

Are Hall of Famers Getting Better Over Time?
Final Report for Exploratory Data Analysis Class
Eric Webb
Ph.D. Student in Decision Sciences
December 9, 2013

Executive Overview

The caliber of players elected to the National Baseball Hall of Fame appears to be increasing over time. Players tend to have more career hits and home runs over time. This analysis seeks to investigate this trend to discover its cause. Upon review, it appears that much of the trend in increased caliber of Hall of Fame players can be explained by the fact that recent second-tier Hall of Fame players have not yet been inducted. Once these potential future candidates are granted admission, most of the trend in increased career hits over time disappears. The trend of increased home runs over time remains, but is largely due to the general increase across Major League Baseball in home run hitting over time.

Introduction

The National Baseball Hall of Fame (HOF), headquartered in Cooperstown, NY, has enshrined the achievements of the best baseball players since its inception in 1936. As of January 2013, 300 people have been elected to the HOF, including 208 former Major League Baseball players, 19 managers, 10 umpires, 28 pioneers, executives, and organizers, and 35 Negro league baseball players and executives [1]. Because statistics are readily available only for Major League Baseball (MLB) players, the analysis presented here will focus on these HOF members. Only about 1% of players who have ever played in MLB are in the HOF [2]. To be elected to the HOF, players must have at least ten years of major league experience. They must be retired for at least five years or deceased for at least six months. If they meet these requirements, players can be voted into the HOF by the Baseball Writers' Association of America. If, after 15 years of voting by the Baseball Writers' Association of America, they have not gained entry into the HOF, they may be elected by the Veterans Committee [1]. This description over-simplifies the election process a bit, and the rules have changed over time. However, there has always been a formal election process focused upon electing the best ballplayers from any given era.

MLB has changed significantly over time. For example, fielding gloves were not used when MLB was founded in 1869. It was rare for players to hit home runs before the 1930's. Black players were not allowed in MLB until 1947. The pitching mound was lowered in 1969 to reduce the dominance of pitchers. Home run frequency shot up in the 1990's, which many people credit to illegal steroid use among some players. With all these changes, it is often difficult to directly compare different players from different eras.

The statistics that analysts and fans use to track the game have also changed greatly over time. While the staple statistics of hits and home runs have been around since MLB's inception, more nuanced statistics such as sacrifice fly balls, pitcher saves, and intentional walks have been added over time. Recently, analysts have begun to combine traditional statistics in ways that allow them to better estimate the value of players. Wins above replacement player (WAR) is a metric developed to estimate the value of a player's total contribution to their team. It estimates the number of additional games the player's team won by using the player at his position instead of replacing him with a replacement-level player. Different websites calculate WAR in different ways, though most take batting, pitching, and fielding statistics into account. WAR is also adjusted by year, as league averages and the quality of replacement players vary by year. Baseball-Reference.com is one website that publishes WAR statistics [2]. Figure 1 shows the career WAR for all non-pitcher MLB HOF members [2], plotted against the mid-points of the players' careers. Two players who played in the Negro leagues for multiple years before moving to MLB were removed as outliers due to their limited time in MLB.

Figure 1 suggests that the value of the players who have gained admittance to the HOF has increased over time. With a best-fitting line having a slope of 0.2, players who played in the 1990's are worth about 20 more wins for their teams than players who played in the 1890's. Since the average Hall of Famer (HOFer) has a career WAR of 66.5, a difference of 20 is quite large. What is causing the apparent increase in player value of HOFers over time? Are the best players becoming increasingly better? Is the HOF being more selective in its admissions and holding players to a higher standard? Are fewer players being elected over time? The goal of this report is to discover what is driving the increase in player value among HOFers over time.

