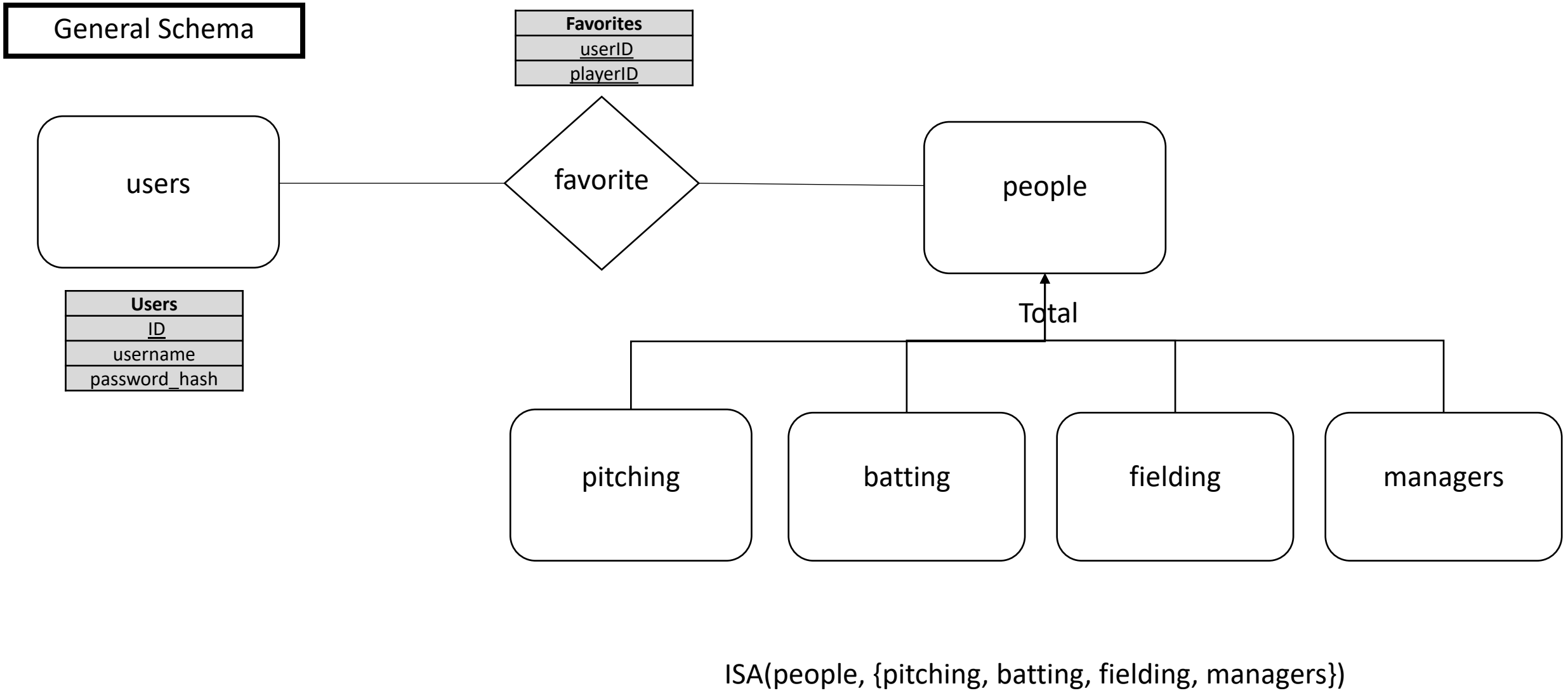
ER Diagrams for CSI 5302 Group Project

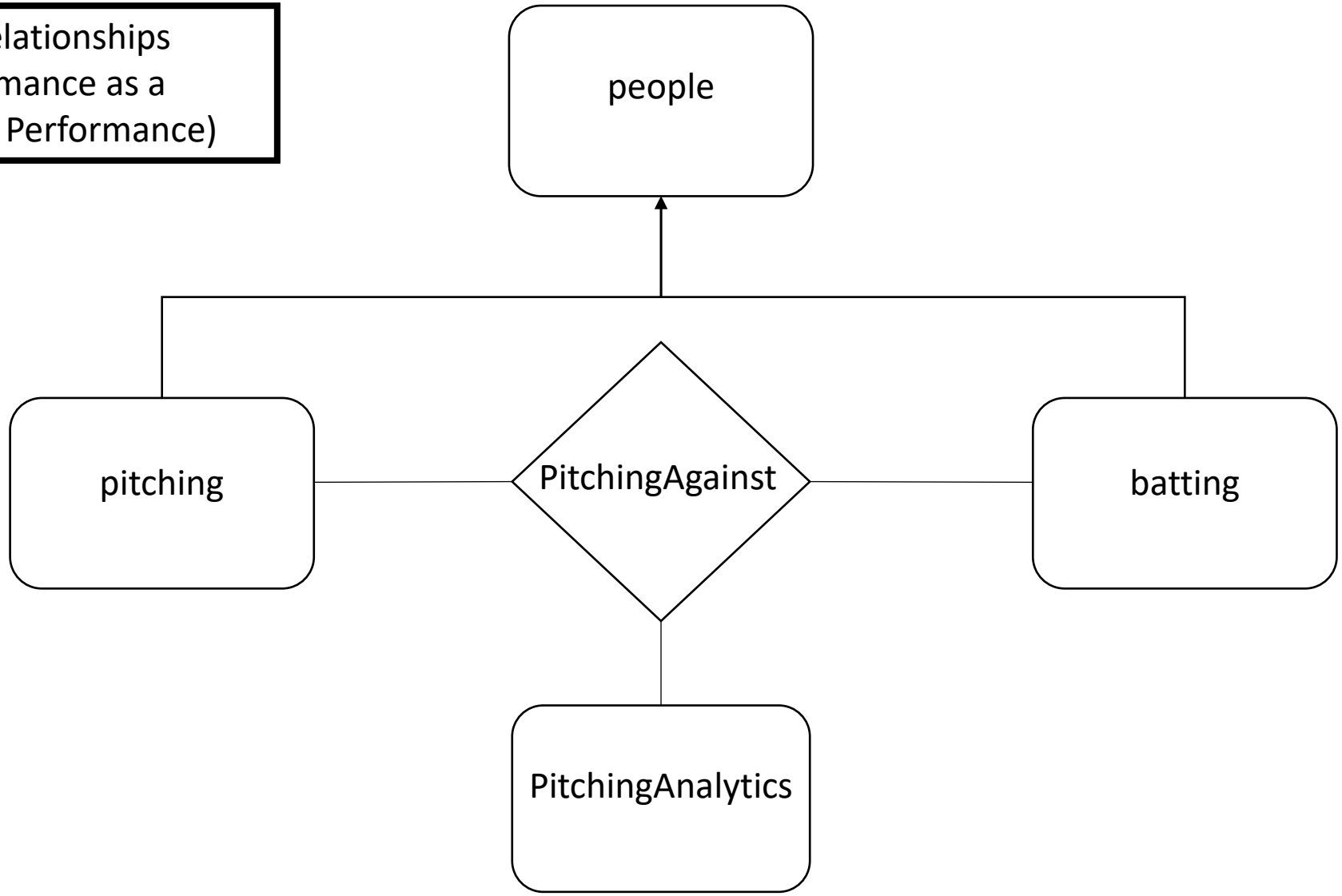
Team: Matthew Hayes, Sarah Smallwood, Joshua Wellman



ISA(people, {pitching, batting, fielding, managers})

