

1. atbats

- a. Pk: ab_id
- b. Fk:
 - i. g_id -> games.g_id
 - ii. batter_id -> players_names.id
 - iii. pitcher_id -> players_names.id
- c. comments:
 - i. we can combine 2019_atbats and atbats(which has 2015~2018 data)
 - ii. "p_score" means the score for the pitcher's team (which is somewhat confusing imo lmao)
 - iii. "top" means whether it is top inning or not
 - iv. "o" means outs
 - v. "stand" means whether the batter is left or right-handed
 - vi. we can probably create two attributes "home_Score" or "away_score" to indicate their current score.
Or not, we can just use "top" and "p_score" to determine the current score for either team

2. ejections

- a. Pk: ab_id, player_id but they are foreign keys...
Or perhaps "des" and "date"?
- b. Fk:
 - i. ab_id -> atbats.ab_id
 - ii. player_id -> players_names.id
 - iii. g_id -> games.g_id
 - iv. dates -> games.date
- c. comments:
 - i. might need to come up with a proper pk
 - ii. not data for 2019
 - iii. not sure what "event_num" means. They are linked to event_num in "pitches" table but lots of data are missing so I am not sure if we should still keep this
 - iv. BS: 'Y' if ejection was for arguing balls and strikes, empty otherwise
 - v. CORRECT: if BS ejection is correct. C is correct, I is not
 - vi. Hmm this is probably the most problematic one

3. games

- a. Pk: g_id
- b. Fk: none
- c. comments:
 - i. we can combine 2019_games and games(which has 2015~2018 data). However, 2019 is missing some data including umpires, wind, elapsed time, attendance, start time, weather, delay

- ii. we may need to discuss how to deal with missing data in 2019. Assigning them to NULL?
- iii. We don't have a database for teams and what we have now is just a short term of the team like "TOR" which represents the Toronto bluejays. Should we create one just like the players_names?

4. pitches

- a. Pk: event_num (they are meant for comparing the data with ejections so I believe they are unique)
Or we can have ab_id, and pitch_num as pk
- b. Fk:
 - i. ab_id -> atbats.ab_id
- c. comments:
 - i. we can combine 2019_pitches and pitches(which has 2015~2018 data).
 - ii. Lots of unnecessary data. We can get rid of all the xyz coordinates. We can keep the "start speed" and "end speed" and "spin rate" (but 2019 does not have spin rate).
 - iii. "code" and "type" are identical in most cases but we can keep both. "code" can be used if the user is looking for something specific. "type" is just a simplified version of it
 - iv. Attributes after "code" are worth keeping
 - v. If we don't want to use event_num then we can get rid of it

5. players_names

- a. Pk: id