Times HR/BB above Season Average

Alfredo Nasiff

September 13, 2019

Introduction

The barrage of Home runs this 2019 MLB season has produced a stratospheric number of records broken. Each year has its own peculiarities, though not as remarkable as this 2019, when even the MLB was prompted to admit that some changes were made to the ball. If we look at the HR/AB mean per season we would see the differences from one year to another that, independently of the many causes, must be taken into consideration by statisticians when drawing comparisons among sluggers from different decades.

Dataset

Using the Batting table of the Lahman Database from R programming language:

```
HR_Average_Season <- Batting %>%
  select(playerID, yearID, AB, HR) %>%
  group_by(yearID) %>%
  mutate(HR_Ave = HR/AB,
         HR_Sea_Ave = sum(HR)/sum(AB),
         HR_Sea_Ave_Rate = HR_Ave / HR_Sea_Ave)

season_avg <- HR_Average_Season %>%
  select(yearID, HR_Sea_Ave) %>%
  group_by(yearID) %>%
  ggplot(aes(yearID, HR_Sea_Ave)) +
  geom_line() + geom_point() + theme_economist() +
  xlab("Season") + ylab("HR/AB Season Mean") + ggtitle("HR/AB mean per Season")

plot(season_avg)
```

HR/AB mean per Season
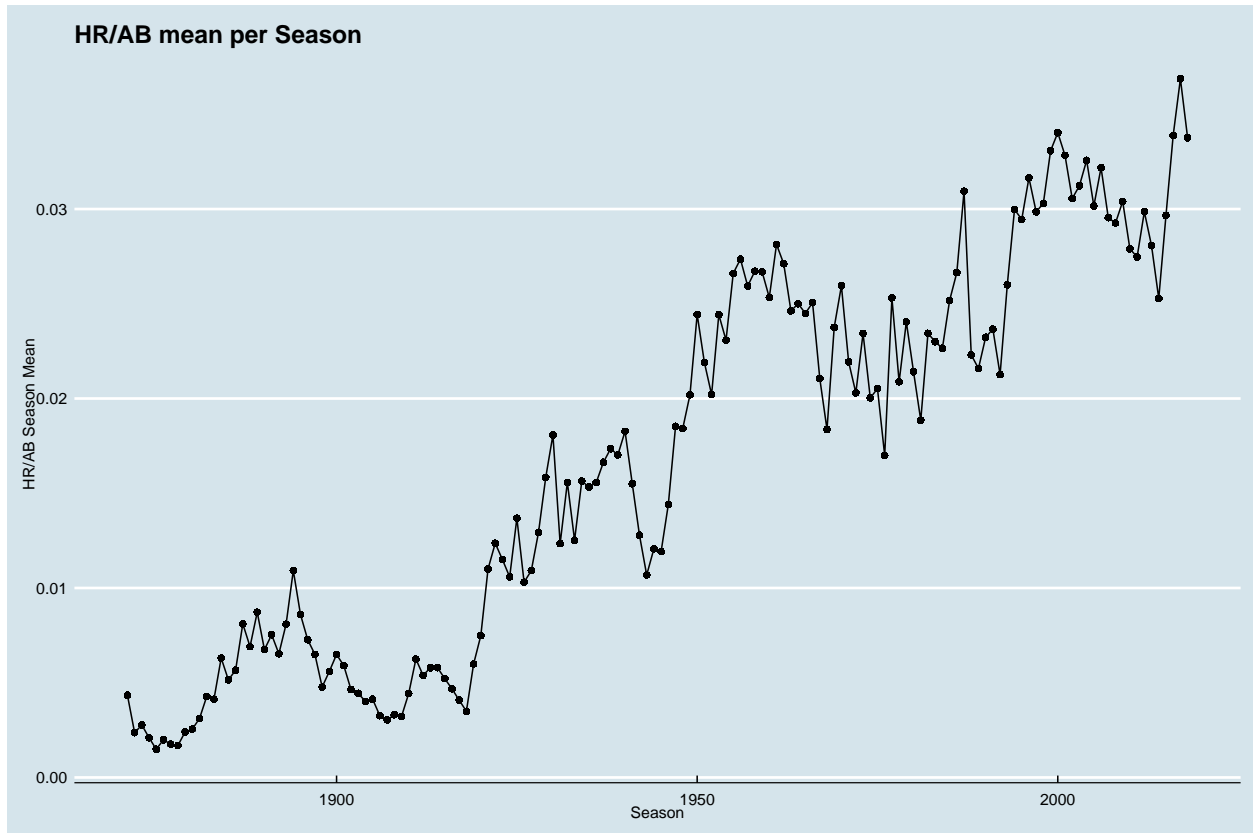
The trend indicates that nowadays is easier to hit Home runs than in the beginning of MLB history. We can speculate that if Babe Ruth would have played in modern times the total of his Home runs would have been higher, but we will never know that. Nevertheless, comparisons can be made among players from different times, weighing in their performances against the rest of the players of each Season played, meaning dividing the individual performance against the Season mean of the parameter we are studying. The purpose of this paper is to make the comparison of the greatest Home run sluggers of all times against the mean of each season played by them.

