Dataset is available here: <http://www.seanlahman.com/baseball-archive/statistics/>

**Dataset Schema:**

***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:

People - 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

FieldingOFsplit - LF/CF/RF splits

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

Parks - list of major league ballparls

HomeGames - Number of homegames played by each team in each ballpark

--------------------------------------------------------------------------

***People table***

playerID A unique code asssigned 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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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')

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***Salaries table***

yearID Year

teamID Team

lgID League

playerID Player ID code

salary Salary

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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 numner of points possible

votesFirst Number of first place votes

------------------------------------------------------------------------------

***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 numner of points possible

votesFirst Number of first place votes

------------------------------------------------------------------------------

***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)

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***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

------------------------------------------------------------------------------

***CollegePlaying table***

playerid Player ID code

schoolID school ID code

year year

------------------------------------------------------------------------------

***FieldingOFsplit 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

------------------------------------------------------------------------------

***Parks table***

park.key ballpark ID code

park.name name of ballpark

park.alias alternate names of ballpark

city city

state state

country country

------------------------------------------------------------------------------

***HomeGames table***

year.key year

league.key league

team.key team ID

park.key ballpark ID

span.first date of first game played

span.last date of last game played

games total number of games

openings total number of dates played

attendance total attendaance