Project Part 1

Topic of Interest

For my project, I want to explore what makes the top MLB power hitters different from the rest of the hitters There are many factors that go into the evaluation of a hitter. Batting average (the amount of times the hitter hits the ball over the number of times the hitter is at the plate), on base percentage (number of times the hitter reaches at least first base over the number of at bats), and slugging percentage (total number of bases a hitter records per at-bat) are just some of these many factors. What I specifically want to look at is the relationship between variables that are collected when a batter hits a home run. By looking specifically at home run statistics, I can start to make out what separates a standard hitter from a very good power hitter.

Research Questions

The three research questions I have come with are as follows: 1) What is the relationship between the average hit angle and average home run distance? Answering this question can provide insight on how batters can set up in the box to either raise or lower their hit angle (if there is a relationship between the two variables). 2) Does a batter's hit speed play a role in the distance they hit the ball? Here I can see whether or not a more powerful hitter tends to hit the ball with a lot of speed or not. 3) Is a batter's hit speed or launch angle more important to consider when trying to hit a home run? This question combines the first two and explores the possibility of one metric being more important when training to hit home runs.

Populations of Interest

There are many groups of people that could be interested in this type of analysis. The first group is hitters. If a hitter is trying to become more valuable and earn a new contract, they might want to tailor their swing to hit more home runs. Finding out whether to change their swing angle or speed could help with that. Another group of interest could be coaches, managers, and/or owners. Management would be especially interested in this when trying to recruit players of value to the team.

Appropriate Data

Appropriate data for this analysis should come from MLB data sources that keep track of hitting statistics. These statistics come directly from the games over a season, and are a primary source. For this project, I have found a dataset from the MLB at this link: https://baseballsavant.mlb.com/leaderboard/statcast.

Data

```
## Rows: 247 Columns: 18
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr  (2): last_name, first_name
## dbl (16): player_id, attempts, avg_hit_angle, anglesweetspotpercent, max_hit...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## # A tibble: 6 x 18
##   last_name  first~1 playe~2 attem~3 avg_h~4 angle~5 max_h~6 avg_h~7  fbld     gb
##   <chr>      <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>  <dbl>
## 1 Semien     Marcus   543760     482    19.7    32      110.     87.1  91.5   83.7
## 2 Turner     Trea     607208     467     9.9    35.5    112.     88.9  92.1   86.2
## 3 Freeman    Freddie  518692     466    14.1    44.2    112.     91.3  94.7   87.1
## 4 Guerrero ~ Vladim~  665489     464     4.6    28.9    118.     93    98.5   90.5
## 5 Verdugo    Alex     657077     463     8.9    37.8    111      89.1  93.1   86.1
## 6 Ramírez    José     608070     457    20.5    33.7    114.     88.1  91.5   87.1
## # ... with 8 more variables: max_distance <dbl>, avg_distance <dbl>,
## #   avg_hr_distance <dbl>, ev95plus <dbl>, ev95percent <dbl>, barrels <dbl>,
## #   brl_percent <dbl>, brl_pa <dbl>, and abbreviated variable names
## #   1: first_name, 2: player_id, 3: attempts, 4: avg_hit_angle,
## #   5: anglesweetspotpercent, 6: max_hit_speed, 7: avg_hit_speed
```