete Palmer, the legendary statistician, who has had a hand in most of the baseball encyclopedias published since 1974. He is largely responsible for bringing the batting, pitching, and fielding data out of the dark ages and into the computer era. Without him, none of this would be possible. For more on Pete's work, please read his own account at: <http://sabr.org/cmsfiles/PalmerDatabaseHistory.pdf>

Two people have been key contributors to the work that followed, first by taking the raw data and creating a relational database, and later by extending the database to make it more accesible to researchers.

Sean Lahman launched the Baseball Archive's website back before most people had heard of the world wide web. Frustrated by the lack of sports data available, he led the effort to build a baseball database that everyone could use. Baseball researchers everywhere owe him a debt of gratitude. Lahman served as an associate editor for three editions of Total Baseball and contributed to five editions of The ESPN Baseball Encyclopedia. He has also been active in developing databases for other sports.

The work of Sean Forman to create and maintain an online encyclopedia at "baseball-reference.com" has been remarkable. Recognized as the premier online reference source, Forman's site provides an outstanding interface to the raw data. His efforts to help streamline the database have been extremely helpful. Most importantly, Forman has spearheaded the effort to provide standards that enable several different baseball databases to be used together. He was also instrumental in launching the Baseball Databank, a forum for researchers to gather and share their work.

Since 2001, these two Seans have led a group of researchers who volunteered to maintain and update the database.

A handful of researchers have made substantial contributions to maintain this database in recent years. Listed alphabetically, they are: Derek Adair, Mike Crain, Kevin Johnson, Rod Nelson, Tom Tango, and Paul Wendt. These folks did much of the heavy lifting, and are largely responsible for the improvements made in the last decade.

Others who made important contributions include: Dvd Avins, Clifford Blau, Bill Burgess, Clem Comly, Jeff Burk, Randy Cox, Mitch Dickerman, Paul DuBois, Mike Emeigh, F.X. Flinn, Bill Hickman, Jerry Hoffman, Dan Holmes, Micke Hovmoller, Peter Kreutzer, Danile Levine, Bruce Macleod, Ken Matinale, Michael Mavrogiannis, Cliff Otto, Alberto Perdomo, Dave Quinn, John Rickert, Tom Ruane, Theron Skyles, Hans Van Slooten, Michael Westbay, and Rob Wood.

Many other people have made significant contributions to the database over the years. The contribution of Tom Ruane's effort to the overall quality of the underlying data has been tremendous. His work at retrosheet.org integrates the yearly data with the day-by-day data, creating a reference source of startling depth. It is unlikely than any individual has contributed as much to the field of baseball research in the past five years as Ruane has.

Sean Holtz helped with a major overhaul and redesign before the 2000 season. Keith Woolner was instrumental in helping turn a huge collection of stats into a relational database in the mid-1990s. Clifford Otto & Ted Nye also helped provide guidance to the early versions. Lee Sinnis, John Northey & Erik Greenwood helped supply key pieces of data. Many others have written in with corrections and suggestions that made each subsequent version even better than what preceded it.

The work of the SABR Baseball Records Committee, led by Lyle Spatz has been invaluable. So has the work of Bill Carle and the SABR Biographical Committee. David Vincent, keeper of the Home Run Log and other bits of hard to find info, has always been helpful. The recent addition of colleges to player bios is the result of much research by members of SABR's Collegiate Baseball committee.

Salary data has been supplied by Doug Pappas, who passed away during the summer of 2004. He was the leading authority on many subjects, most significantly the financial history of Major League Baseball. We are grateful that he allowed us to include some of the data he compiled. His work has been continued by the SABR Business of Baseball committee.

Thanks is also due to the staff at the National Baseball Library in Cooperstown who have been so helpful -- Tim Wiles, Jim Gates, Bruce Markusen, and the rest of the staff.

A special debt of gratitude is owed to Dave Smith and the folks at Retrosheet. There is no other group working so hard to compile and share baseball data. Their website (www.retrosheet.org) will give you a taste of the wealth of information Dave and the gang have collected.

The 2013 database benefitted from the work of Ted Turocy and his Chadwick baseball Bureau. For more details on his tools and services, visit: <http://chadwick.sourceforge.net/doc/index.html>

Thanks to all contributors great and small. What you have created is a wonderful thing.

1.4 Using this Database

This version of the database is available in Microsoft Access format, SQL files or in a generic, comma delimited format. Because this is a relational database, you will not be able to use the data in a flat-database application.

Please note that this is not a stand alone application. It requires a database application or some other application designed specifically to interact with the database.

If you are unable to import the data directly, you should download the database in the delimited text format. Then use the documentation in sections 2.1 through 2.22 of this document to import the data into your database application.

1.5 Revision History

Version	Date	Comments
1.0	December 1992	Database ported from dBase
1.1	May 1993	Becomes fully relational
1.2	July 1993	Corrections made to full database
1.21	December 1993	1993 statistics added
1.3	July 1994	Pre-1900 data added
1.31	February 1995	1994 Statistics added
1.32	August 1995	Statistics added for other leagues
1.4	September 1995	Fielding Data added
1.41	November 1995	1995 statistics added
1.42	March 1996	HOF/All-Star tables added
1.5-MS	October 1996	1st public release - MS Access format
1.5-GV	October 1996	Released generic comma-delimited files
1.6-MS	December 1996	Updated with 1996 stats, some corrections
1.61-MS	December 1996	Corrected error in MASTER table
1.62	February 1997	Corrected 1914-1915 batters data and updated
2.0	February 1998	Major Revisions-added teams & managers
2.1	October 1998	Interim release w/1998 stats
2.2	January 1999	New release w/post-season stats & awards added
3.0	November 1999	Major release - fixed errors and 1999 statistics added
4.0	May 2001	Major release - proofed & redesigned tables
4.5	March 2002	Updated with 2001 stats and added new biographical data
5.0	December 2002	Major revision - new tables and data
5.1	January 2004	Updated with 2003 data, and new pitching categories
5.2	November 2004	Updated with 2004 season statistics.
5.3	December 2005	Updated with 2005 season statistics.
5.4	December 2006	Updated with 2006 season statistics.
5.5	December 2007	Updated with 2007 season statistics.
5.6	December 2008	Updated with 2008 season statistics.
5.7	December 2009	Updated for 2009 and added several tables.
5.8	December 2010	Updated with 2010 season statistics.
5.9	December 2011	Updated for 2011 and removed obsolete tables.
2012	December 2012	Updated with 2012 season statistics
2013	December 2013	Updated with 2013 season statistics

2.0 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 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 SchoolsPlayers - list of players and the colleges they attended

Sections 2.1 through 2.24 of this document describe each of the tables in detail and the fields that each contains.

2.1 MASTER 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 G_batting Game as batter 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 G_Old Old version of games (deprecated)

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 IP 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 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 IP 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 IP 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 Whetehr team is currently active (Y or N) NAassoc ID of National Association team franchise played as

2.12 FieldingOF 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 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

2.14 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

2.15 Salaries table

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

2.16 SeriesPost 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

2.17 AwardsManagers 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

2.18 AwardsPlayers 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

2.19 AwardsShareManagers table

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

2.20 AwardsSharePlayers table

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

2.21 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)

2.22 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 Schools 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 SchoolsPlayers playerid Player ID code schoolID school ID code yearMin year player's college career started yearMax year player's college career started