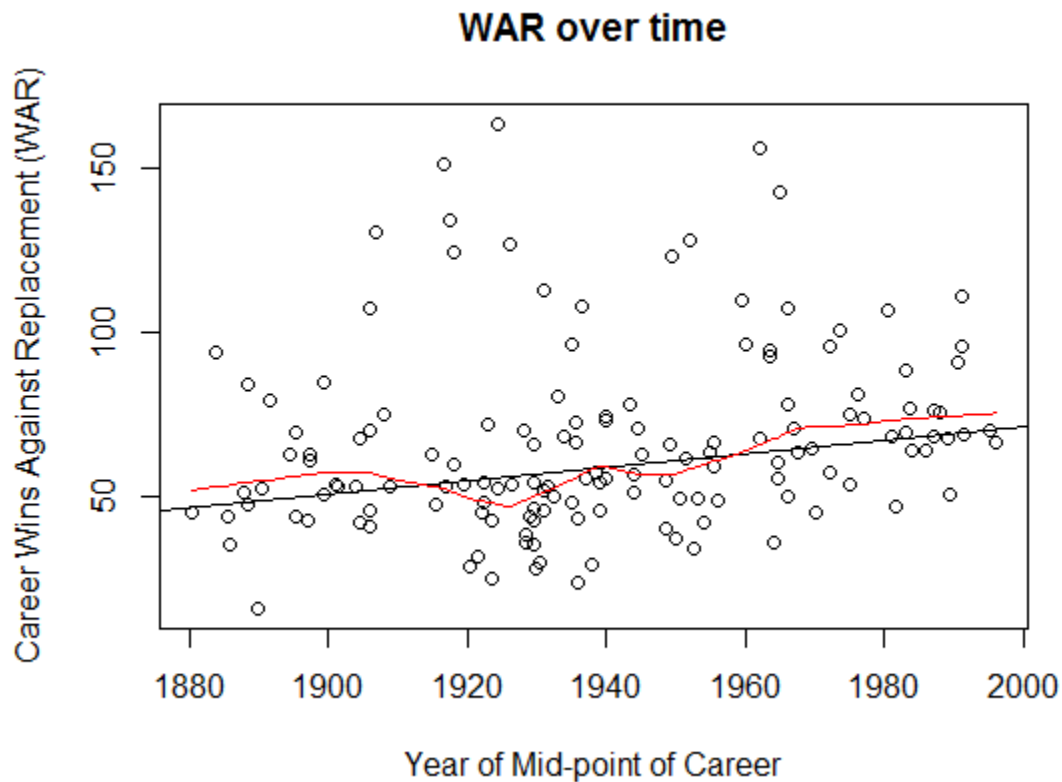


Figure 1: Career WAR of non-pitcher MLB HOFers plotted against the year of the mid-point of a player's career. The best-fitting robust-resistant line is shown in black and has a slope of .2. An effort to smooth the points with loess ($f=.3$) is shown in red and agrees well with the linear fit.

Data Set

The statistics for this analysis were taken from Baseball-Reference.com [2]. This analysis focused upon batting statistics, as opposed to pitching statistics, so only non-pitchers are included. After removing two HOFers who spent much of their careers in the Negro Leagues before moving to MLB, there are 146 MLB non-pitcher HOFers in this data set. Baseball-Reference.com provides the career starting year, career ending year, and year inducted into the HOF. It also provides career WAR, hits, and home runs, among other batting statistics. Data for all 146 players are shown in Appendix A.

While WAR was a motivating statistic for this project, the analysis from this point will focus on two traditional statistics: Hits and Home Runs. This is because the definition of WAR changes over time and is not standardized across sources. Hits and home runs, however, are defined in the same way now as they were when MLB was founded. The number of league-leading hits in a season has been fairly consistent all the way back to the 1880's at around 200 +/- 40. Home runs have varied more between eras of baseball, with home runs before 1920 being very rare.

Figure 2 gives an overview of hits and home runs among HOFers. Generally speaking, career hits have moderately increased over time among HOFers and career home runs have greatly increased over time. While not all hitters are power hitters who regularly hit home runs, almost every positional player in the HOF was very good at getting hits.

Home runs, as shown in Figure 2A and 2C, varied greatly among HOFers. As described in Appendix B, taking the logarithm of career home runs simplifies comparisons across time by eliminating variation in spread of the home run distribution across time and by removing the right-skewness of the home run distribution. For the rest of this analysis, we will use the logarithm of career home runs by HOF instead of unscaled career home runs.

