# RETROSHEET: STEP-BY-STEP EXAMPLE

The following step-by-step example is in response to several requests for more detailed instructions on using Retrosheet data and the programs on this web site. The tasks illustrated probably will not be exactly what you want to do, but they should show you how to generate box scores and extract data that can be imported into spreadsheet and database software for further processing and analysis. You may want to print this page for future reference.

Comments, suggestions, and improvements are welcomed. You can send an e-mail using the mail link at the bottom of our home page.

Press the back button on your browser to return to the previous page when done with this one.

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

I don't think I can take the time to write a complete and fairly foolproof set of instructions that a 5th grader can follow. However, here is a list of steps to show you how to generate the Tigers box scores at Fenway in 1982.

1) If you haven't done so already, download the following files from our site. You should put them in the just one directory on your hard disk, which probably should be one set up to deal with retrosheet files. I'll use C:\retro as the example, but you can put it in any directory that suits you.

a) the program files (under Tools on the site)

bevent.zip (which expands to bevent.exe)

box.zip

bgame.zip

[Note you click on bevent.exe, box.exe, bgame.exe to do the downloads.]

If you don't know how to "unzip" files, also download the program file pkunzip.exe. It won't hurt to download it if you are not sure.

b) data files (under Game Data), choose a league & year, and then your teams or the whole league. Since your message makes it sound like you want to work with all teams, after choosing the 1982 AL, click on Entire American League. This will put the file 1982al.zip on your hard disk

2) If you did not download to the C:\retro directory, move the downloaded files to this directory. You can do this using Windows.

3) Now you will need to use DOS. From the Windows Start button in the lower left, after clicking move up to Programs and then over to MS-DOS Prompt and select it. A "DOS Window" should open on your screen with a prompt that looks like C:\WINDOWS> (it may look a little different). If it is small, you may be able to enlarge it by clicking on the box just to the left of the X in the upper right corner. This assumes Windows 95 or 98; Windows 3.1 works a bit differently, but there is a way to open a DOS window.

4) Move to the retro directory (following my example directory name) by typing "cd \retro" and pressing the Enter key. The prompt should now look like "C:\retro>".

5) The next step is to unzip the files you downloaded if you have not already done so. You can do so with the following four commands entered from the DOS prompt. Press enter after each one:

[Note: DOS commands can be fussy; be sure to leave spaces where shown and no spaces where shown. If you get a message indicating that something does not work right, then retype the command and try again.]

a) pkunzip bgame.zip

b) pkunzip bevent.zip

c) pkunzip box.zip

d) pkunzip 1982al.zip

6) Use bgame to get a list of all games played at Boston:

bgame -y 1982 82bos.eva >82bos.txt (press Enter)

The ">82bos.txt" causes the output to be written to the file 82bos.txt for further reference. You can use a different name if you prefer. Also if you end the command with ">prn" instead, the list of games will be printed on your printer. You may want to do it both ways. If your file name ends in ".txt" you can view the file by finding it a Windows file list (keep the DOS window open) by double clicking on its name or icon. At that point, another window should open up showing the contents of the file.

7) Find the games the Tigers played. You can look at the file or printout or use the search feature in the program (NotePad or WordPad) that opened the file to look for DET. I have already done it, and the Tigers played at Fenway June 21-23 (three games) and Sept 10-12 (three games). I can tell the dates by the game IDs at the left of each line: BOS198206210. The zero at the end indicates that it was a single game. Double header games end in a 1 or a 2.

8) We can make the box scores with two commands in the DOS window, one for each series:

box -y 1982 -s 0621 -e 0623 82bos.eva >box1.txt

box -y 1982 -s 0910 -e 0912 82bos.eva >box2.txt

As before, you can use >prn at the end instead to print the box scores.

To see what the -y 1982, etc. means (if you can't guess), you can enter the command box -h and it will display a brief list of the program options and command syntax. For example, if you wanted to make all the box scores for Red Sox home games, you could enter the command

box -y 1982 82bos.eva >bos82box.txt

[Note: due to how DOS works, do not use more than 8 characters before the .txt]

9) If you want to try to put together the data for how the Tigers did in Fenway, the first step would be to extract the play-by-play records for the six games. This can be done using bevent, and the syntax is similar to that for box:

bevent -y 1982 -s 0621 -e 0623 82bos.eva >detdata1.txt

bevent -y 1982 -s 0910 -e 0912 82bos.eva >detdata2.txt

That will produce two files that probably look almost nonsensical if you view them as above. However, they can be read into database or spreadsheet programs for further processing. At this point, I am not going to be able to give you directions. I will refer you to the documentation on the web site for information about the output of the bevent program in the two files, but it may not help you a whole lot. If I were going to get data on team performance in each park, I would either write a program in BASIC to pull everything together or do a fair amount of sorting and aggregating using a spreadsheet or database program.

One final note. Instead of downloading the league file, you may find it better to download the team files. The league file contains 14 files with all the home games for the AL teams, one per team. However, if you are interested in how the Tigers did in each road park, the Detroit file has all of their games, and the home and road teams are identified in the record for each game. It will be easier to work with the road games in the Detroit file than the 13 home game files for the other teams.

I hope this gets you at least part of the way to your goal. I can try to answer questions, but I am not going to be able to provide any more detailed directions like those above. As I said before, our data are not easy to use (except for making box scores) and require a fair amount of additional processing in most cases.

Good luck. I hope you have some fun doing this.

Mark

# RETROSHEET: HOW TO USE OUR EVENT FILES

The ".zip" file archives you can download contain several files that can be expanded and written to your hard disk using the program PKUNZIP.EXE, which is available for downloading under the Tools selection. Several other programs can expand and manage .zip files. The game data archives contain event files, roster files for the league, and a team names and abbreviations file. All files are standard ASCII text files.

## Event file – eight different record types

For each game there are eight different record types. Each record type has a unique designator, which is followed by several fields separated by commas. The eight record types and their possible fields are described below. There are two other record types not described here which are used to indicate a batter or pitcher batting or throwing other than his normal handedness.

### ID

Each game begins with a twelve character ID record which identifies the date, location, and number of the game. For example, BAL198304040 should be read as follows. The first three characters identify the home team (the Orioles). The next two are the last two digits of the year (1983). The next two are the month (April), the next two are the day (04). The last digit is the number of the game: here it is a 0 for a single game; it would be 1 for the first game of a double header, 2 for the second game.

### VERSION

The version statement is next and tracks when the file was created. The current version is '1'.

### INFO

There are up to 30 info records, each of which contains a single piece of information, such as the temperature, the attendance, the identity of each umpire, etc.

z

### START

There are 18 (for the NL or pre-DH AL) or 20 (for the AL with the DH) start records, which identify the starting lineups for the game. Each start record has five fields.

1. The first field is the Retrosheet ID code, which is unique for each player. This 8 digit code is constructed from the first four letters of the player's last name, the first initial of his common name, and a three digit number.
2. The second field is the player's name.
3. The next field is either 0 (for visiting team), or 1(for home team).
4. The next field is the position in the batting order.
5. The last field is the starting fielding position. The numbers are the standard notation, with designated hitters being identified as position 10.

### PLAY

The play records contain the events of the game. Each play record has 7 fields.

1. The first field is the inning.
2. The second field is either 0 (for visiting team) or 1 (for home team).
3. The third field is the Retrosheet ID code.
4. The fourth field is the count on the batter when this particular event (play) occurred. Most Retrosheet games do not have this information, and in such cases, "??" appears in this field.
5. The fifth field is of variable length and contains all pitches to this batter in this plate appearance. The standard pitches are: C for called strike, S for swinging strike, B for ball, F for foul ball. In addition, pickoff throws are indicated by the number of the base the throw went to. For example, "1" means the pitcher made a throw to first, "2" a throw to second, etc. If the base number is preceded by a "+" sign, the pickoff throw was made by the catcher. Some of the less common pitch codes are L:foul bunt, M:missed bunt,  Q:swinging strike on a pitchout, R:foul ball on a pitchout, I:intentional ball, P:pitchout, H:hit by pitch, K:strike of unknown type, U:unknown or missing pitch. Most Retrosheet games do not have pitch data and consequently this field is blank for such games.
6. There is occasionally more than one event for each plate appearance, such as stolen bases, wild pitches, and balks in which the same batter remains at the plate. On these occasions the pitch sequence is interrupted by a period, and there is another play record for the resumption of the batter's plate appearance.
7. The sixth field describes the play which occurred. This field is variable in length and has three main portions which follow the Retrosheet scoring system. The scoring procedure description also contains a diagram that explains clearly how each are of the playing field is designated. (The diagram is available under Docs, but most Retrosheet game accounts do not contain detailed location codes.)

