**RETROSHEET: HOW TO USE OUR EVENT FILES**

**Download to Hard disk:**

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.

**Identifying 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. |
| **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. |

**Ordered List of Field Names:**

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.

**Table for Play Records**

|  |  |
| --- | --- |
| 1. **Inning** | The first field is the inning. |
| 1. **Identify Team** | The second field is either **0 (for visiting team) or 1 (for home team).** |
| 1. **Retrosheet ID** | The third field is the Retrosheet ID code. |
| 1. **Count of Batter** | The fourth field is the count on the batter when this event occurred. Missing information is denoted by, "??" which appears in this field. |
| 1. **Sequence of Pitches** | 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. A blank field indicates no information. There is occasionally more than one event for each plate appearance, such as stolen bases. The pitch sequence is interrupted by a period, and there is another play record for the resumption of the batter's plate appearance. |
| **6. Indicate Event** | 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 area of the playing field is designated. See Diagram 1.1. |
| **6.1 Indicate Event Further Broken Down** | **a**. 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.  **b.** 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.  **c.** 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. |

**Additional Syntax to Note:**

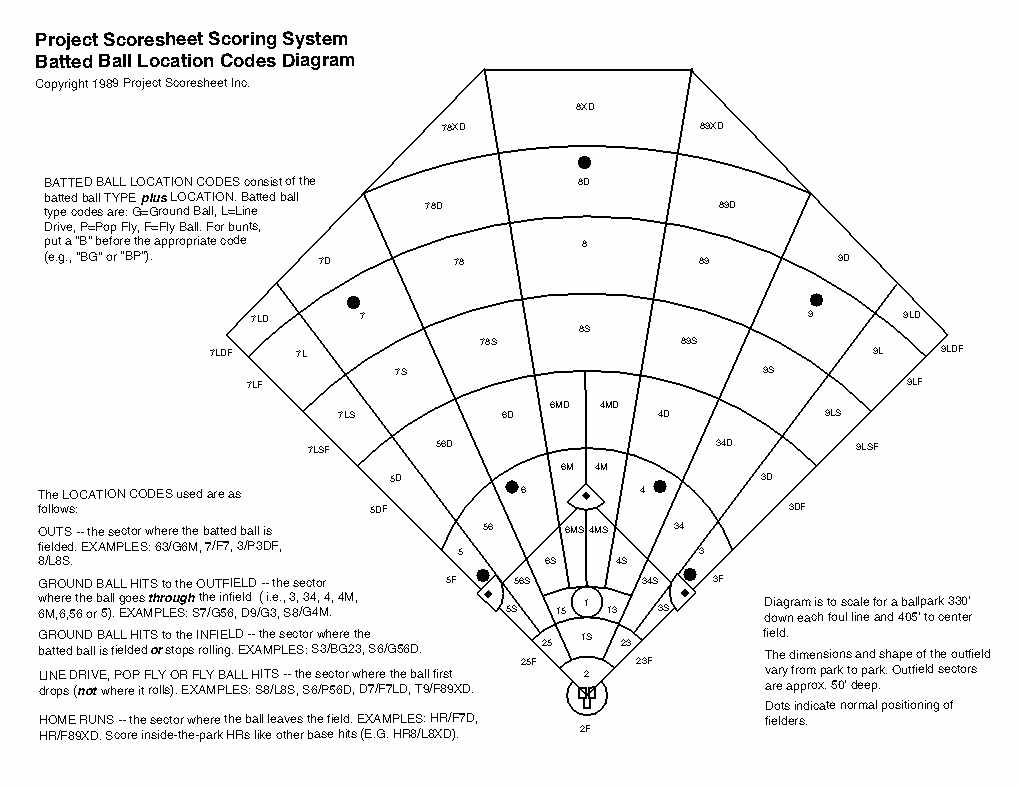
**com** 🡪 This record is used for those plays which require a special comment to describe the event.

**sub**  🡪 Whenever any change in the lineup occurs, there is a substitute record. There are five fields for this 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.



**Figure 1.1 Field Location**

**Using the Programs:**

\*It is important to note that 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.

**Using 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". 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

2. 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.

3. 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 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.

**Flags with BOX.EXE**

|  |  |
| --- | --- |
| -h | Print help |
| -vn | Print no status messages |
| -vf | Print status message for each file (default). |
| -vg | Print status message for each game. |
| -i id | Only process game id. |
| -y year | Year to process (for teamyyyy and aaayyyy.ros). |
| -s start | Earliest date to process (mmdd). |
| -e end | Last date to process (mmdd). |
| -q | Ask about each game. |
| -m | Use master player file instead of local roster files. |
| -D | Print debugging information |

**Using BEVENT and BGAME:**

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 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. 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. Please see “Event File Explained” for details about the formatting of the data.

**There are more than 95 different fields which may be created for each BEVENT** record. You can 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. (Type "BEVENT -h" to see these fields). 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.