Career hits, as described in Appendix C, do not suffer from the same distributional problems as career home runs. We will not scale career hits.

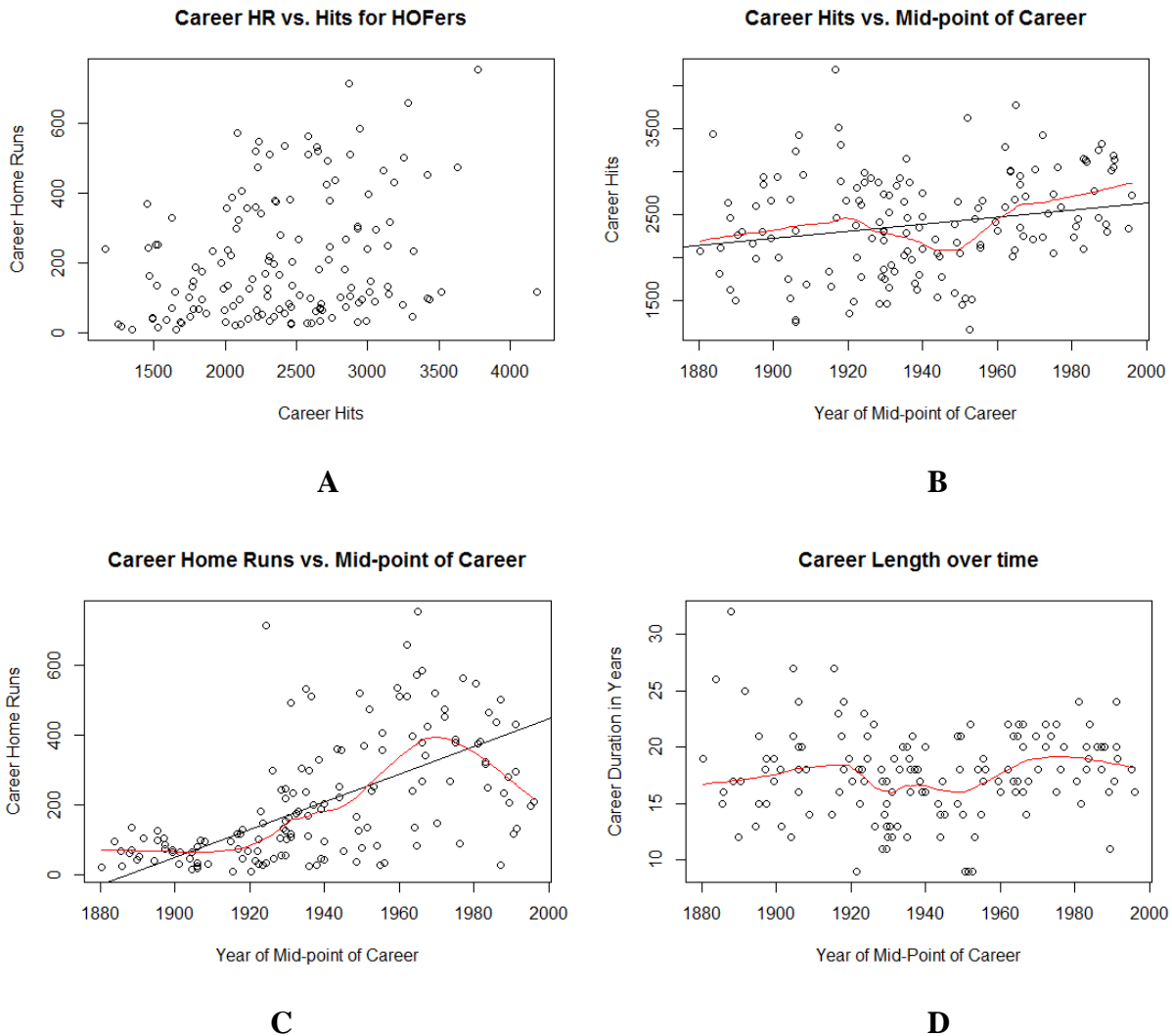


Figure 2: **A** shows a scatterplot of career home runs against career hits for HOFers. There is a large variability in home run numbers, as certain baseball eras did not hit many home runs and many players were not power hitters. **B** shows career hits vs. the year of the mid-point of the player’s career. The best-fitting robust-resistant line, in black, has a slope of 4.1. Shown in red is a smoothing with loess ($f=.3$).