1. The first portion is a description of the basic play, following standard baseball scoring notation. For example, a fly ball to center field is "8", a ground ball to second is "43", etc. Base hits are abbreviated with a letter (S for singles, D for doubles, T for triples, H for home runs) and (usually) a number identifying the fielder who played the ball. Therefore "S7" is a single fielded by the left fielder.
2. The second portion is a modifier of the first part and is separated from it with a forward slash, "/". In fact, there may be more than one second portion. Typical examples are hit locations. For example, "D8/78" indicates a double fielded by the center fielder on a ball hit to left center. Other possible second portion modifiers are "SH" for sacrifice hits, GDP for grounding into double plays, etc.
3. The third portion describes the advancement of any runners, separated from the earlier parts by a period. For example, "S9/L9S.2-H;1-3” should be read as: single fielded by the right fielder, line drive to short right field. The runner on 2nd scored (advanced to home), and the runner on first advanced to third. Note that any advances after the first are separated by semicolons.

### COM

This record is used for those plays which require a special comment to describe the event.

### SUB

Whenever any change in the lineup occurs, there is a substitute record. There are five fields in this type of record.

1. The first field is the Retrosheet ID code.
2. The second field is the player's name.
3. The third field is either 0 (visiting team) or 1 (home team).
4. The fourth field is batting order position of the substitute.
5. The fifth field is the position the substitute is going to play. Code "11" is for pinch hitters and "12" is for pinch runners.

Note that whenever a substitution occurs, it is preceded by a play record with the play event described as "NP", meaning No Play. The purpose of this record is to "mark the place” of the substitution for other programs.

### DATA

The several data records appear after all plays from the game and contain the number of earned runs allowed by each pitcher. Each record contains the pitcher's Project Scoresheet code and the number of earned runs he allowed.

# Utility Programs for data

There are three programs available for downloading under Tools that facilitate your analysis of the raw play by play information. These programs are called BOX, BEVENT and BGAME. Their functions are described below. It is important to note that in order to use these programs, you must have the TEAM file and the ROSTER files in the directory with the program and the event files. The name of the TEAM file includes the year, TEAM1967 for example. These files are included in the .zip archive files.

## BOX

The purpose of BOX is to generate newspaper (or Sporting News) style box scores from the event file. It is designed so that you can print box scores of specific games or all games in an event file. There are three specific ways in which BOX may be run, each of which involves invoking a different "switch". All of these programs must have the year specified after the -y switch on the command line.

1. The simplest use of BOX is to print all the games in the file. For example, the file 1967NYA.EVA contains all the Yanks home games. To print a box score of all games in this file, be sure your printer is on and ready and type:

BOX -y 1967 1967NYA.EVA>PRN

If you leave out the ">PRN" at the end of the command, then all the box scores will be displayed on your monitor in a continuous unreadable stream. As an alternative you may direct the output to a new file on your disk for later examination. In this example such a file might have the name "Yanksbox". In that case the command would be:

BOX -y 1967 67NYA.EVA>YANKSBOX

1. The second choice is to print a box score for a specific game for which you know the GameID, for example, NYA6704140, the game we considered above. In this case, prepare your printer and type:

BOX -y 1967 -i NYA6704140 67NYA.EVA>PRN

Note that you must specify both the GameID as well as the name of the file which contains the game.

When this -i switch is used, the BOX program searches the entire event file from the start for the GameID you specify. Depending on your equipment, this search may take many seconds or even a few minutes. After your requested game has been printed, the BOX program will continue to search to the end of the file, another potentially time-consuming process.

1. The third choice for using BOX is important if you don't remember a desired GameId. In this option, the BOX program searches through the file you specify, displays each GameID on the screen for you, and asks if you want that game printed. You must respond yes or no for each game. The format of this command is:

BOX -y 1967 -q 67NYA.EVA>PRN

It is important to note that in order to use BOX, you must have the TEAM file and the ROSTER files in the directory with the BOX program and the event files. These files are included in the .zip archive.

There are two utility programs which will greatly facilitate your analysis of the play-by-play data. These programs are called BEVENT and BGAME. Instructions for using both of these programs are contained in this file.

BEVENT, and BGAME have small help screens which can be displayed by typing the program name followed by a space and the characters "-h" at the DOS prompt. These help screens are useful when you cannot remember the correct syntax for each program or when you want to quickly display the switch options available for each program.

Both of these utility programs must have the year specified after the -y switch on the command line. Sample syntax for use:

BEVENT -y 1967 67BOS.EVA > 67BOS.BEV

In addition, these programs work with the standard DOS wildcards "\*" and "?" in the names of the input files.

## BEVENT

The purpose of BEVENT is to prepare a report of a game in a format that is suitable for importing into a data base or spreadsheet. It would be reasonable to think of BEVENT functioning as a parser to extract information from the event file and report the game state after each play. BEVENT functions by converting each play of a game into a record that contains several items of situational information such as number of outs, score of game, identity of pitcher, handedness of batter and pitcher, identity of any runners, etc., along with the play results. In order to run the BEVENT program, you must have the "team" and the appropriate roster files in the same directory with the BEVENT program and the event files.

There are more than 95 different fields which may be created for each BEVENT record. You have the ability to control which of these fields is created. The list of all possible fields may be obtained by typing the command "BEVENT -d". If you do not specify which fields you want BEVENT to create, it will default to the fields noted by the help screen (type "BEVENT -h"). These default fields are also noted with an asterisk in the list generated by "BEVENT -d".

There are several other switches which may be used with BEVENT. To obtain a list of these options, enter the command "BEVENT -h". The output of any BEVENT command may be directed to a printer or to a new file.

Some BEVENT fields are numbers and some are strings (that is, non-numeric characters). In all cases, the strings are enclosed by double quotes so that the records can be correctly interpreted as numbers or strings when imported into spreadsheet and data base programs. Following are descriptions of each field.

|  |  |
| --- | --- |
| Field Name | Description |
| game id | Game ID following the format described in the "data.doc" file |
| event num | All events are numbered consecutively throughout each game for easy reference |
| inning | Inning in which this play took place |
| batting team | A one-character identification of the team at bat ("0" for the visiting team and "1" for the home team) |
| outs | Number of outs before this play |
| balls, strikes, pitch sequence | These three consecutive fields present the pitch information for this play |
| vis score | Number of runs for the visiting team before this play |
| home score | Number of runs for the home team before this play |
| batter | Player ID code for the batter |
| batter hand | One character which describes how the batter batted for this event (L or R) |
| res batter and res batter hand | These fields are almost always the same as batter and batter hand. They only differ if the batter is replaced during the time at bat and the final event is charged to the previous batter. For example, if a pinch-hitter is inserted with two strikes and then takes strike three, the strikeout is charged to the first batter (the responsible batter) |
| pitcher | Player ID code for the pitcher. |
| pitcher hand | The hand with which the pitcher throws (L or R). |
| res pitcher and res pitcher hand | Counterparts to res batter and res batter hand for those occasions when a pitcher is changed during an at-bat and the first pitcher is charged with the result. For example, if a relief pitcher enters with a three-ball, no-strike count and throws ball four, then the walk is charged to the first pitcher. |
| Positions | The next eight fields contain the Player ID codes for the players at each of the eight fielding positions, in numerical sequence by position number. |
| first runner, second runner, third runner | These three consecutive fields contain the Player ID codes for the runner at each base. If a base is not occupied, then the field has no width and there will be a pair of double quotes with no space between them. For example, Bill Ripken on first as the only runner would look like this:  "ripkb001","","",  With Joe Orsulak on first and Cal Ripken on third, these fields would look like::  "orsuj001","","ripkc001" |
| event text | The complete description of the play using the format described for the event files. |
| leadoff flag | A one character descriptor which is T for the first batter of each inning and F for all others. |
| pinchhit flag | Another one character flag which is T for pinch-hitters and F for all others. |
| defensive position | The defensive position currently being played by this batter. It is pinch-hitter (position 11) for pinch-hitters. |
| lineup position | Position in the batting order for this batter. |
| event type | There are 25 different numeric codes to describe the type of event. They are:  0 Unknown event  1 No event  2 Generic out  3 Strikeout  4 Stolen base  5 Defensive indifference  6 Caught stealing  7 Pickoff error  8 Pickoff  9 Wild pitch  10 Passed ball  11 Balk  12 Other advance  13 Foul error  14 Walk  15 Intentional walk  16 Hit by pitch  17 Interference  18 Error  19 Fielder's choice  20 Single  21 Double  22 Triple  23 Home run  24 Missing play |
| batter event flag | A one character indication of whether or not the event terminated the batter's appearance. T = yes, which is most common; F = no, meaning the same batter stayed at the plate, such as after a stolen base. |
| ab flag | A one character indication of whether batter was charged with at-bat (T = yes, F = no). |
| hit value | One number indicating value of hit (0 = no hit; 1 = single; 2 = double; 3 = triple; 4 = home run). |
| SH flag | One character indicating sacrifice hit (T = yes; F = no). |
| SF flag | One character indicating sacrifice fly (T = yes; F = no). |
| outs on play | Number of outs recorded on this play. |
| double play flag | One character field of DP or not. |
| triple play flag | Once character field of TP or not. |
| RBI on play | Number of RBI credited to batter on this play. |
| wild pitch flag, passed ball flag | Two records with indication of whether there was a WP or PB on this play. |
| fielded by | Identity of the fielder who played the ball. This is especially important for base hits when no formal fielding credit is given. |
| batted ball type | Descriptor which is either F (fly ball), L (line drive), P (pop-up), or G (ground ball). |
| bunt flag | Descriptor for whether or not play was a bunt. |
| foul flag | Descriptor for whether or not ball was played in foul ground. |
| hit location | The zone on the field where the ball was hit. Refer to the Scoring System attachments for a diagram of all locations. |
| num errors | Number of errors on this play (a maximum of three is allowed). |
| error players and types | These are 6 consecutive fields which identify the player committing the 1st, 2nd or 3rd errors on the play and the type of error each was (throw or drop). |
| batter dest | The base which the batter reached at the conclusion of the play. If he was out, the base is 0. |
| runner dest | The next three fields contain the base reached by each of the three runners at the conclusion of the play. If there was no advance, then the base shown will be the one where the runner started. Note that these runner fields are not updated on plays which end an inning, even if the inning-ending play would have resulted in an advance of one or more runners had it occurred earlier in the inning. |
| Plays | The next four fields indicate the play (if any) made on the batter and each of the runners (if any). |
| SB, CS, PO flags | The next nine fields contain single character descriptors for each of the runners indicating whether he had a stolen base, was caught stealing or was picked off. |
| responsible pitcher for runner | The next three fields indicate which pitcher was responsible for the runners on each base, if any. This assignment reflects responsibility should the runner score. |
| new game and end game flags | The next two fields set a flag if this is the first record of a new game or the last record of the game. |
| Pinchrunners | The next three fields indicate if a pinchrunner has entered the game and at which base. |
| removed runners | The next three fields contain the player ID of the runner who was just run for, one field for each base. If there is no pinchrunner at that base, the field contains the NULL string "". |
| removed batter | If there is a pinchhitter, this field contains the player ID of the batter removed. If there is no pinchhitter, this field contains the NULL string "". |
| removed batter position | If there is a pinchhitter, this field contains the fielding position of the removed batter. If there is no pinchhitter, this value is 0. |
| fielder putouts | The next three fields indicate the first, second, and third fielders credited with putouts on the play. |
| fielder assists | The next five fields indicate which fielders got credited with assists on the play (maximum of five fielders). |