Analysis and Results

We begin by presenting the case of Babe Ruth as an example. The following graph shows HR/AB by year with three lines drawn, one represents Babe Ruth's index, another one represents the mean of the Season and the last one the Rate, which is [Babe Ruth (HR/AB)]/ [Season (HR/AB)]. In 1920 Babe Ruth HR/AB index was 0.118, the Season mean was 0.007, therefore he had a frequency of HR/AB 15.8 times above average (the Rate).

```
HR_Ave_Sea_BabeRu01 <- HR_Average_Season %>%
  filter(playerID == "ruthba01")

# The chart
previous <- HR_Ave_Sea_BabeRu01 %>%
  ggplot(aes(x = yearID)) +
  geom_line(aes(y = HR_Ave, color = 'HR_Ave'), size = 1) +
  geom_line(aes(y = HR_Sea_Ave, color = 'HR_Sea_Ave'), size = 1) +
  geom_line(aes(y = HR_Sea_Ave_Rate, color = 'HR_Sea_Ave_Rate'), size = 1) +
  geom_point(aes(y = HR_Ave, color = 'HR_Ave'), size = 1.5) +
```

```
geom_point(aes(y = HR_Sea_Ave, color = 'HR_Sea_Ave'), size = 1.5) +
geom_point(aes(y = HR_Sea_Ave_Rate, color = 'HR_Sea_Ave_Rate'), size = 1.5) +
geom_text(aes(y = HR_Ave, label = ifelse(yearID == '1920', round(HR_Ave,3), '')),
          nudge_y = 0.2) +
geom_text(aes(y = HR_Sea_Ave, label = ifelse(yearID == '1920', round(HR_Sea_Ave,3),'')),
          nudge_y = -0.2) +
geom_text(aes(y = HR_Sea_Ave_Rate, label = ifelse(yearID == '1920',
                                                   round(HR_Sea_Ave_Rate,1), '')),
          nudge_y = 0.2) +
theme_economist() +  scale_y_log10() +
xlab("year") + ylab("HR Rate") + ggtitle("Babe Ruth")

plot(previous)
```
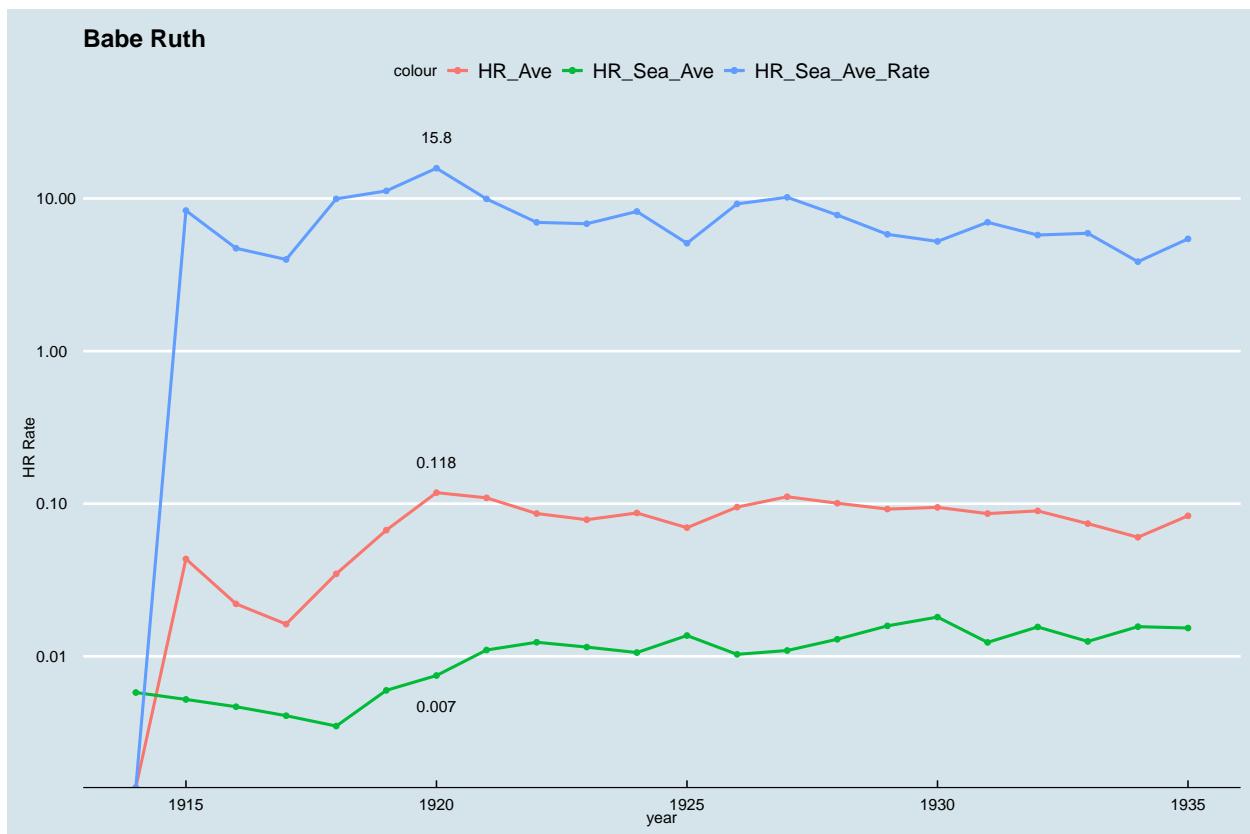