**Flags with BEVENT.EXE**

|  |  |
| --- | --- |
| -h | Print help |
| -i id | Only process game given by id |
| -y year | Year to process (for teamyyyy and aaayyyy.ros). |
| -s start | Earliest date to process (mmdd). |
| -e end | Last date to process (mmdd). |
| -a | Generate Ascii-delimited format files (default) |
| -q | Ask about each game. |
| -m | Use master player file instead of local roster files. |
| -d | Print list of field numbers and descriptions |
| -ft | generate Fortran format files |
| -f flist | give list of fields to output  Default is 0-6,8-9,12-13,16-17,26-40,43-45,51,58-61 |

**Description of Each Field in BEVENT:**

|  |  |
| --- | --- |
| 1. **Game ID:** | Game ID description is located above. |
| 1. **Event Number:** | See description below for what these event files mean. |
| 1. **Inning:** | Inning in which this play took place. |
| 1. **Batting Team:** | A binary indication of the team at bat ("0" for the visiting team and "1" for the home team). |
| 1. **Outs:** | Number of outs before this play. |
| 1. **Balls, Strikes, Pitch Sequence** | These three consecutive fields present the pitch information for this play. |
| 1. **Vis Score** | Number of runs for the visiting team before this play. |
| 1. **Home Score** | Number of runs for the home team before this play. |
| 1. **Batter** | Player ID code for the batter. |
| 1. **Batter Hand** | One character describing how the batter batted for this event (L or R). |
| 1. **Res Batter & 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). |
| 1. **Pitcher** | Player ID code for the pitcher. |
| 1. **Pitcher Hand** | The hand with which the pitcher throws (L or R). |
| 1. **Res Pitcher & 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. |
| 1. **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. |
| 1. **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"** |
| 1. **Event Text:** | The complete description of the play using the format described for the event files. |
| 1. **Leadoff Flag:** | A T indicates the first batter of each inning and F for all others. |
| 1. **Pinch hit Flag:** | Another one-character flag which is T for pinch-hitters and F for all others. |
| 1. **Defensive Position:** | The defensive position currently being played by this batter. It is pinch-hitter (position 11) for pinch-hitters. |
| 1. **Lineup Position:** | Position in the batting order for this batter. |
| 1. **Event Type:** | 25 different numeric codes describe each event. They are: 0 Unknown event, 1 No event, 2 Generic out, 3 Strikeout, 4 Stolen bases, 5 Defensive indifference, 6 Caught stealing, 7 Pickoff errors, 8 Pickoff, 9 Wild pitch, 10 Passed ball, 11 Balk, 12 Other advance, 13 Foul error, 14 Walk, 15 Intentional walks, 16 Hit by pitch, 17 Interference, 18 Error, 19 Fielder's choice, 20 Single, 21 Double, 22 Triple, 23 Home run, 24 Missing play |
| 1. **Batter Event Flag:** | A one-character indication of whether 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. |
| 1. **AB Flag** | A one-character indication of whether batter was charged with at-bat (T = yes, F = no). |
| 1. **Hit Value** | One number indicating value of hit (0 = no hit; 1 = single; 2 = double; 3 = triple; 4 = home run). |
| 1. **SH Flag** | One-character indicating sacrifice hit (T = yes; F = no). |
| 1. **SF Flag** | One-character indicating sacrifice fly (T = yes; F = no). |
| 1. **Outs on Play** | Number of outs recorded on this play. |
| 1. **Double Play Flag** | One-character field of DP or not. |
| 1. **Triple Play Flag** | One-character field of TP or not. |
| 1. **RBI on Play** | Number of RBI credited to batter on this play. |
| 1. **Wild Pitch Flag, Passed Ball Flag** | Two records with indication of whether there was a WP or PB on this play. |
| 1. **Fielded By** | Identity of the fielder who played the ball. This is especially important for base hits when no formal fielding credit is given. |
| 1. **Batted Ball Type** | Descriptor which is either F (fly ball), L (line drive), P (pop-up), or G (ground ball). |
| 1. **Bunt Flag** | Descriptor for whether play was a bunt. |
| 1. **Foul Flag** | Descriptor for whether ball was played in foul ground. |
| 1. **Hit Location** | The zone on the field where the ball was hit. Refer to figure 1.1. |
| 1. **Num Errors** | Number of errors on this play (a maximum of three is allowed). |
| 1. **Error Players and Type** | 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). |
| 1. **Batter Dest** | The base which the batter reached after the play. If he was out, the base is 0. |
| 1. **Runner Dest** | The next three fields contain the base reached by each of the three runners after the play. If there was no advance, then the base show 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. |
| 1. **Plays** | The next four fields indicate the play (if any) made on the batter and each of the runners (if any). |
| 1. **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. |
| 1. **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. |
| 1. **New Game, 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. |
| 1. **Pinch Runners** | The next three fields indicate if a pinch runner has entered the game and at which base. |
| 1. **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 pinch runner at that base, the field contains the NULL string "". |
| 1. **Removed Batter** | If there is a pinch hitter, this field contains the player ID of the batter removed. f there is no pinch hitter, this field contains the NULL string "". |
| 1. **Removed Batter Position** | If there is a pinch hitter, this field contains the fielding position of the removed batter. If there is no pinch hitter, this value is 0. |
| 1. **Fielder Putouts** | The next three fields indicate the first, second, and third fielders credited with putouts on the play. |
| 1. **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 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 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 of each Field in BEVENT:**

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 pinch-hit 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 scores and unearned, 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

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

**Number of each Field in BGAME:**

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 8m 80 home batter 9, 81 home position 9

### Retrosheet makes no guarantees of accuracy for the information that is supplied. Much effort is expended to make our website as correct as possible, but Retrosheet shall not be held responsible for any consequences arising from the use the material presented here. All information is subject to corrections as additional data are received. We are grateful to anyone who discovers discrepancies and we appreciate learning of the details.