IS4300 Final Project

Cabot Boyd

2022-04-27

## Introduction

We will be examining a data set found on Github called “Yearly batting records for all major league baseball players.” It contains a total of 21,699 records, covering 1,228 players from 1871 to 2007 and only contains players with more than 15 seasons played.

The data contained in this .csv is various batting related statistics, such as at bats, home runs, stolen bases, and hits by pitch.

This report will be examining Gary Sheffield’s case for making the MLB Hall of Fame. We will doing this by comparing his offensive performance to other Hall of Famers from the same era.

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

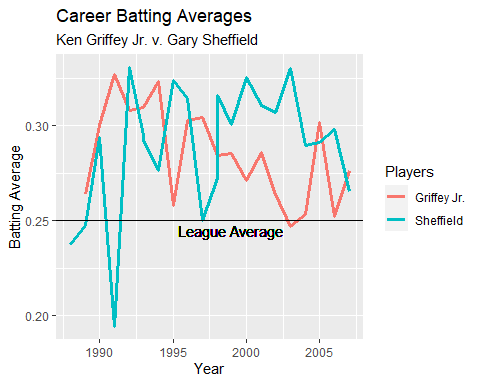
## New names:  
## \* `` -> ...1

## Rows: 21699 Columns: 23  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (3): id, team, lg  
## dbl (20): ...1, year, stint, g, ab, r, h, X2b, X3b, hr, rbi, sb, cs, bb, so,...  
##   
## 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.

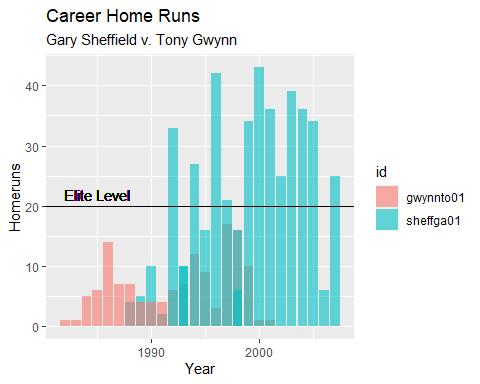
## spec\_tbl\_df [21,699 x 23] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ...1 : num [1:21699] 4 44 68 99 102 106 113 121 143 167 ...  
## $ id : chr [1:21699] "ansonca01" "forceda01" "mathebo01" "startjo01" ...  
## $ year : num [1:21699] 1871 1871 1871 1871 1871 ...  
## $ stint: num [1:21699] 1 1 1 1 1 1 1 1 1 1 ...  
## $ team : chr [1:21699] "RC1" "WS3" "FW1" "NY2" ...  
## $ lg : chr [1:21699] NA NA NA NA ...  
## $ g : num [1:21699] 25 32 19 33 29 29 29 46 37 25 ...  
## $ ab : num [1:21699] 120 162 89 161 128 146 145 217 174 130 ...  
## $ r : num [1:21699] 29 45 15 35 35 40 36 60 26 40 ...  
## $ h : num [1:21699] 39 45 24 58 45 47 37 90 46 53 ...  
## $ X2b : num [1:21699] 11 9 3 5 3 6 5 10 3 11 ...  
## $ X3b : num [1:21699] 3 4 1 1 7 5 7 7 0 0 ...  
## $ hr : num [1:21699] 0 0 0 1 3 1 2 0 0 0 ...  
## $ rbi : num [1:21699] 16 29 10 34 23 21 23 50 15 16 ...  
## $ sb : num [1:21699] 6 8 2 4 3 2 2 6 0 2 ...  
## $ cs : num [1:21699] 2 0 1 2 1 2 2 6 1 2 ...  
## $ bb : num [1:21699] 2 4 2 3 1 4 9 16 1 1 ...  
## $ so : num [1:21699] 1 0 0 0 0 1 1 3 1 0 ...  
## $ ibb : num [1:21699] NA NA NA NA NA NA NA NA NA NA ...  
## $ hbp : num [1:21699] NA NA NA NA NA NA NA NA NA NA ...  
## $ sh : num [1:21699] NA NA NA NA NA NA NA NA NA NA ...  
## $ sf : num [1:21699] NA NA NA NA NA NA NA NA NA NA ...  
## $ gidp : num [1:21699] NA NA NA NA NA NA NA NA NA NA ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ...1 = col\_double(),  
## .. id = col\_character(),  
## .. year = col\_double(),  
## .. stint = col\_double(),  
## .. team = col\_character(),  
## .. lg = col\_character(),  
## .. g = col\_double(),  
## .. ab = col\_double(),  
## .. r = col\_double(),  
## .. h = col\_double(),  
## .. X2b = col\_double(),  
## .. X3b = col\_double(),  
## .. hr = col\_double(),  
## .. rbi = col\_double(),  
## .. sb = col\_double(),  
## .. cs = col\_double(),  
## .. bb = col\_double(),  
## .. so = col\_double(),  
## .. ibb = col\_double(),  
## .. hbp = col\_double(),  
## .. sh = col\_double(),  
## .. sf = col\_double(),  
## .. gidp = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