If you run BEVENT and generate all the possible fields, the output will be more than three times the size of the event file you specified to the program. It is strongly suggested that you generate only a subset of all possible fields at any single time, since most studies will not need all of the information at one time. BEVENT may then be run again, specifying different fields for the output for a subsequent study.

The following list presents all of the above options with the numbers to use with the -f option to specify them. Those marked with an asterisk are produced by the default option when the user specifies no fields.

|  |  |
| --- | --- |
| number | field |
| 0 | game id\* |
| 1 | visiting team\* |
| 2 | inning\* |
| 3 | batting team\* |
| 4 | outs\* |
| 5 | balls\* |
| 6 | strikes\* |
| 7 | pitch sequence |
| 8 | vis score\* |
| 9 | home score\* |
| 10 | batter |
| 11 | batter hand |
| 12 | res batter\* |
| 13 | res batter hand\* |
| 14 | pitcher |
| 15 | pitcher hand |
| 16 | res pitcher\* |
| 17 | res pitcher hand\* |
| 18 | catcher |
| 19 | first base |
| 20 | second base |
| 21 | third base |
| 22 | shortstop |
| 23 | left field |
| 24 | center field |
| 25 | right field |
| 26 | first runner\* |
| 27 | second runner\* |
| 28 | third runner\* |
| 29 | event text\* |
| 30 | leadoff flag\* |
| 31 | pinchhit flag\* |
| 32 | defensive position\* |
| 33 | lineup position\* |
| 34 | event type\* |
| 35 | batter event flag\* |
| 36 | ab flag\* |
| 37 | hit value\* |
| 38 | SH flag\* |
| 39 | SF flag\* |
| 40 | outs on play\* |
| 41 | double play flag |
| 42 | triple play flag |
| 43 | RBI on play\* |
| 44 | wild pitch flag\* |
| 45 | passed ball flag\* |
| 46 | fielded by |
| 47 | batted ball type |
| 48 | bunt flag |
| 49 | foul flag |
| 50 | hit location |
| 51 | num errors\* |
| 52 | 1st error player |
| 53 | 1st error type |
| 54 | 2nd error player |
| 55 | 2nd error type |
| 56 | 3rd error player |
| 57 | 3rd error type |
| 58 | batter dest\* (5 if scores and unearned, 6 if team unearned) |
| 59 | runner on 1st dest\* (5 if scores and unearned, 6 if team unearned) |
| 60 | runner on 2nd dest\* (5 if scores and unearned, 6 if team unearned) |
| 61 | runner on 3rd dest\* (5 if socres and uneanred, 6 if team unearned) |
| 62 | play on batter |
| 63 | play on runner on 1st |
| 64 | play on runner on 2nd |
| 65 | play on runner on 3rd |
| 66 | SB for runner on 1st flag |
| 67 | SB for runner on 2nd flag |
| 68 | SB for runner on 3rd flag |
| 69 | CS for runner on 1st flag |
| 70 | CS for runner on 2nd flag |
| 71 | CS for runner on 3rd flag |
| 72 | PO for runner on 1st flag |
| 73 | PO for runner on 2nd flag |
| 74 | PO for runner on 3rd flag |
| 75 | Responsible pitcher for runner on 1st |
| 76 | Responsible pitcher for runner on 2nd |
| 77 | Responsible pitcher for runner on 3rd |
| 78 | New Game Flag |
| 79 | End Game Flag |
| 80 | Pinch-runner on 1st |
| 81 | Pinch-runner on 2nd |
| 82 | Pinch-runner on 3rd |
| 83 | Runner removed for pinch-runner on 1st |
| 84 | Runner removed for pinch-runner on 2nd |
| 85 | Runner removed for pinch-runner on 3rd |
| 86 | Batter removed for pinch-hitter |
| 87 | Position of batter removed for pinch-hitter |
| 88 | Fielder with First Putout (0 if none) |
| 89 | Fielder with Second Putout (0 if none) |
| 90 | Fielder with Third Putout (0 if none) |
| 91 | Fielder with First Assist (0 if none) |
| 92 | Fielder with Second Assist (0 if none) |
| 93 | Fielder with Third Assist (0 if none) |
| 94 | Fielder with Fourth Assist (0 if none) |
| 95 | Fielder with Fifth Assist (0 if none) |
| 96 | event num |

\*\*\*IMPORTANT: If you run BEVENT and generate all the possible fields, the output will be more than three times the size of the event file you specified to the program. It is suggested that you generate only a subset of all possible fields at any single time, since most studies will not need all of the information at one time. BEVENT may then be run again, specifying different fields for the output, for a subsequent study.

## BGAME

The purpose of this program is to generate a summary of those items which are constant for each game, such as date, indication of day or night, identity of umpires, etc. There is also summary data of the game, such as the final score.

The output of BGAME may be coordinated with that from BEVENT since they use the same Game ID code. The use of this program is similar to BEVENT in that there are several switches to create the precise output desired. To see the options enter the command "BGAME -h"; for more general instructions on how to use BGAME to analyze play-by-play data, see the instructions for BEVENT.