Adding up the Rate index accumulated by Babe Ruth along his entire career will make a total of 157.2 times above Seasons mean:

```
final <- HR_Ave_Sea_BabeRu01 %>%
  ggplot(aes(x = yearID, y = HR_Sea_Ave_Rate, label = round(HR_Sea_Ave_Rate,1))) +
  geom_line() +
  geom_point() +
  theme_economist() +
  xlab("year") + ylab("HR Rate") + ggtitle("Babe Ruth") +
  annotate('text', x = 1930, y = 14, label = 'sum = 157.2', color = 'blue') +
  geom_text_repel()

plot(final)
```
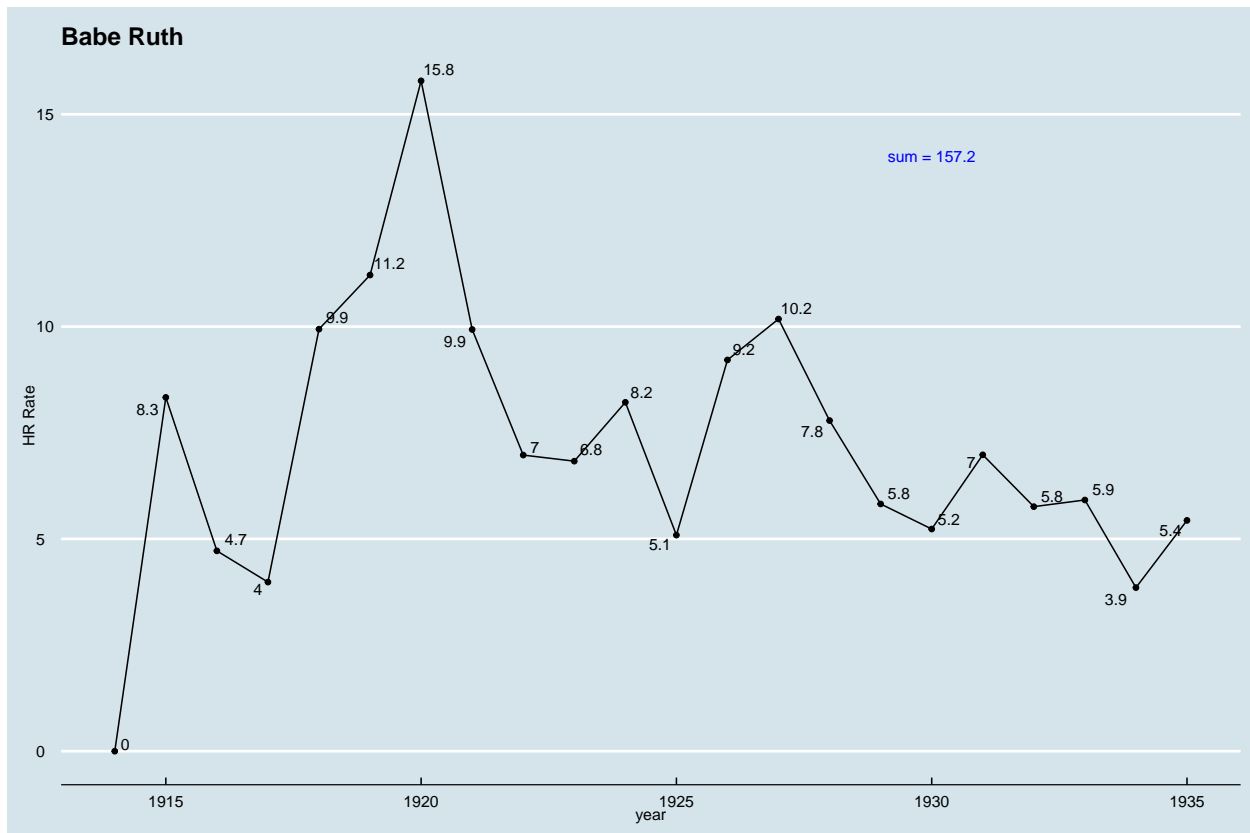
How does this compare with the other players listed among the top HR sluggers, the following table shows it in descending order by number of HR:

```
Best_HR <- HR_Average_Season %>%
  group_by(playerID) %>%
  summarise(AB = sum(AB),
            HR = sum(HR),
            HRperAB_Mean = round(sum(HR_Sea_Ave_Rate),digits = 1))%>%
  filter(AB>2999)

Master_PlayerName <- Master[,c("playerID","nameFirst","nameLast","debut","finalGame",
                                "weight","height")]

Merge_Tables <- merge(Best_HR,Master_PlayerName,by = "playerID")%>%
  mutate(Name = paste(nameFirst, nameLast, sep = " "),
        Years = paste(substr(debut,0,4),substr(finalGame,0,4),sep = "-")) %>%
  select(Name, Years,AB,HR, HRperAB_Mean)%>%
  arrange(-HR)

kable(Merge_Tables[1:42,])
```

