

# Appendix

To see all of the data and code that went into this project, you can go to the [Github page](#) here.

## Data Set Columns

Full Batter Data set (Suffix not included for brevity):

- Player - Current year and player name (key identifier, current only)
- Age - Current age of player (current only)
- G - Games Played
- PA - Plate Appearances
- HR - Home Runs
- R - Runs Scored
- RBI - Runs Batted In
- SB - Stolen Bases
- BB\_rate - Walk Rate
- K\_rate - Strikeout Rate
- ISO - Isolated Power
- BABIP - Batting Average Balls in Play
- AVG - Batting Average
- OBP - On-Base Percentage
- SLG - Slugging Percentage
- wOBA - Weighted On-Base Average
- wRC\_plus - Weighted Runs Created Plus
- BsR - Baserunning
- Off - Offense Rating
- Def - Defense Rating
- WAR - Wins Above Replacement
- MLS - Major League Service (current only)
- Salary - That year's Salary
- Salary\_Y - Next year's salary (Response)

Tuned Batter Data set (Suffix not included for brevity):

- Player - Current year and player name (key identifier, current only)
- Age - Current age of player (current only)
- PA - Plate Appearances
- HR - Home Runs
- RBI - Runs Batted In
- wOBA - Weighted On-Base Average
- WAR - Wins Above Replacement
- MLS - Major League Service (current only)
- Salary - That year's Salary (current only)
- Salary\_change - Difference between last year and current year's salary (P1 and P2)
- Salary\_Y - Next year's salary (Response)

- Interactions between Salary\_C and all other predictors

Full Pitcher Data set (Suffix not included for brevity):

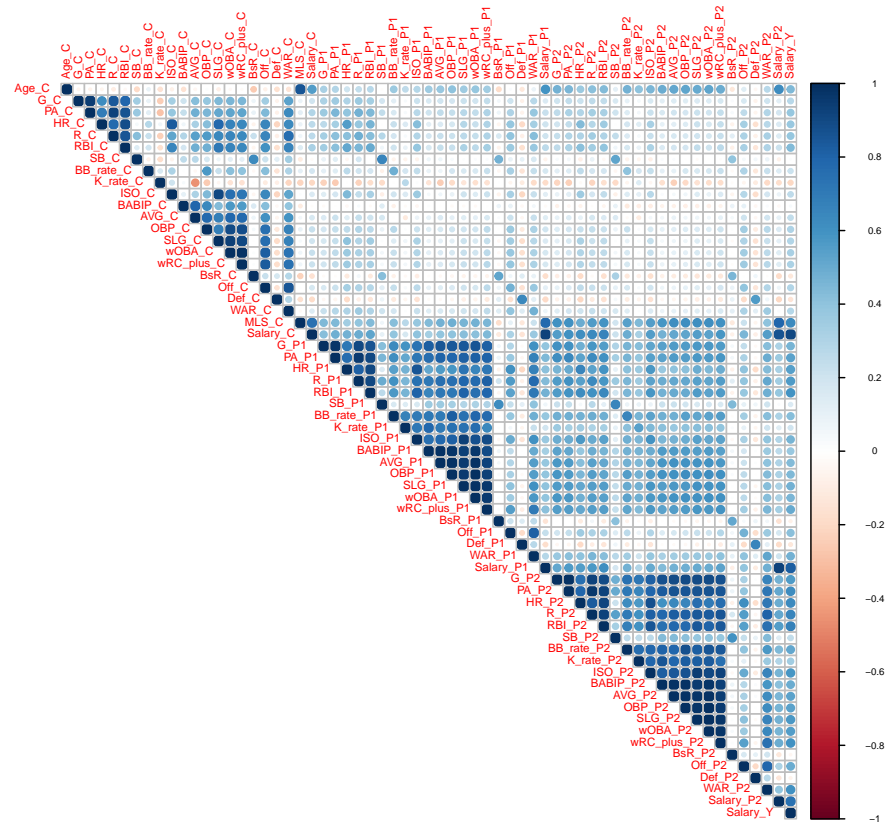
- Player - Current year and player name (key identifier, current only)
- Age - Current age of player (current only)
- W - Wins
- L - Losses
- SV - Saves
- G - Games Pitched
- GS - Games Started
- IP - Innings Pitched
- K\_9 - Strikeouts per 9 Innings
- BB\_9 - Walks per 9 Innings
- HR\_9 - Home Runs per 9 Innings
- BABIP - Batting Average Balls in Play
- LOB\_rate - Rate of runners left on base
- GB\_rate - Groundball rate
- HR\_FB\_rate - Home Run to Flyball rate
- vFA - Average Fastball Velocity
- ERA - Earned Run Average
- ERA\_minus - ERA minus
- FIP - Fielding Independent Pitching
- FIP\_minus - FIP minus
- xFIP - Expected FIP
- xFIP\_minus - Expected FIP minus
- WAR - Wins Above Replacement
- SIERA - Skill-Interactive ERA
- MLS - Major League Service (current only)
- Salary - That year's Salary
- Salary\_Y - Next year's salary (Response)