The following list presents all of the options for BGAME with their numerical identification. Note that the default for BGAME is that all fields are produced. The -f option similar to that in BEVENT can be used to output selected fields.

|  |  |
| --- | --- |
| number | field |
| 0 | game id |
| 1 | date |
| 2 | game number (0 = no double header) |
| 3 | day of week |
| 4 | start time |
| 5 | DH used flag |
| 6 | day/night flag |
| 7 | visiting team |
| 8 | home team |
| 9 | game site |
| 10 | vis. starting pitcher |
| 11 | home starting pitcher |
| 12 | home plate umpire |
| 13 | first base umpire |
| 14 | second base umpire |
| 15 | third base umpire |
| 16 | left field umpire |
| 17 | right field umpire |
| 18 | attendance |
| 19 | PS scorer |
| 20 | translator |
| 21 | inputter |
| 22 | input time |
| 23 | edit time |
| 24 | how scored |
| 25 | pitches entered? |
| 26 | temperature |
| 27 | wind direction |
| 28 | wind speed |
| 29 | field condition |
| 30 | precipitation |
| 31 | sky |
| 32 | time of game |
| 33 | number of innings |
| 34 | visitor final score |
| 35 | home final score |
| 36 | visitor hits |
| 37 | home hits |
| 38 | visitor errors |
| 39 | home errors |
| 40 | visitor left on base |
| 41 | home left on base |
| 42 | winning pitcher |
| 43 | losing pitcher |
| 44 | save for |
| 45 | GW RBI |
| 46 | visitor batter 1 |
| 47 | visitor position 1 |
| 48 | visitor batter 2 |
| 49 | visitor position 2 |
| 50 | visitor batter 3 |
| 51 | visitor position 3 |
| 52 | visitor batter 4 |
| 53 | visitor position 4 |
| 54 | visitor batter 5 |
| 55 | visitor position 5 |
| 56 | visitor batter 6 |
| 57 | visitor position 6 |
| 58 | visitor batter 7 |
| 59 | visitor position 7 |
| 60 | visitor batter 8 |
| 61 | visitor position 8 |
| 62 | visitor batter 9 |
| 63 | visitor position 9 |
| 64 | home batter 1 |
| 65 | home position 1 |
| 66 | home batter 2 |
| 67 | home position 2 |
| 68 | home batter 3 |
| 69 | home position 3 |
| 70 | home batter 4 |
| 71 | home position 4 |
| 72 | home batter 5 |
| 73 | home position 5 |
| 74 | home batter 6 |
| 75 | home position 6 |
| 76 | home batter 7 |
| 77 | home position 7 |
| 78 | home batter 8 |
| 79 | home position 8 |
| 80 | home batter 9 |
| 81 | home position 9 |

# The Event File – detailed description and the scoring system

The event files contain game descriptions using the Retrosheet scoring system. This page will describe the scoring system in sufficient detail to allow working with these full play-by-play descriptions.

The files containing the play-by-play data follow a naming convention. Each file has one team's home games and has a name of the form YYYYTTT.EVX. In this format, YYYY is the four digit year and TTT is a three character team code. The zip archive downloaded contains a file named TEAMYYYY that contains the team codes and team names in the particular season. Each file contains the home games in chronological order for the specified team.

Files are ASCII text files consisting of a series of records. Each record is a single line starting with a type designator and ending with the DOS new line sequence (newline, carriage return characters).

For each game as many as eleven different record types may be used. Each record type has a unique designator, which is followed by several fields separated by commas. These are discussed in detail below.

The record type is not considered to be a field and starts in column 1. Following the record type are the record fields which are separated from the record type and each other by commas ‘, '.

Field data such as names are normally enclosed in double quotes ' " '. Commas used in quoted fields are not field separators.

### Retrosheet player id

All players are represented by a code that is unique for each player. This 8 character code is constructed from the first four letters of the player's last name, the first initial of his common name, and a three digit number. If a player's last name is less than 4 characters long a dash "-" is used as a placeholder. Numbers starting with 0 are used for players appearing in games in or after 1983. Players completing their careers before 1983 are assigned numbers starting with 100.

joner002 is the Retrosheet player id for Ruppert Jones.

### ID

Each game begins with a twelve character ID record which identifies the date, home team, and number of the game. For example, ATL198304080 should be read as follows. The first three characters identify the home team (the Braves). The next four are the year (1983). The next two are the month (April) using the standard numeric notation, 04, followed by the day (08). The last digit indicates if this is a single game (0), first game (1) or second game (2) if more than one game is played during a day, usually a double header The *id* record starts the description of a game thus ending the description of the preceding game in the file.

id,ATL198304080

### VERSION

The version record is next, but is obsolete and can be ignored.

version,1

### INFO

There are up to 34 info records, each of which contains a single piece of information, such as the temperature, attendance, identity of each umpire, etc. The record format is info, type, data. The complete list of [info record types](http://www.retrosheet.org/eventfile.htm#1) is given below.

info, attendance,32737

### START and SUB

There are 18 (for the NL and pre-DH AL) or 20 (for the AL with the DH) start records, which identify the starting lineups for the game. Each start or sub record has five fields. The sub records are used when a player is replaced during a game. The roster files that accompany the event files include throwing and batting handedness information.

1. The first field is the [Retrosheet player id](http://www.retrosheet.org/eventfile.htm#2), which is unique for each player.
2. The second field is the player's name.
3. The next field is either 0 (for visiting team), or 1 (for home team).
4. The next field is the position in the batting order, 1 - 9. When a game is played using the DH rule the pitcher is given the batting order position 0.
5. The last field is the fielding position. The numbers are in the standard notation, with designated hitters being identified as position 10. On sub records 11 indicates a pinch hitter and 12 is used for a pinch runner.

start,richg001,"Gene Richards",0,1,7

### PLAY

The play records contain the events of the game. Each play record has 7 fields.

1. The first field is the inning, an integer starting at 1.
2. The second field is either 0 (for visiting team) or 1 (for home team).
3. The third field is the [Retrosheet player id](http://www.retrosheet.org/eventfile.htm#2) of the player at the plate.
4. The fourth field is the count on the batter when this particular event (play) occurred. Most Retrosheet games do not have this information, and in such cases, "??" appears in this field.
5. The fifth field is of variable length and contains all pitches to this batter in this plate appearance and is [described below](http://www.retrosheet.org/eventfile.htm#3). If pitches are unknown, this field is left empty, nothing is between the commas.
6. The sixth field describes the play or event that occurred.

play,5,1,ramir001,00,,S8.3-H;1-2

A play record ending in a number sign, #, indicates that there is some uncertainty in the play.

Occasionally, a com record may follow providing additional information.

A play record may also contain exclamation points, "!" indicating an exceptional play and question marks "?" indicating some uncertainty in the play.

These characters can be safely ignored.

play,3,1,kearb001,??,,PB.2-3#  
com,"Not sure if PB, may have been balk"

The event is the most complex of all the fields and is [described in detail below](http://www.retrosheet.org/eventfile.htm#5).

### BADJ

This record is used to mark a plate appearance in which the batter bats from the side that is not expected ("badj" means "batting adjustment"). The syntax is:

badj,playerid,hand

The expectation is defined by the roster file. There are two general cases in which this is used:

1. Many switch-hitters bat right-handed against right-handed knuckle ball pitchers even though the default assumption is that they would be batting left-handed.

badj,bonib001,R

indicates that switch-hitter Bobby Bonilla was batting right-handed against a right-handed pitcher.

1. Occasionally a player will be listed in a roster as batting "R" or "L" but will bat the other way. For example, Rick Dempsey did this 13 times in 1983.

The syntax this is:

badj,dempr101,L

### PADJ

This record covers the very rare case in which a pitcher pitches to a batter with the hand opposite the one listed in the roster file. To date this has only happened once, when Greg Harris of the Expos, a right-hander, pitched left-handed to two Cincinnati batters on 9-28-1995.

The syntax is parallel to that for the badj record:

padj,harrg001,L

### LADJ

This record is used when teams bat out of order.

### DATA

Data records appear after all play records from the game.

At present, the only data type, field 2, that is defined specifies the number of earned runs allowed by a pitcher.

Each such record contains the pitcher's [Retrosheet player id](http://www.retrosheet.org/eventfile.htm#2) and the number of earned runs he allowed. There is a data record for each pitcher that appeared in the game.

data,er,showe001,2

### COM

The final record type is used primarily to add explanatory information for a play.

However, it may occur anywhere in a file. The second field of the com record is quoted.

com,"ML debut for Behenna"

There is a standard record ordering for each game.

1. An *id* record starts the description of a particular game.
2. This is followed by the *version* and *info* records.
3. The *start* records follow the *info* records.
4. The game is described by a series of *play*, *sub* and *com* records.
5. A *sub* record is always preceded by a *play np* record.
6. *data* records follow the last *play* record for the game.
7. A game description is terminated by an *id* record starting another game or the end of the file.

## 

## Info Record types

Complete records are shown. *info* records are of two general kinds, game-related and administrative. The order of these records, which appear after the game id, may not be in the order shown below. Game-related *info* records are:

The home and visiting teams are specified by their [Retrosheet team codes](http://www.retrosheet.org/team_codes.html).

info,visteam,SDN  
info,hometeam,ATL

The date is given in conventional yyyy/mm/dd style:

info,date,1983/04/08

The number record indicates if this is a single game (0), first game (1) or second game (2) if more than one game is played during a day, usually this is a double header:

info,number,0

The hometeam, date and number records duplicate the information in the *id* record.