The drop from the linear fit around 1935-1955 is partially due to an era of pitching dominance and partially due to many HOF players losing multiple years while they served in World War II. **C** shows career home runs vs. the year of the mid-point of the player’s career. The best fitting robust-resistant line, in black, has a slope of 4.0. Shown in red is a smoothing with loess ($f=.3$). The smoothing follows the line until perhaps 1980. Many of the recent admissions to the HOF are not power hitters. With the cloud of steroid use hanging over the heads of many sluggers from the 1990’s, there are many power hitting players who have not yet been elected to the HOF. **D** shows career length over time. Career duration is fairly consistent over time, with a slight drop around WWII. A loess smoother ($f=.3$) is shown in red.

Analysis

With career hits and $\log(\text{career home runs})$ as our statistics of interest, we now turn to the question posed above: What is causing the apparent increase in player value of HOFers over time? To answer this question, two related questions must be answered:

- Can we expect more players from recent decades to be elected to the HOF?
- Will the inclusion of more players from recent decades eliminate the upward trends we see in Figure 1, Figure 2B, and Figure 2C?

To answer the first question, let us examine the distribution of the year of the mid-point of HOF careers. Figure 3 shows a stem plot and histogram of these mid-points.

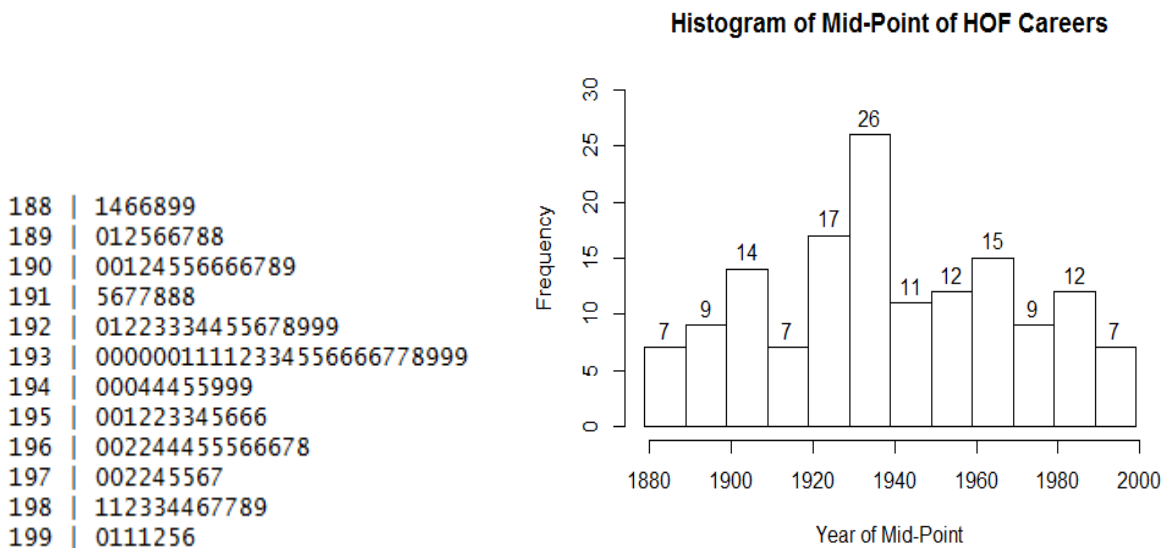


Figure 3: A stem plot of the year of the career mid-point for our 146 HOFers is shown on the left. A histogram of the career mid-points is shown on the right, binning by decade. The most populous decades of HOFers are the 1920's and 1930's.