Name	Years	AB	HR	HRperAB_Mean
Barry Bonds	1986-2007	9847	762	61.2
Hank Aaron	1954-1976	12364	755	59.3
Babe Ruth	1914-1935	8398	714	157.2
Alex Rodriguez	1994-2016	10566	696	43.0

Name	Years	AB	HR	HRperAB_Mean
Willie Mays	1951-1973	10881	660	51.7
Albert Pujols	2001-2018	10196	633	36.6
Ken Griffey	1989-2010	9801	630	46.7
Jim Thome	1991-2012	8422	612	51.5
Sammy Sosa	1989-2007	8813	609	39.4
Frank Robinson	1956-1976	10006	586	53.9
Mark McGwire	1986-2001	6187	583	57.4
Harmon Killebrew	1954-1975	8147	573	54.5
Rafael Palmeiro	1986-2005	10472	569	36.8
Reggie Jackson	1967-1987	9864	563	51.3
Manny Ramirez	1993-2011	8244	555	41.2
Mike Schmidt	1972-1989	8352	548	50.2
David Ortiz	1997-2016	8640	541	37.4
Mickey Mantle	1951-1968	8102	536	47.8
Jimmie Foxx	1925-1945	8134	534	71.5
Willie McCovey	1959-1980	8197	521	56.9
Frank Thomas	1990-2008	8199	521	43.1
Ted Williams	1939-1960	7706	521	67.4
Ernie Banks	1953-1971	9421	512	40.7
Eddie Mathews	1952-1968	8537	512	43.4
Mel Ott	1926-1947	9456	511	70.4
Gary Sheffield	1988-2009	9217	509	42.8
Eddie Murray	1977-1997	11336	504	39.4
Lou Gehrig	1923-1939	8001	493	63.7
Fred McGriff	1986-2004	8757	493	41.0
Adrian Beltre	1998-2018	11068	477	29.1
Stan Musial	1941-1963	10972	475	43.2
Willie Stargell	1962-1982	7927	475	50.5
Carlos Delgado	1993-2009	7283	473	31.9
Chipper Jones	1993-2012	8984	468	30.0
Miguel Cabrera	2003-2018	8456	465	28.0
Dave Winfield	1973-1995	11003	465	39.5
Jose Canseco	1985-2001	7057	462	44.5
Adam Dunn	2001-2014	6883	462	34.7
Carl Yastrzemski	1961-1983	11988	452	38.0
Jeff Bagwell	1991-2005	7797	449	28.1
Vladimir Guerrero	1996-2011	8155	449	27.4
Dave Kingman	1971-1986	6677	442	58.6

At first glance one figure jumps out from the table: Babe Ruth "(HR/AB) / Season (HR/AB)" more than double the next player in the list.

If we look carefully, will see that in recent years it is tougher for players to excel above the mean, in fact, the only two active players on the list are doing very badly in the Index, which can be explained by the rise in the mean, trend that in turn has multifactorial causes. Among them we have the hypothesis that these days players hit more Home runs thanks to the rise in competitiveness due to the larger number of players, Teams, Leagues and Training Camps, also the advancements made in the technology applied to enhance performance plays a major role.

Can we measure these factors like competitiveness and enhanced performance? The proposition is to use the weight and height of the players as a expression of how these factors have improved their pysical traits and therefore quantify how has this affected the media of HR per Season.