Game starting time is given by the two records (0:00 and unknown indicate missing information):

info,starttime,7:44PM  
info,daynight,night

Use of the designated hitter is indicated with true or false:

info,usedh,false

The presence or absence of pitch information is given. For some games, the bal-strike counts of the plays are shown, but no pitch detail is provided. (pitches, count or none):

info,pitches,pitches

Each umpire and his position on the field are indicated individually by his Retrosheet ID. For games where umpires are stationed in the outfield, umplf and umprf are used. Retrosheet has umpire assignments for all games in history, except some games in 1979 in which replacement umpires were used.

info,umphome,quicj901  
info,ump1b,palld901  
info,ump2b,engeb901  
info,ump3b,rungp901

Various field conditions are given:

info,fieldcond,unknown  
info,precip,unknown  
info,sky,night  
info,temp,69  
info,winddir,unknown  
info,windspeed,-1

Values used -

For fieldcond are: dry, soaked, wet, unknown;

For precip: drizzle, none, rain, showers, snow, unknown;

For sky: cloudy, dome, night, overcast, sunny, uknown;

For winddir: fromcf, fromlf, fromrf, ltor, rtol, tocf, tolf, torf, unknown.

Temp(erature) is in degrees Fahrenheit with 0 being the not known value.

An unknown windspeed is indicated by -1.

The BGAME.EXE program outputs these fields using numeric codes:

FieldCond: 0 Unknown, 1 Soaked, 2 Wet, 3 Damp, 4 Dry  
Precip: 0 Unknown, 1 None, 2 Drizzle, 3 Showers, 4 Rain, 5 Snow  
Sky: 0 Unknown, 1 Sunny, 2 Cloudy, 3 Overcast, 4 Night, 5 Dome  
WindDir: 0 Unknown, 1 ToLeft, 2 ToCenter, 3 ToRight, 4 LeftToRight, 5 FromLeft, 6 FromCenter, 7 FromRight, 8 RightToLeft  
WindSpeed: 0 Unknown, 1 Known, other value is the wind speed

The length of the game in minutes and the attendance (0 used if these are not known) are given:

info,timeofgame,134  
info,attendance,10356

The game site is provided. The site symbols are defined in the file [parkcode.txt:](http://www.retrosheet.org/parkcode.txt)

info,site,SFO02

Pitcher win, loss and save data are given as info records. The Retrosheet player id is used for identification. If no save is credited, the player id field is empty.

info,wp,beher001  
info,lp,sotom001  
info,save,forst001

When it was used as an official statistic, game winning RBI credit is given:

info,gwrbi,chamc001

If this information is unknown or a gwrbi was not credited, the data field is left empty.

info records that pertain to how the game account was obtained and processed (administrative data) are:

info,edittime,2000/03/31 11:00AM  
info,howscored,park  
info,inputprogvers,"version 7RS(19) of 07/07/92"  
info,inputter,"C. Chestnut"  
info,inputtime,1995/02/07 9:01PM  
info,scorer,"Braves"  
info,translator,"C. Chestnut"

## Play Record – fifth field - the pitches field of the play record

synopsis: play,inning,home/visitor,player id,count,pitches,event

The fifth field, pitches, is a string of variable length and contains all pitches to this batter in this plate appearance. Most Retrosheet games do not have pitch data and consequently this field is blank for such games.

|  |  |
| --- | --- |
| + | following pickoff throw by the catcher |
| \* | indicates the following pitch was blocked by the catcher |
| . | marker for play not involving the batter |
| 1 | pickoff throw to first |
| 2 | pickoff throw to second |
| 3 | pickoff throw to third |
| > | Indicates a runner going on the pitch |
|  |  |
| B | ball |
| C | called strike |
| F | foul |
| H | hit batter |
| I | intentional ball |
| K | strike (unknown type) |
| L | foul bunt |
| M | missed bunt attempt |
| N | no pitch (on balks and interference calls) |
| O | foul tip on bunt |
| P | pitchout |
| Q | swinging on pitchout |
| R | foul ball on pitchout |
| S | swinging strike |
| T | foul tip |
| U | unknown or missed pitch |
| V | called ball because pitcher went to his mouth |
| X | ball put into play by batter |
| Y | ball put into play on pitchout |

## Play Record – sixth field - the event field of the play record

The sixth field, event, describes the play which occurred. This field is variable in length and has three main portions which define the Retrosheet scoring system.

The first part of an event is a description of the basic play.

The second part is a modifier for the first part and is separated from it with a forward slash, "/". In fact, there may be more than one modifier. A typical use of modifiers is to specify [hit locations](http://www.retrosheet.org/location.htm). For example, "D8/78" indicates a double fielded by the center fielder on a ball hit to left center. A complete list of modifiers excepting hit locations [is given below](http://www.retrosheet.org/eventfile.htm#6). When more than one modifier is used, each is introduced by a "/".

The third part describes the advance of any runners, separated from the earlier parts by a period. A successful advance is indicated by a dash, "-". An out made while advancing is indicated by an X. 2-3 indicates a runner has advanced from second to third on the play. 1X2 indicates the runner was out at second advancing from first. If a base runner is not listed as advancing he remains on the base he was on. In some cases lack of advance is indicated explicitly by an advance starting and ending on the same base such as 3-3 . When put outs are made on base runners the advance field indicates fielding data and errors if they occur. See below for a [complete description for advances](http://www.retrosheet.org/eventfile.htm#4). Note that any advances after the first are separated by semicolons.

For example, the event "S9/L9S.2-H;1-3" should be read as: single fielded by the right fielder, line drive to short right field. The runner on 2nd scored (advanced to home), and the runner on first advanced to third.

Many event descriptions require information in the form of numbers. The meaning of a particular number depends on where it appears in the event. For the descriptions that follow the following notation will be used:

Fielders will be represented by a number in the range 1 (pitcher) to 9 (right fielder) using a dollar sign, "$". When two $ symbols are used, $$, this is understood to mean a sequence of two or more fielders.

Bases are represented by a percent sign, "%", representing one of five characters, 1, 2 and 3 for first through third; B or H for home. B is used when a batter advance must be explicitly given. Scoring is indicated by an advance that reaches home, H.

Many examples of plays scored using the Retrosheet system will be given in this document. For some interesting and extreme cases check the Retrosheet [strange and unusual plays](http://www.retrosheet.org/strange.htm) listing.

The example plays have been chosen to illustrate how events are coded. Some of these events are exceedingly rare.

There is occasionally more than one event for each plate appearance, such as stolen bases, wild pitches, and balks in which the same batter remains at the plate. On these occasions the pitch sequence is interrupted by a period, and there is another play record for the resumption of the batter's plate appearance.

For purposes of description, it is convenient to separate the event types into two categories: those involving the [batter at the plate](http://www.retrosheet.org/eventfile.htm#8) and [base running](http://www.retrosheet.org/eventfile.htm#9) plays that do not involve the batter.

## Play Record - events made by the batter at the plate

**$** A single fielder represents a fly ball out made by the specified fielder. Modifiers can be added to indicate the fly ball trajectory: G for ground ball, L for line drive, P for pop up, F for a fly ball, BG for bunt grounder, BP for bunt pop up. The ball trajectory code may be followed by a hit location code.

play,7,0,saboc001,01,CX,8/F78

01 – pitch count

C – called strike

X - ball put into play by hitter

8 – caught by center fielder

F - fly

7 – left field

8 – center field

Indicates a fly ball caught by the center fielder in left center field.

play,5,0,grifk001,10,.BX,9/SF.3-H

10 – pitch count

. – play not involving the batter

B - ball

X - ball put into play by hitter

9 – caught by right fielder

SF - sacrifice fly

3-H – runner advances from third to home

A sacrifice fly is indicated by the modifier SF following a fly out play.

The runner scoring because of the sacrifice is coded in the advance part of the play.

play,5,0,duncm001,00,X,3/G.2-3

00 – pitch count

X - ball put into play by hitter

3 – caught by first baseman unassisted

G - ground ball

2-3 – runner advances from second to third

In the case that a fielder makes an unassisted out on a ground ball a modifier G follows the event.

Indicates an unassisted out made by the first baseman with the runner on second advancing to third.