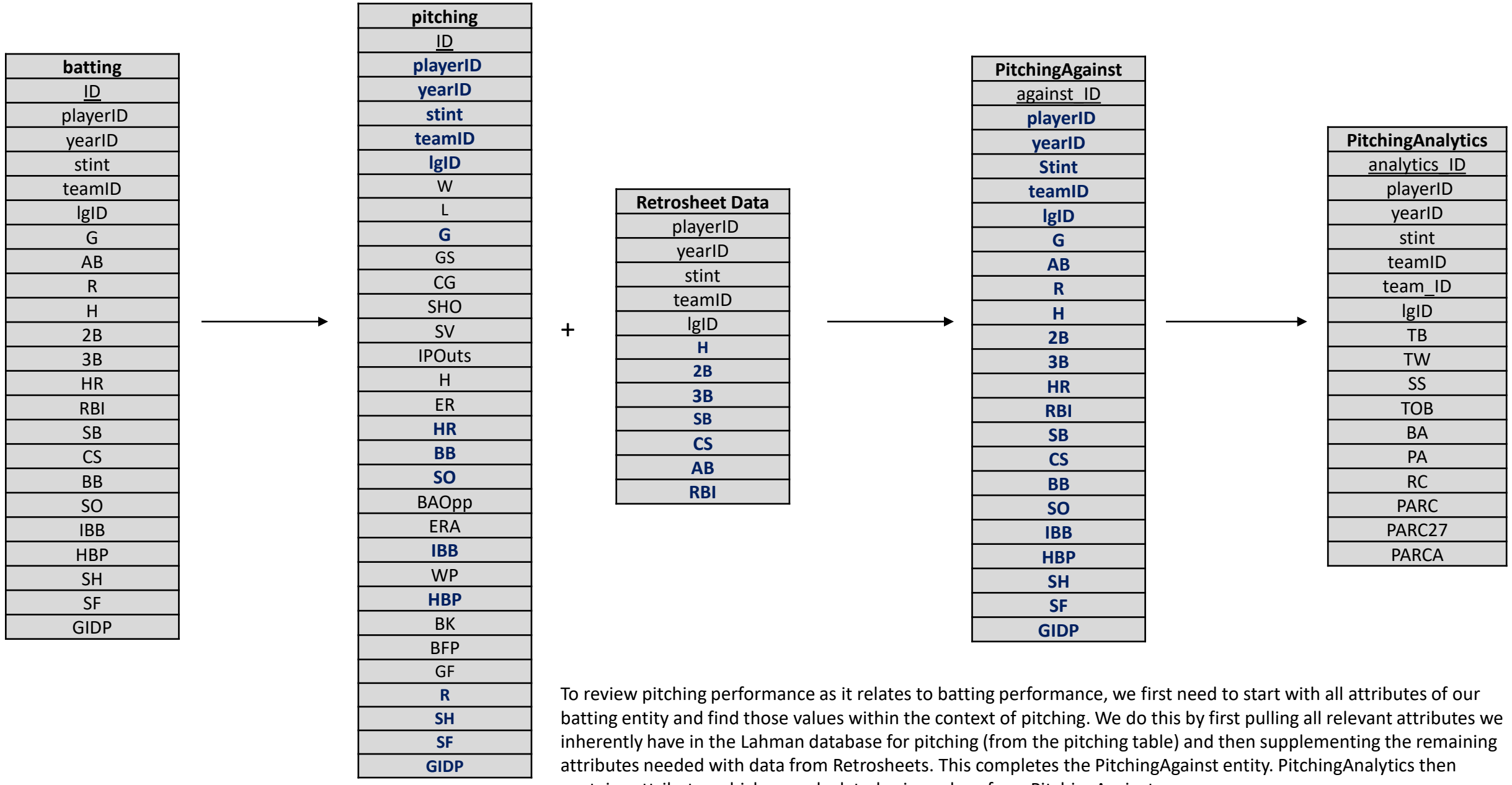
For this project, we will have users that will access our web application and favorite people, specifically pitchers, within the Lahman database. Both users and favorites are new entities we add to our existing Lahman database. Inherently in the Lahman database, all people are categorized into pitchers, batters, fielders, and managers. Individual players can be sorted into more than 1 category.

Project Entity Relationships
(Pitching Performance as a Function of Batting Performance)