Tuned Pitcher Data set (Suffix not included for brevity):

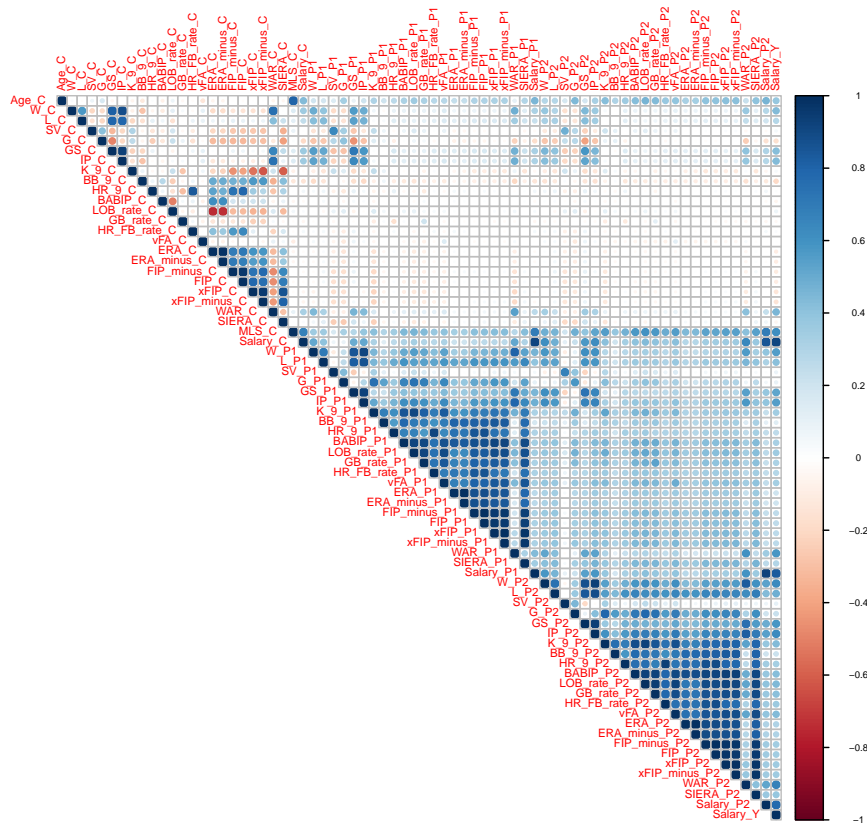
- Player - Current year and player name (key identifier, current only)
- Age - Current age of player (current only)
- W - Wins
- GS - Games Started
- IP - Innings Pitched
- ERA\_minus - ERA minus
- FIP\_minus - FIP minus
- WAR - Wins Above Replacement
- MLS - Major League Service (current only)
- Salary - That year's Salary
- Salary\_Y - Next year's salary (Response)
- Interactions between Salary\_C and all other predictors

## Correlation Plots

### Batter Correlation Plot



Pitcher Correlation Plot



## Model Hyperparameters

### Batter Full Models

- Regularization
  - penalty = 2.05e-10
  - mixture = 0.314 (0 indicates Ridge, 1 indicates LASSO)
- Random Forest
  - mtry = 58
  - trees = 1000
  - min\_n = 16
- XGBoost
  - mtry = 56
  - trees = 1000
  - min\_n = 9
  - tree\_depth = 9
  - learn\_rate = 0.0023
  - loss\_reduction = 6.04e-08
  - sample\_size = 0.76

### Batter Tuned Models

- Regularization
  - penalty = 0.0047
  - mixture = 0.03 (0 indicates Ridge, 1 indicates LASSO)
- Random Forest
  - mtry = 33
  - trees = 1000
  - min\_n = 2
- XGBoost
  - mtry = 28
  - trees = 1000
  - min\_n = 12
  - tree\_depth = 13
  - learn\_rate = 0.0025
  - loss\_reduction = 1.17e-07
  - sample\_size = 0.63

### Pitcher Full Models

- Regularization
  - penalty = 1.28e-09
  - mixture = 0.062 (0 indicates Ridge, 1 indicates LASSO)
- Random Forest
  - mtry = 63
  - trees = 1305

- min\_n = 10
- XGBoost
  - mtry = 56
  - trees = 1823
  - min\_n = 4
  - tree\_depth = 9
  - learn\_rate = 0.0018
  - loss\_reduction = 2.08e-08
  - sample\_size = 0.56

## Pitcher Tuned Models

- Regularization
  - penalty = 7.11e-05
  - mixture = 0.97 (0 indicates Ridge, 1 indicates LASSO)
- Random Forest
  - mtry = 126
  - trees = 799
  - min\_n = 6
- XGBoost
  - mtry = 119
  - trees = 870
  - min\_n = 6
  - tree\_depth = 10
  - learn\_rate = 0.0021
  - loss\_reduction = 2.34e-10
  - sample\_size = 0.64