**$$** Strings of two or more fielders as an event specify a ground out where the put out is credited by the last fielder in the string. Other fielders are credited with assists.

play,6,0,davie001,01,FX,63/G6M

01 – pitch count

F - foul

X - ball put into play by hitter

6 – short stop fielder

3 – first baseman makes the out

G - ground ball

6 – short stop assist

M - ??? assist?

indicates a ground ball out at first on a ball fielded by the shortstop.

play,9,1,pendt001,00,X,143/G1

00 – pitch count

X – ball put into play by hitter

1 – pitcher deflects ball

4 – second base catches ball and throws to first

3 – first base makes the out

G – ground ball

1 – ground ball to pitcher

More than one player can touch the ball before an out is made. In this case, the pitcher has deflected the ball before the second baseman threw to first base.

play,7,1,tempg001,00,X,54(B)/BG25/**SH**.1-2

00 – pitch count

X – ball put into play by hitter

5 – third base throws to second

4 – second base makes the out

(B) – batter out at first by second baseman

BG – ground ball bunt

2 – ground ball bunt between catcher

5 – ground ball bund between catcher and third base

SH – sacrifice hit (bunt)

1-2 – runner advances from first to second

If the putout is made at a base not normally covered by the fielder the base runner, batter in this example, is given explicitly.

Force outs are indicated by adding the FO modifier and indicating the base runner forced.

play,5,0,gileb001,10,BX,54(1)/**FO**/G5.3-H;B-1

10 – pitch count

B – ball

X – ball put into play by hitter

5 – third base

4 – second base

1. – first base runner out at second on throw from third to second

FO – Force out

G – ground ball

5 – ground ball to third base

3-H – runner advances from third to home

B-1 – batter safe at first

The runner on first is forced at second by a throw from the third baseman.

The runner on third scores and the batter is safe at first.

The explicit advance indicated for the batter is optional.

A second modifier is used to indicate the batted ball trajectory and location.

play,6,1,camik001,00,X,23/**SH**.1-2

00 – pitch count

X – ball put into play by hitter

2 - catcher

3 – first base

SH – sacrifice hit (bunt)

1 – 2 runner advances from first to second

With the addition of a SH modifier this form is used to indicate sacrifice hits or bunts that advance a runner.

**$(%)$ $$(%)$** Events of this form are used to code grounded into double plays.

play,7,0,backw001,11,FBX,64(1)3/GDP/G6  
indicates a grounded into double play. The parenthesized 1 indicates the base runner on first was the initial out on the play. The GDP modifier is followed by a another / and a hit type and location.

play,8,1,smito001,22,BFCBX,4(1)3/G4/GDP  
An unassisted ground ball out by the second baseman starts this double play.

**$(B)$(%)** followed by the modifier LDP is used to indicate a lined into double play.

play,7,0,leonj001,01,CX,8(B)84(2)/LDP/L8  
indicates a fly ball out to the center fielder with the runner on second doubled up.

play,7,0,fernt001,10,BX,3(B)3(1)/LDP  
indicates an unassisted double play by the first baseman who fielded the line drive and caught the runner off first base.

The double play notation can be extended in obvious ways to describe triple plays.

play,7,1,randw001,00,.>X,1(B)16(2)63(1)/LTP/L1

**Note:** the double digit combination 99, which cannot arise in play, is used to code unknown plays including forms that otherwise describe force outs and the double plays. Additional fielders in the double play are assigned 9. No assist or putout credits are given.  
  
**C/E2**codes catcher interference. Implicitly, the batter is awarded first unless overridden by an advance indicating otherwise. A redundant B-1 is allowed.

play,9,1,cruzj002,??,,C/E2.1-2

**C/E1** or **C/E3** are used when the pitcher or first baseman are called for interfering with the batter putting him on first without being charged with an at bat. In these cases C is interpreted as interference by the fielder specified following the E, not the catcher.

**S$** single  
**D$** double  
**T$** triple  
A hit (excepting a home run) is indicated by one of S, D and T optionally followed by the fielder, $, initially handling the ball. If more than one fielder handles the ball the appropriate sequence of fielders is given. The fielder designation is omitted if that information is not known. The batter advance to the designated base is implicit.

play,8,0,pacit001,??,,S7  
is a minimal coding of a single showing that the left fielder first handled the ball. The ?? in the count field indicates the count at the time of the hit is unknown.

play,2,1,santn001,12,CFBX,D7/G5.3-H;2-H;1-H  
codes a bases loaded double fielded by the left fielder, a modifier showing the hit location code and advances for each of the base runners.

play,3,0,raint001,11,CBX,T9/F9LD.2-H  
describes a triple to right field, a hit location and a runner on second scoring.

**DGR** is the code for a ground rule double. No fielding player is specified.

play,3,0,surhb001,10,.BX,DGR/L9LS.2-H

**E$** is the code for an error allowing a batter to get on base. The fielder making the error is given by $. The batter advance to first is implicit but may be given explicitly.

play,2,0,ruffb001,10,BX,E1/TH/BG15.1-3  
indicates a throwing error (modifier "/TH") error on the pitcher with the runner on first advancing to third. The batter advance to first is implicit.

play,5,1,young001,00,X,E3.1-2;B-1  
indicates a fielding error by the first baseman. In this case the batter advance to first has been explicitly given.

**FC$** Fielder's choice. $ is the fielder first fielding the ball. The batter advance to first is understood if it is not given explicitly.

play,4,0,harpb001,22,BBFSFX,FC5/G5.3XH(52)  
The first baseman fielded the ball and threw home in time to retire the runner attempting to score. The batter was safe at first.

play,5,1,jordr001,00,X,FC3/G3S.3-H;1-2  
The first baseman fielded the ball and attempted to throw an unspecified runner out. No outs were made and the batter is safe at first.

Note that even though force outs are considered fielder's choices, the notation distinguishes between force outs and non-forced fielder's choices.

**FLE$** Error on foul fly ball.

play,5,0,murre001,00,F,FLE5/P5F

**H** or **HR** is the code for a home run leaving the park. The location modifier can be used to indicate where the ball left the playing field.

play,8,0,bellg001,21,CBBX,H/L7D  
indicates a solo home run into left field.

play,12,1,bichd001,02,FFFX,HR/F78XD.2-H;1-H  
shows a home run into center field with the runners on first and second scoring.

**H$** or **HR$** indicates an inside-the-park home run by giving a fielder as part of the code.

play,4,0,younr001,32,FBFFFBBX,HR9/F9LS.3-H;1-H

**HP** Batter hit by a pitch. The batter advance to first is implicit. Other advances are given if needed.

play,1,1,lansc001,00,H,HP.1-2

**K** Strike out

play,1,1,steit001,12,C2FBS,K

play,6,1,wynnm001,22,..BBFCFS,K23  
A dropped third strike with a putout at first base is given by the event K23.

**K+event** On third strikes various base running play may also occur. The event can be SB%, CS%, OA, PO%, PB, WP and E$.

play,2,0,roomr001,12,1BF1S11S,K+PB.1-2  
A passed ball on strike three allowed the runner on first to go to second.

play,5,1,whitd001,02,FLFFS,K+WP.B-1  
An explicit batter advance is given when he reaches first on a third strike miscue. An [alternative notation](http://www.retrosheet.org/eventfile.htm#7) for WP and PB is given below.

play,8,1,davic001,12,CFB.S,K23+WP.2-3  
Of course, a base running event can occur when the third strike is dropped.

**NP** no play. This event is used as a marker when substitutions are made.

play,8,0,puckk001,00,,NP  
sub,kutcr001,"Randy Kutcher",1,5,8

**I** or **IW** intentional walk  
**W** walk. In both cases base runner advances are given if needed. The batter advance to first base is implicit.

play,6,1,ripkc001,32,CFBBFB>B,W.1-2

play,8,0,sciom001,30,B+22.III,IW

**W+event**, **IW+event**. On ball four various base running plays may also occur. The event can be SB%, CS%, PO%, PB, WP and E$.

play,1,1,sandr001,32,C1FBB.BFB,W+WP.2-3  
The fourth ball was a wild pitch allowing the runner on second to advance.

## Play Record - base-running events not involving the batter

The player specified in these plays is the batter at the plate, not the base runner or runners affected by the play.

The play pitches and count fields (if given) are for the batter at the time of the event. Unless the event is a inning or game ending out it will be followed by another event listing the batter.

**BK** indicates a balk.

play,6,0,niekp001,??,,BK.3-H;1-2

**CS%($$)** is the event code for caught stealing. The bases, %, for this play are 2,3 and H. The fielding data, $$, is considered part of the play. Other advances may be given.

play,5,1,ceror001,??,,CSH(12)

play,1,0,bayld001,??,,CS2(24).2-3

play,6,0,beneb001,??,,CS2(2E4).1-3  
The error negates the out with the advance field indicating a two base advance on the play.