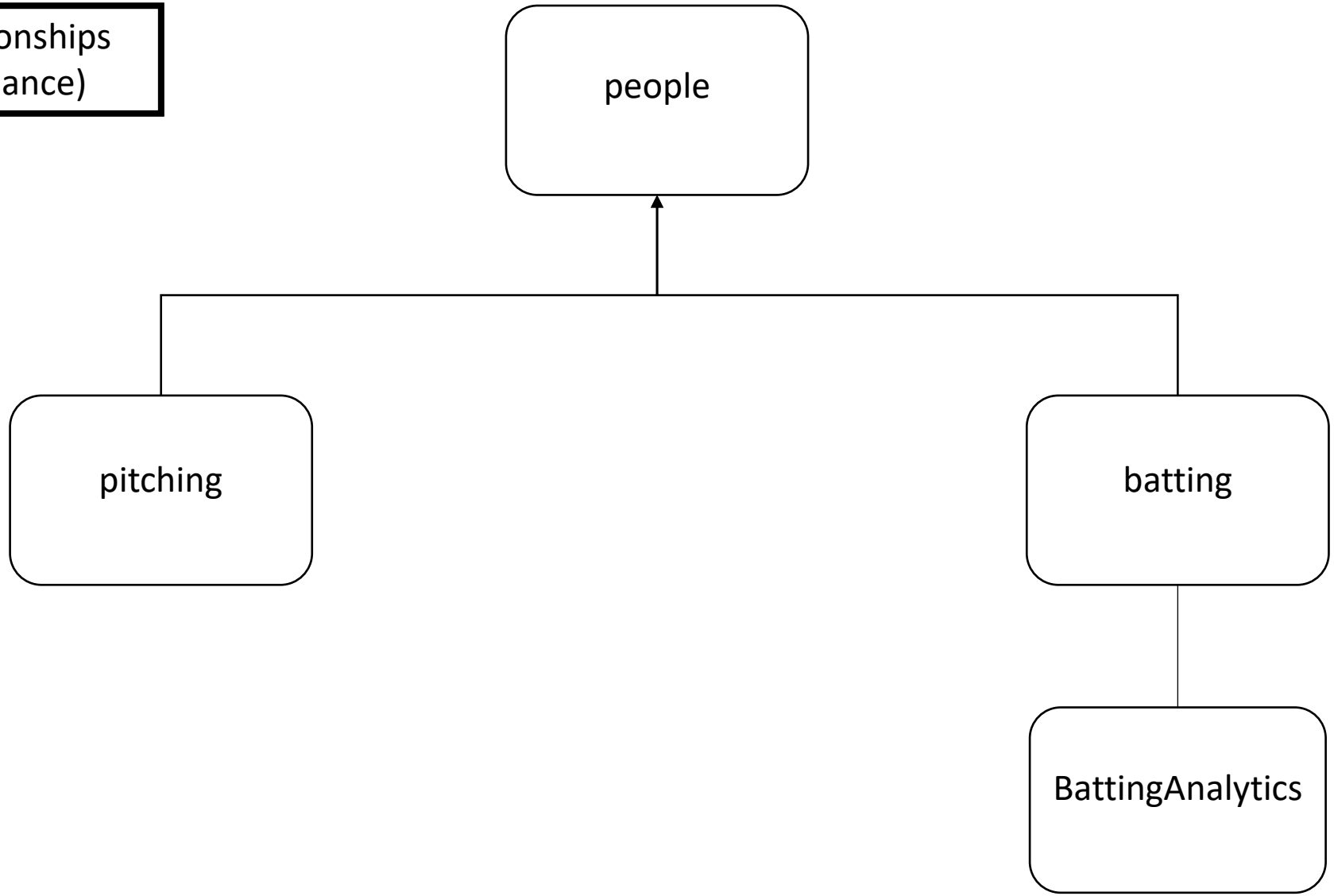
Because pitchers play against batters, we can analyze the relationship between pitchers and batters with a set of statistics that shows pitchers' performance specifically as it relates to that of batters'. We call this relationship PitchingAgainst and it becomes its own entity. To analyze the performance described in this relationship, another entity is needed to record statistics that are not directly related to batting, but stem from pitching performance. We call this entity PitchingAnalytics.

Project Tables (Pitching Performance as Related to Batting Performance)



To review pitching performance as it relates to batting performance, we first need to start with all attributes of our batting entity and find those values within the context of pitching. We do this by first pulling all relevant attributes we inherently have in the Lahman database for pitching (from the pitching table) and then supplementing the remaining attributes needed with data from Retrosheets. This completes the PitchingAgainst entity. PitchingAnalytics then contains attributes which are calculated using values from PitchingAgainst.

Extra Entity Relationships
(Batting Performance)



During lesson exercises, we also analyzed batting performance, but not in the same way that we are measuring pitching performance for this project (as a function specifically of how batters performed when directly compared to pitchers). Therefore, there is no direct relationship to the pitching entity here, and only 1 entity is needed to describe these batting statistics, which we call BattingAnalytics.

Project Tables (Batting Performance)

batting
ID
playerID
yearID
stint
teamID
lgID
G
AB
R
H
2B
3B
HR
RBI
SB
CS
BB
SO
IBB
HBP
SH
SF
GIDP

BattingAnalytics
analysis_ID
playerID
yearID
lgID
G
AB
R
H
2B
3B
HR
RBI
SB
CS
BB
SO
IBB
HBP
SH
SF
GIDP
OBP
TB
PA
RC
RC27

To review batting performance, we first need to start with all attributes of our batting entity and calculate the desired statistics. We do this by first pulling all attributes from the Lahman batting table and then using these attributes to calculate Total Bases (TB), Plate Appearances (PA), Runs Created (RC), Runs Created per 27 Outs (RC27). The difference between creating this table and creating the aforementioned PitchingAnalytics table is that BattingAnalytics is measuring performance of batters using data on batting performance alone, while PitchingAnalytics is measuring pitching performance within the context of batting performance.