Overview

This project is designed to take around **four hours** of your time. We prefer you actually perform as much of the analysis as possible, but we understand that the recommended **three hours** of analysis and **one hour** writing up your findings may not be enough time to do everything you intended to do. If you do not have time to implement your entire analysis plan, please provide a **detailed description** of how you would attack this problem in the written report.

You will also need to provide a detailed write-up that explains the methods you used and includes a copy of any code you used. Separate files containing code and the text of your write-up are certainly acceptable, but "notebook" style documents (e.g. RMarkdown or Jupyter) are preferred.

Data Analysis

Suppose you have been asked to evaluate hitter performance based on measurements of batted ball metrics. For each batted ball, you have measurements of the following two quantities:

1. The **speed-off-bat (miles per hour; or "exit velocity")**, which is a measure of how hard a ball was hit.
2. The **initial vertical angle (degrees; sometimes called "launch angle")**
    a. A vertical angle of 0 is a ball whose initial trajectory is parallel to the ground
    b. Fly balls and popups have large, positive vertical angles
    c. Ground balls have negative vertical angles

Speed-off-bat and vertical angle are measured by ***two different measurement systems*** (system A and system B).

- For some batted balls, you have a speed-off-bat and vertical angle measurement from both systems
- For some batted balls, you have a measurement from system A, but no measurement from system B, and vice versa. Missing data points are coded as NA
- Additionally, for each batted ball you are given a hit type (either fly_ball, ground_ball, line_drive, popup) recorded by a human observer watching the game (not derived from the data from sensor A or B)

You will be provided with a csv containing measurements for many batted balls. Each row corresponds to an individual batted ball. For each batted ball, you are given an ID for the batter and pitcher, as well as a speed-off-bat and vertical angle measurement from system A and system B (4 example rows are shown below).

Batter	Pitcher	hittype	speed_A	vangle_A	speed_B	vangle_B
393	405	ground_ball	110.9876	4.194081	103.8426	3.164307
366	405	ground_ball	NA	NA	28.0922	-28.3241
448	518	line_drive	102.7576	11.75185	97.846	11.6588
140	518	fly_ball	61.95209	33.48815	NA	NA

- Although we suspect that system A is much more accurate, each measurement system has its own quirks. We recommend you begin by investigating the difference between the two systems.
- Your approach for estimating speed-off-bat should be sound for players who only have data from system A, as well as players who only have data from system B.

Your task: We would like you to project the "true" average speed-off-bat (on the scale of system A, or system B, or some combination of the two systems) for each batter in the following season.

Evaluation

Through this project, we hope to evaluate your communication skills, ability to manage imperfect data, ability to project future player performance, and creativity.