**DI** is the defensive indifference code and is given when there is no attempt to prevent a stolen base. The advance field specifies which base the runner went to.

play,9,0,bencj101,??,,DI.1-2

**OA** is coded for a base runner advance that is not covered by one of the other codes. A comment may be given explaining the advance.

play,3,1,parkr001,??,,OA.2X3(25)  
com,"Thompson out trying to advance after ball eluded catcher"

**PB** passed ball  
**WP** wild pitch. In both cases the catcher is unable to handle a pitch and a base runner advances.

play,1,1,jackb001,12,FBSFFB,WP.2-3;1-2

play,1,1,evand002,01,CB,PB.2-3

**PO%($$)** picked off of base % (1, 2 or 3) with the ($$) indicating the throw(s) and fielder making the putout.

play,4,0,guerp001,00,22,PO2(14)  
indicates the runner on second was out by a pick off throw from the pitcher to second baseman.

play,1,1,wallt001,10,B11,PO1(E3).1-2  
shows an attempt at a pick off at first with the first baseman committing an error that allows the runner to advance to second. The presence of the error (E3) negates the out normally associated with the pickoff play.

**POCS%($$)** picked off off base % (1, 2 or 3) with the runner charged with a caught stealing. The ($$) is the sequence of throws resulting in the out.

play,6,1,javis001,10,B1,POCS2(1361)

**SB%** is the event code for a stolen base. The bases, %, for this play are 2,3 and H.

play,6,0,benzt001,11,BSB,SB2

play,4,1,waltj001,10,BB,SB3;SB2  
play,4,1,shefg001,12,SP1CB,SBH;SB2  
show double steals, second and third in one case, second and home in the other.

## Play Record - play modifiers and explanations

Each modifier is preceded by / in a play record. As always, % indicates one the four bases and $ indicates a fielder.

|  |  |
| --- | --- |
| AP | appeal play |
| BP | pop up bunt |
| BG | ground ball bunt |
| BGDP | bunt grounded into double play |
| BINT | batter interference |
| BL | line drive bunt |
| BOOT | batting out of turn |
| BP | bunt pop up |
| BPDP | bunt popped into double play |
| BR | runner hit by batted ball |
| C | called third strike |
| COUB | courtesy batter |
| COUF | courtesy fielder |
| COUR | courtesy runner |
| DP | unspecified double play |
| E$ | error on $ |
| F | fly |
| FDP | fly ball double play |
| FINT | fan interference |
| FL | foul |
| FO | force out |
| G | ground ball |
| GDP | ground ball double play |
| GTP | ground ball triple play |
| IF | infield fly rule |
| INT | interference |
| IPHR | inside the park home run |
| L | line drive |
| LDP | lined into double play |
| LTP | lined into triple play |
| MREV | manager challenge of call on the field |
| NDP | no double play credited for this play |
| OBS | obstruction (fielder obstructing a runner) |
| P | pop fly |
| PASS | a runner passed another runner and was called out |
| R$ | relay throw from the initial fielder to $ with no out made |
| RINT | runner interference |
| SF | sacrifice fly |
| SH | sacrifice hit (bunt) |
| TH | throw |
| TH% | throw to base % |
| TP | unspecified triple play |
| UINT | umpire interference |
| UREV | umpire review of call on the field |

## Play Record - event advances

In addition to base runner movements, the advance portion of an event indicates fielding, errors and has the indicators indicating if a run is unearned and if an RBI is or is not credited.

Bases are represented by one of five characters, 1 for first, 2, 3 and B or H for home. B is used when a batter advance must be explicitly given. Scoring is indicated by a successful advance that reaches home, H.

Separate advances are given for each runner on base and are separated by a semicolon, ";". When more than one runner advance is given for a play they are ordered starting with the runner on third base and ending with the batter.

Advances may include additional information in the form of one or more parameters specified as a parenthesized strings of characters. When more than one parameter is given on an advance they are individually parenthesized.

A successful advance is given in the form 1-2. The dash "-" indicates a successful advance. Multiple base advances are indicated with the same notation: B-2, 1-3, 1-H, 2-H.

play,1,1,marte001,32,BBCBFFB,W.2-3;1-2

play,3,1,stilk001,11,CBX,S7/F7S.2-H;B-2

A runner put out at a particular base is indicated by the "X": 2X3, 1XH. When a runner is out the advance gives the fielding information as a parameter specifying the fielders. The last fielder gets credit for the put out and the others get assists.

play,4,1,stubf001,32,CBFBBFFS,K/DP.1X2(26)

play,6,0,murre001,22,BSFFBX,9/F9LS/FDP.3XH(92)

play,4,0,blauj001,01,CX,S8/L78.BX2(8434)

Fielding errors are indicated by including an E in the parameter following an advance. The fielder following the E is charged with the error.

play,3,0,fielc001,00,X,S7/L7LD.3-H;2-H;BX2(7E4)

Following a second baseman error the batter is safe at second. The error indicator negates the out. The left fielder is credited with an assist.

play,7,0,puckk001,01,CX,S5/G5.1-3(E5/TH)

The parameter in this play attributes a throwing error to the third baseman. A base indicator may follow TH, TH2 for example.

Parameters are used to indicate if a run is unearned (UR) and if RBI is to be credited (RBI) or not (NR), (NORBI). When these parameters are not present, normal rules are followed.

play,9,0,davie001,30,BBBB,W+PB.3-H(NR);1-3

The run scored on the passed ball is not credited as an RBI to the batter.

play,8,1,sax-s001,22,BCFBFX,S4/G34.2-H(E4/TH)(UR)(NR);1-3;B-2

Three parameters are given on the 2-H advance. The first indicates a second baseman throwing error, the second indicates it is an unearned run and the third indicates no RBI.

play,2,1,willk001,11,BFX,E6/G6.3-H(RBI);2-3;B-1

In this play an RBI is given to the batter.  Interference can be indicated with an advance parameter. An alternative way of writing this is (5/INT).

play,2,0,stanp001,12,CCBX,S/L9S.3-H;2X3(5/INT);1-2  
com,"$Gonzalez out for grabbing coach on way back to 3B"

Team unearned runs are indicated by TUR in cases with more than one pitcher in the inning and the current pitcher is to be charged with an earned run.

play,5,1,ashba001,??,,S9.3-H(TUR);2-H(TUR);1-3;BX2(93)

A U appearing in a fielding sequence indicates the fielder handling the ball is unknown.

play,7,0,perrg001,21,B.BFX,S8.2-H;BX2(8U3)

In the 8U3 sequence most likely the U is the shortstop or second baseman.

Advance parameters provide an alternative way of indicating wild pitches and passed balls.

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Ladj Record – teams bat out of order

This record is used when teams bat out of order. The normal assumption is that proper lineup sequence is followed, therefore, it is necessary to have some special indication when this is violated.

The format is:

ladj,hv,pos

where "hv" is 0 for visiting or 1 for the home team and "pos" is 1-9 for the batting order position. Retrosheet has discovered quite a few cases of batting out of turn. You can see them in the Special Lists section: [Batting Out of Turn](http://www.retrosheet.org/outturn.htm).

Here are some examples.

play,2,1,hortw101,??,,63

ladj,1,7

play,2,1,simpj101,??,,D7/BOOT

ladj,1,6

play,2,1,steib001,??,,HP/BOOT

ladj,1,8

play,2,1,cox-l101,??,,S9/BOOT.2-3;1-2

play,2,1,mendm101,??,,NP

sub,robel001,"Leon Roberts",1,9,11

play,2,1,robel001,??,,64(1)3/GDP

play,5,1,talbf101,??,,NP

sub,rollr101,"Rich Rollins",1,9,11

play,5,1,rollr101,??,,S8

play,5,1,harpt101,??,,S/B.1-2

ladj,1,4

play,5,1,simpd102,??,,K/BOOT

ladj,1,5

play,5,1,comew101,??,,8/BOOT

ladj,1,4

play,5,1,simpd102,??,,2/BOOT

com,"$Davis is called out for batting out of order;"

com,"he doubled in 2 runs which triggered the protest;"

com,"since Simpson was the one due up, he was charged with the out"

Note that every batting out of turn situation has its own character, including whether or not it is detected by the opposition and whether or not the incorrect batter makes an out or reaches safely.

