Batting data

* 1900 – 2016
* Cut-off criteria
  + Plate appearances (PA): average plate appearances per game over a 162 game season
    - \*MLB requirements
* **Main goals**
  + Create a “Cumulative rate of failure” term (**SURVIVAL ANALYSIS**)
    - A value that determines the player’s cumulative success
    - Allows for single value comparison of players
  + **Career predictions**
    - Develop models to predict batter’s success for the remainder of their career
      * “Age curves”
  + **Season predictions**
    - Predict batting MVPs using data from the first half of a season
    - Predict stats for second half of season
* Models to apply
  + OLS
    - Attempt to find correlations using multivariate OLS
      * Two inputs: age & ‘input X’
  + Regularized regression
    - Repeat OLS calculations
  + Non-linear Regression using Loss functions
    - Predict players cumulative rate of failure based on previous stats and age
  + Kriging || Neural Network
    - Predict average stats for of players based on first half of their season