Scripts

bevent.sh

bevent.sh - uses BEVENT.EXE to parse the event files looking for pitcher statistics and outputs all_events.csv.

When run against all of the event files, this generated **291MB** worth of data.

```
0    game id
3    batting team

    0   visiting team
    1   home team

14   pitcher
34   event type

    Code Meaning

    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
36   AB flag
43   RBI on play
```

The resulting data will need to be summed to get season numbers for each pitcher.

bgame.sh

bgame.sh - uses BGAME.EXE to parse the event file for game statistics and outputs all_games.csv.

When run against all of the eent files, this generated **4.6MB** worth of data.

```
0    game id
7    visiting team
8    home team
9    game site
25   pitchers entered?
34   visitor final score
35   home final score
36   visitor hits
37   home hits
```

teams.sh

teams.sh - reads in the teams files (TEAMYYYY), places the year as the first entry, and outputs teams_combined.csv.

stats.py

stats.py - reads from the various csv files generated with the shell scripts from above and outputs pitchingAgainst_update.sql. This script attempts to match players to the lahman playerIDs and updates the PitchingAgainst table.

Known Issues

If a player, through multiple stints, ends back up on the same team, stats.py will incorrectly assume they are part of the same stint and combine them. This was discovered late in the process of this project and could not be resolved in a timely manner. This affects roughly 6+ players.

There are players within the Lahman dataset that do not exist in the retrosheets dataset and therefore cannot be updated in the pitchingAgainst table. This affects affects roughly 3+ players.