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Comment Record – instant replay

Instant replay of home run calls was instituted on 8/28/2008. It was expanded at the start of the 2014

season to include many other types of plays. For a more complete explanation and list all replays, see

the following two pages.

http://www.retrosheet.org/ReplayHR.htm

http://www.retrosheet.org/Replay.htm

Each time the replay system is used, a slash tag is added to the play string. This will be /UREV for an

umpire-initiated review and /MREV for a manager challenge. Immediately after that play there will be a

comment record with details of the replay/challenge. The fields in this string are:

com,"replay,inning,Batter ID,Batter Team ID,Umpire ID,Ballpark ID,Reason,Reversed,Initiator,Team,Type

Code"

Inning: inning in which the replay occurred

Batter ID: batter for the replay instance (not necessarily the player involved in the replay)

Batter Team ID: the team at bat for the replay

Umpire ID: crew chief’s ID

Ballpark ID: the ballpark in which the game was played

Reason (home run replay only):

O - Over the fence

F - Fair/foul

I - Fan interference

Reversed: Y or N

Initiator: I (home run instant replay, 2008-13), U (umpire, 2014-present), M (manager, 2014-present)

Team: team which challenged (only for M initiator)

Type code:

|  |  |
| --- | --- |
| Code | Desc |
| H | Home run |
| G | Grounds rule |
| N | Fan interference |
| S | Boundary call |
| C | Force play |
| A | Tag play |
| O | Fair/foul (outfield) |
| T | Trap play (outfield) |
| I | Hit by pitch |
| M | Timing play |
| B | Touching a base |
| R | Passing runners |
| K | Record keeping |
| L | Multiple issues |
| P | Home plate collision |
| X | Other |

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Com Record - ejections

Each time someone is ejected by an umpire, there will be multiple comments about the incident. This ejected person could be a player, coach, manager, trainer, mascot or fan. The first line will contain

details in the following record format.

com,"ej,Ejectee,Job Code,Umpire ID,Reason"

Ejectee – the ID of the person ejected

Job Code

P - Player

M - Manager

C - Coach

T - Trainer

N - Non-uniformed person

Umpire ID – the umpire who ejected the person

Reason: Short description

We have made an effort to standardize the test used in the reason field.

All following comments with be text describing the incident. It is most usual for it to be as simple as this.

com,” Babe Ruth ejected by HP umpire Tommy Connolly”

However, there are many ejections with more detailed text describing the incident.

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Com Record - umpire changes during games

Occasionally, an umpire will be injured or develop an illness during a game. When there are changes to

the umpire alignment during a game, there will be multiple comments with standardized fields to

describe the change.

com,"umpchange,Inning,Position,Umpire ID"

umpchange – standard text

Inning – the inning in which the change took place

Position:

umphome

ump1b

ump2b

ump3b

umplf

umprf

Umpire ID – the umpire who changed positions

The following comment will contain a text description of the incident.

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Comment Record - protests

When a manager protests an umpire ruling to the league office, a comment is added to the game to

indicate the details of that protest. This comment is usually at the start of the game.

com,"Protest=Code"

|  |  |
| --- | --- |
| P | unidentified team |
| V | disallowed protest by visiting team |
| H | disallowed protest by home team |
| X | upheld protest by visiting team |
| Y | upheld protest by home team |

Usually, there is a detailed comment at the spot in the game where the protest occurred.

play,5,0,feldm001,22,1LPB>F1S,K.1-2(WP)

## Com Record - suspensions

When a game is suspended by weather or other conditions, a comment is added to the game to indicate

the details of the suspension. This comment is usually at the start of the game.

com,"Suspend=YYYYMMDD,ParkID,Vis,Home,Outs"

YYYYMMDD - completion date

ParkID – if the game was resumed in another park from where it started

Vis - visitor score at time of suspension

Home - Home score at time of suspension

Outs - Length of game in outs at time of suspension

# Game Log File Layout

Field(s) Meaning

1 Date in the form "yyyymmdd"

2 Number of game:

"0" -- a single game

"1" -- the first game of a double (or triple) header

including seperate admission doubleheaders

"2" -- the second game of a double (or triple) header

including seperate admission doubleheaders

"3" -- the third game of a triple-header

"A" -- the first game of a double-header involving 3 teams

"B" -- the second game of a double-header involving 3 teams

3 Day of week ("Sun","Mon","Tue","Wed","Thu","Fri","Sat")

4-5 Visiting team and league

6 Visiting team game number

For this and the home team game number, ties are counted as

games and suspended games are counted from the starting

rather than the ending date.

7-8 Home team and league

9 Home team game number

10-11 Visiting and home team score (unquoted)

12 Length of game in outs (unquoted). A full 9-inning game would

have a 54 in this field. If the home team won without batting

in the bottom of the ninth, this field would contain a 51.

13 Day/night indicator ("D" or "N")

14 Completion information. If the game was completed at a

later date (either due to a suspension or an upheld protest)

this field will include:

"yyyymmdd,park,vs,hs,len" Where

yyyymmdd -- the date the game was completed

park -- the park ID where the game was completed

vs -- the visitor score at the time of interruption

hs -- the home score at the time of interruption

len -- the length of the game in outs at time of interruption

All the rest of the information in the record refers to the

entire game.

15 Forfeit information:

"V" -- the game was forfeited to the visiting team

"H" -- the game was forfeited to the home team

"T" -- the game was ruled a no-decision

16 Protest information:

"P" -- the game was protested by an unidentified team

"V" -- a disallowed protest was made by the visiting team

"H" -- a disallowed protest was made by the home team

"X" -- an upheld protest was made by the visiting team

"Y" -- an upheld protest was made by the home team

Note: two of these last four codes can appear in the field

(if both teams protested the game).

17 Park ID

18 Attendance (unquoted)

19 Time of game in minutes (unquoted)

20-21 Visiting and home line scores. For example:

"010000(10)0x"

Would indicate a game where the home team scored a run in

the second inning, ten in the seventh and didn't bat in the

bottom of the ninth.

22-38 Visiting team offensive statistics (unquoted) (in order):

at-bats

hits

doubles

triples

homeruns

RBI

sacrifice hits. This may include sacrifice flies for years

prior to 1954 when sacrifice flies were allowed.

sacrifice flies (since 1954)

hit-by-pitch

walks

intentional walks

strikeouts

stolen bases

caught stealing

grounded into double plays

awarded first on catcher's interference

left on base

39-43 Visiting team pitching statistics (unquoted)(in order):

pitchers used ( 1 means it was a complete game )

individual earned runs

team earned runs

wild pitches

balks

44-49 Visiting team defensive statistics (unquoted) (in order):

putouts. Note: prior to 1931, this may not equal 3 times

the number of innings pitched. Prior to that, no

putout was awarded when a runner was declared out for

being hit by a batted ball.

assists

errors

passed balls

double plays

triple plays

50-66 Home team offensive statistics

67-71 Home team pitching statistics

72-77 Home team defensive statistics

78-79 Home plate umpire ID and name

80-81 1B umpire ID and name

82-83 2B umpire ID and name

84-85 3B umpire ID and name

86-87 LF umpire ID and name

88-89 RF umpire ID and name

If any umpire positions were not filled for a particular game

the fields will be "","(none)".

90-91 Visiting team manager ID and name

92-93 Home team manager ID and name

94-95 Winning pitcher ID and name

96-97 Losing pitcher ID and name

98-99 Saving pitcher ID and name--"","(none)" if none awarded

100-101 Game Winning RBI batter ID and name--"","(none)" if none

awarded

102-103 Visiting starting pitcher ID and name

104-105 Home starting pitcher ID and name

106-132 Visiting starting players ID, name and defensive position,

listed in the order (1-9) they appeared in the batting order.

133-159 Home starting players ID, name and defensive position

listed in the order (1-9) they appeared in the batting order.

160 Additional information. This is a grab-bag of informational

items that might not warrant a field on their own. The field

is alpha-numeric. Some items are represented by tokens such as:

"HTBF" -- home team batted first.

Note: if "HTBF" is specified it would be possible to see

something like "01002000x" in the visitor's line score.

Changes in umpire positions during a game will also appear in

this field. These will be in the form:

umpchange,inning,umpPosition,umpid with the latter three

repeated for each umpire.

These changes occur with umpire injuries, late arrival of

umpires or changes from completion of suspended games. Details

of suspended games are in field 14.

161 Acquisition information:

"Y" -- we have the complete game

"N" -- we don't have any portion of the game

"D" -- the game was derived from box score and game story

"P" -- we have some portion of the game. We may be missing

innings at the beginning, middle and end of the game.

Missing fields will be NULL.

# Franchise/Team IDs

The file linked below is comma delimited. There are 149 team-league combinations listed in the file.

The fields are:

Team Abbreviation  
League  
City  
Nickname   
First year of this combination  
Last year of this combination (most recent season for current franchises)

The league codes are:

NA = National Association  
NL = National League  
AA = American Association  
UA = Union Association  
PL = Players League  
AA = American League  
FL = Federal League

The file has been modified from the version that was posted:

* The league field now contains the two letter entries shown above rather than the single letter codes.
* ID PTU (Pittsburgh in Union Association) removed because there is no data on the site for it. The CHU franchise has been changed to Chicago-Pittsburgh because the team moved in August, 1884.
* Nicknames have been added to replace "Unions" for teams in the UA.
* Current team nicknames replaced early ones for CLE and NYA.

**Note: We are currently revising this file, so it is not available**

[File with nicknames](http://www.retrosheet.org/Nickname.htm) for current teams including historical ones for the franchises as well as league/division memberships.