It is proposed to calculate the "Strength" of players adding both their height and weight, reasoning that the taller and corpulent the player the more prepared and competitive is. Plotting the mean of each season, the graph looks like this:
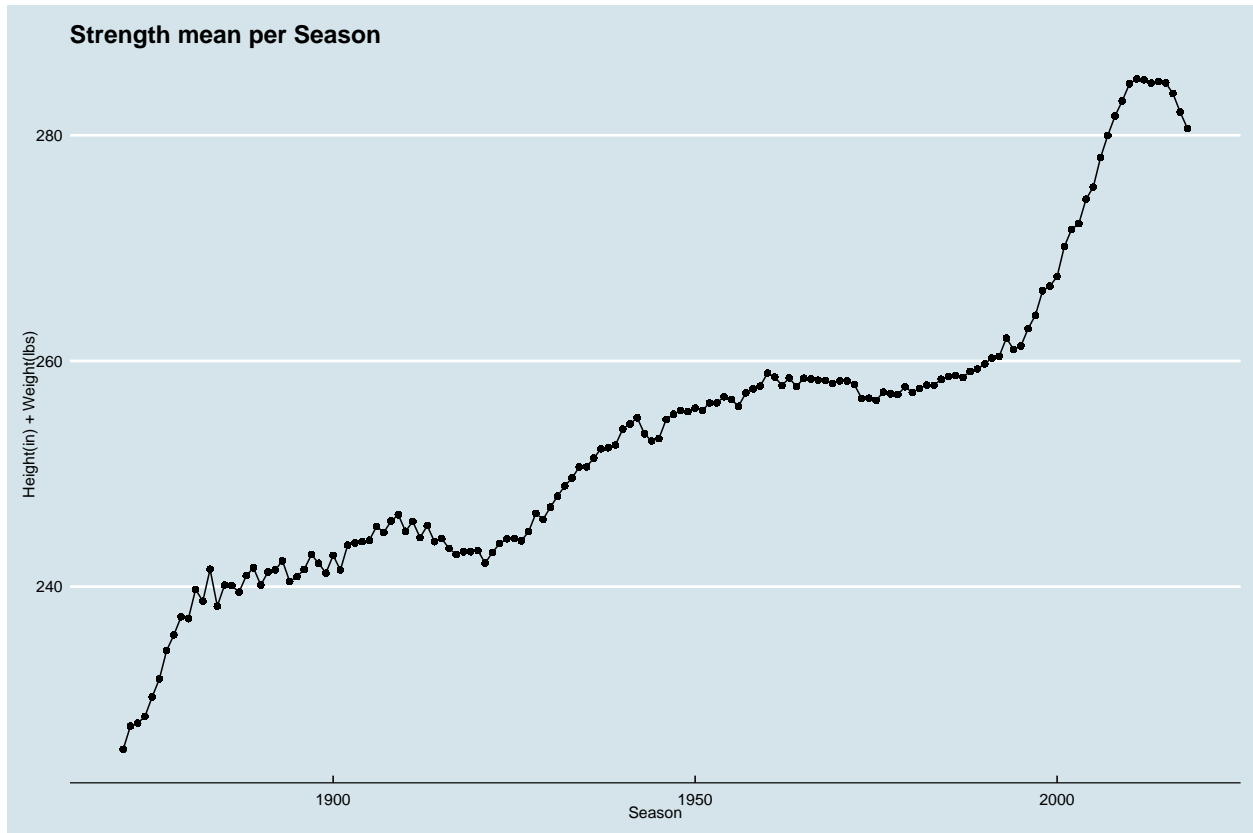
It is effectively seen that in recent times the players are stronger. In 1920 the Strength mean was 243.2 while Babe's Strength was 289, taking over 45 points of advantage. In 2011, the year Mike Trout debuted with a Strength of 309, the mean topped the all-time list with 285, a meager 24 points below.

Recalculating the "HRperAB_Mean" Rate dividing it by the "Strength_Mean" resulting in "HR_Strength", shows the difference in "HR_Strength_diff":

```r
Merge_Tables_4.1 <- Merge_Tables_4.0 %>%
  group_by(playerID) %>%
  summarise(AB = sum(AB),
            HR = sum(HR),
            HRperAB_Mean = round(sum(HR_Sea_Ave_Rate),digits = 1),
            Strength_Mean = round(sum(Strength_Sea_Ave_Rate),digits = 1),
            HR_Strength = round(sum(HR_Rate_vs_Strength_Rate),
                                              digits = 1))%>%
  filter(AB>2999)

Merge_Tables_5 <- merge(Merge_Tables_4.1,Master_PlayerName,by = 'playerID') %>%
  mutate(Name = paste(nameFirst, nameLast, sep = " "),
         Years = paste(substr(debut,0,4),substr(finalGame,0,4),sep = "-"),
         HR_Strength_diff = HR_Strength -
           HRperAB_Mean) %>%
  select(Name,Years,AB,HR,HRperAB_Mean,Strength_Mean,HR_Strength,
         HR_Strength_diff) %>%
```

```
arrange(-HR)

kable(Merge_Tables_5[1:42,])
```

