

Title: History's greatest homerun hitter and their strikeout rate
Author: Yosuke Ishizaka
Class: Coursera, Developing Data Products
Date: 9/25/2015

Introduction:

Top homerun hitters among major league baseball players have special skill sets that allow them to identify when to swing and make hard contact with a bat when a pitch is thrown inside the strike zone. A batter cannot hit a homerun without making a good, hard contact. Thus, it's crucial that a batter doesn't strike out often. Players such as Barry Bonds and Hank Aaron, who lead their eras in homeruns, were able to perform at the highest level because they did not strike out often, and when they saw a pitch they can hit, the ball was hit into the stand frequently. Not very many players even come close to the level of the two players, however, history suggests that top homerun hitters tend to have low strikeout rate.

Methods:

All of my data comes from Lahman database package in R. Documentation can be obtained from <http://seanlahman.com/files/database/readme2014.txt>. Package consists of various baseball datasets. In this project, I used Master and Batting datasets. I joined the datasets into one data and created a new column to store player's full name. I used the formula $PA = AB + BB + HBP + SF + SH$ to create Plate Appearances variable, which I used to calculate the Strikeout Rate. Strikeout rate is obtained by the formula $KRate = SO/PA$. Since Sacrifice Fly was not officially recorded in some years, I found NA entries that I had to convert to zero in order to compute plate appearances. I combined yearly stats for each player to get career homerun totals and career strikeout rates. Finally, I sorted by top 100 homerun hitters to get a table of only the top 100 homerun hitters.

In this App, users can select up to 2 players from a selection of top 100 homerun hitters. When a player is selected, it obtains career stats from the table that I created and outputs player name, homerun totals, strikeout totals, plate appearances, and strikeout rate. It also plots a scatter plot where users can see the selected player's skill relative to all historical baseball players. Additionally, users can select up to 2 players of their choice and compare career stats and see where they rank on the KRate versus Total Homeruns plot.

Conclusion:

We are able to observe association between strikeout rates and total homeruns hit. It's clear that the most successful homerun hitters have low strikeout rate.