## Visualizations

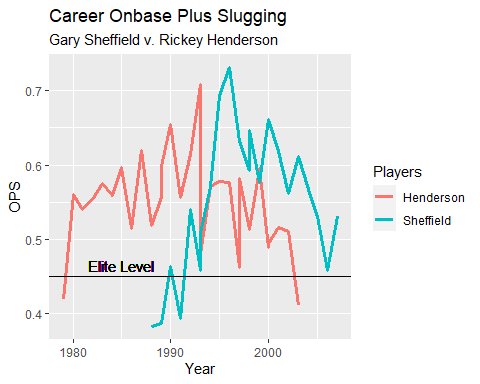
The first visual is a representation of Gary Sheffield’s batting average over his career compared to HOFer Ken Griffey Jr., who played over the same span.

Figure 1 

Next, we will look at another HOFer and teammate on the San Diego Padres, Tony Gwynn, and compare their home run numbers.

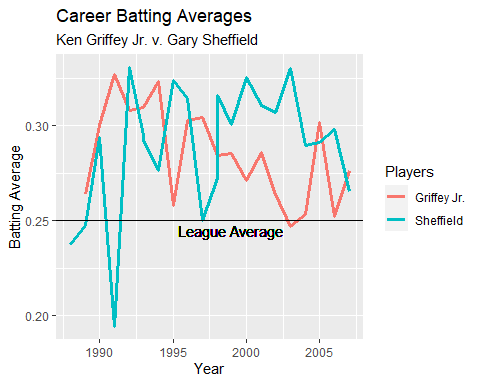
Figure 2 

Finally we will compare Sheffield’s on base plus slugging percent, or OPS, compared to another Hall of Fame Inductee who played in that era, Rickey Henderson.

Figure 3 

## Analysis

Looking at Sheffield’s numbers compared to Hall of Fame Inductees, we can see that he has comparable numbers. Sheffield played at or above an elite level on the offensive side of the ball throughout his entire career.



Looking at Figure 1, we can see that Sheffield’s batting average was comparable to Ken Griffey Jr., and well above the league average of 0.250 for several seasons. Sheffield had a career high batting average of .330 and a career average of .248, with an injured 1991 season bringing the average down.

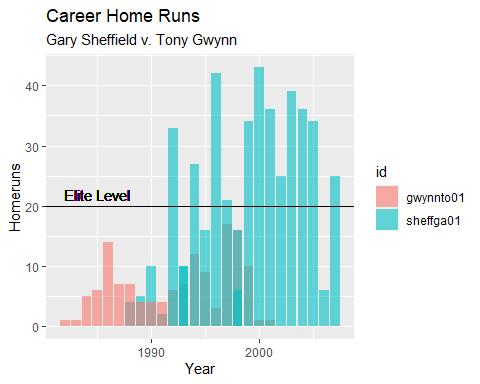
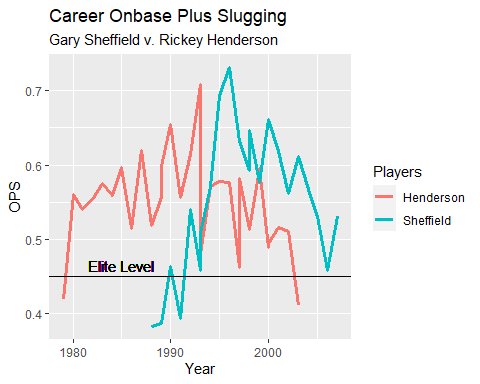


Figure 2 shows that Sheffield had 12 seasons with more than 20 home runs in the season. He also had more home runs than Tony Gwynn, a Hall of Famer. Sheffield hit 509 homeruns over his career, being 1 of 28 players to hit 500 home runs. 14 members of the 500 HR club are in the Hall of Fame. Of the other half, 2 players are active, and 11 players have been accused or found guilty of using steroids. That leaves only Gary Sheffield.



Examining Figure 3 tells us a lot. Rickey Henderson did not play all the same seasons but they had over a decade of overlap. While Henderson, a Hall of Famer, had a more consistent career in terms of OPS, Sheffield still had many seasons with OPS at an elite level, with only 3 seasons below .450. Sheffield also ranks 63rd in career OPS in the MLB, just below Ken Griffey Jr.

## Conclusion

Gary Sheffield has appeared on the Hall of Fame ballot for eight years, with only two more years remaining before he gets left off the ballot and misses out on joining the greats in Cooperstown. Using only the above statistics and visuals, you can see that Sheffield makes a compelling case for entering the halls of Cooperstown. He has comparable numbers to Hall of Famers and has played at an elite level on offense for the majority of his career. If I had a vote, Gary Sheffield, the 9 time All-Star, 5 time Silver Slugger, and World Champ, would definitely make it into the Hall of Fame.