Name	Years	AB	HR	HRperAB_Mean	Strength_Mean	HR_Strength	HR_Strength_diff
Barry Bonds	1986-2007	9847	762	61.2	21.4	63.3	2.1
Hank Aaron	1954-1976	12364	755	59.3	22.5	60.7	1.4
Babe Ruth	1914-1935	8398	714	157.2	25.9	133.2	-24.0
Alex Rodriguez	1994-2016	10566	696	43.0	24.4	38.8	-4.2
Willie Mays	1951-1973	10881	660	51.7	21.4	55.5	3.8
Albert Pujols	2001-2018	10196	633	36.6	20.3	32.5	-4.1
Ken Griffey	1989-2010	9801	630	46.7	23.0	46.4	-0.3
Jim Thome	1991-2012	8422	612	51.5	29.9	43.2	-8.3
Sammy Sosa	1989-2007	8813	609	39.4	16.9	44.4	5.0
Frank Robinson	1956-1976	10006	586	53.9	21.9	54.2	0.3
Mark McGwire	1986-2001	6187	583	57.4	18.9	51.7	-5.7
Harmon Killebrew	1954-1975	8147	573	54.5	22.8	52.6	-1.9
Rafael Palmeiro	1986-2005	10472	569	36.8	19.1	38.7	1.9
Reggie Jackson	1967-1987	9864	563	51.3	21.8	49.5	-1.8
Manny Ramirez	1993-2011	8244	555	41.2	22.9	37.7	-3.5
Mike Schmidt	1972-1989	8352	548	50.2	18.8	48.1	-2.1
David Ortiz	1997-2016	8640	541	37.4	22.0	34.2	-3.2
Mickey Mantle	1951-1968	8102	536	47.8	18.6	46.3	-1.5
Jimmie Foxx	1925-1945	8134	534	71.5	22.4	67.1	-4.4
Willie McCovey	1959-1980	8197	521	56.9	24.4	53.6	-3.3
Frank Thomas	1990-2008	8199	521	43.1	23.6	36.5	-6.6
Ted Williams	1939-1960	7706	521	67.4	20.8	61.6	-5.8
Ernie Banks	1953-1971	9421	512	40.7	18.6	41.4	0.7
Eddie Mathews	1952-1968	8537	512	43.4	18.4	42.5	-0.9
Mel Ott	1926-1947	9456	511	70.4	21.0	73.9	3.5
Gary Sheffield	1988-2009	9217	509	42.8	23.4	43.8	1.0
Eddie Murray	1977-1997	11336	504	39.4	23.4	38.6	-0.8
Lou Gehrig	1923-1939	8001	493	63.7	18.6	58.1	-5.6
Fred McGriff	1986-2004	8757	493	41.0	21.9	39.2	-1.8
Adrian Beltre	1998-2018	11068	477	29.1	22.0	27.9	-1.2
Stan Musial	1941-1963	10972	475	43.2	21.2	44.8	1.6
Willie Stargell	1962-1982	7927	475	50.5	21.4	49.7	-0.8
Carlos Delgado	1993-2009	7283	473	31.9	18.2	29.8	-2.1
Chipper Jones	1993-2012	8984	468	30.0	19.9	28.6	-1.4
Miguel Cabrera	2003-2018	8456	465	28.0	18.5	24.3	-3.7
Dave Winfield	1973-1995	11003	465	39.5	26.5	34.3	-5.2
Jose Canseco	1985-2001	7057	462	44.5	22.9	36.9	-7.6
Adam Dunn	2001-2014	6883	462	34.7	20.8	26.8	-7.9
Carl Yastrzemski	1961-1983	11988	452	38.0	22.0	39.8	1.8
Jeff Bagwell	1991-2005	7797	449	28.1	15.0	28.0	-0.1
Vladimir Guerrero	1996-2011	8155	449	27.4	18.1	24.1	-3.3
Dave Kingman	1971-1986	6677	442	58.6	21.3	52.4	-6.2

The largest differences were accounted by Babe Ruth (-24) who still almost double his closest tracker (Mel Ott, who displaced Jimmie Foxx of the second place thanks to his low 239 of Strength) and Sammy Sosa (+5). So, the physical traits of Babe Ruth (74 in + 215 lbs = 289) impacted negatively in his "Times HR/AB

over Season mean", as the other players of his time were in physical disadvantage with him.

Conclusions

Babe Ruth, despite this later skirmish using the Strength statistic, seems to be once again, immovable as the Greatest Player of All-Time. Big names show up topping the list of the "Times HR_Strength": 1-Babe Ruth; 2-Mel Ott; 3-Jimmie Foxx; 4-Barry Bonds; 5-Ted Williams; 6-Hank Aaron; 7-Lou Gehrig; 8-Willie Mays; 9-Frank Robinson; 10-Willie McCovey; 11-Harmon Killebrew. Make your own judgment.