Figure 3 shows that the most popular decades for HOFers were the 1920's and the 1930's. The number of teams in MLB expanded greatly in the early 1900's, which partially explains why there are only half as many players in the HOF from each decade of the 1800's and early 1900's as the 1920's. After 26 HOFers from the 1930's made it into the HOF, the next three decades provided 11, 12, and 15 HOFers. After the 1970's, though, it seems like fewer players have made the HOF, with just 7 making it from the 1990's. As stated in the introduction, players can be voted into the HOF by the Veteran's Committee long after they have retired. It

could be the case that recent players have not been retired long enough for all of the HOF-worthy players to be elected.

It is simple to examine the amount of time that players had to wait between their retirement and their induction to the HOF. The earliest that most players can enter the HOF is six years after their retirement: 5 years to become eligible and 1 year to get voted in. Because the HOF elected its first class of HOFers in 1936, it seems appropriate to scale the retirement year of early players to be 1930. Thus, if they were elected in the first HOF class ever, in 1936, then their wait until election will be 6 years. Figure 4 shows a scatterplot of time between retirement and election vs. year of retirement.

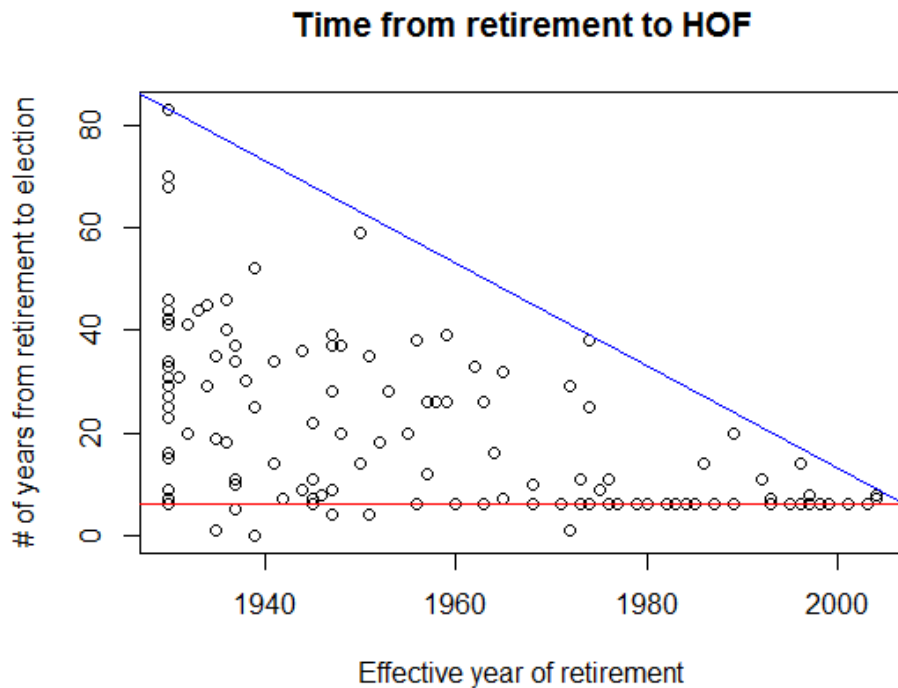


Figure 4: Scatterplot showing the number of years waiting for induction to the HOF against the year of retirement. The x-axis is the effective year of retirement, which is the following: $\max\{1930, \text{year of retirement}\}$. Because the HOF inducted its first class in 1936 and it typically takes at least six years to get inducted, the retirement year for all players who retired before 1930 was scaled to be 1930. The y-axis is the number of years between effective retirement and induction. The red line shows the typical six years that are required to be inducted. The blue line shows the maximum amount of years that a player could be waiting for election in 2013.

Figure 4 suggests that, with more elapsed time, more players who retired after 1980 will be inducted into the HOF. These are likely second-tier HOFers who did not make the initial cut of talented players elected soon after obtaining eligibility for the HOF. To illustrate the relative qualifications of these second-tier HOFers, Figure 5 plots the career hits and home runs of players against the time from their retirement to election.

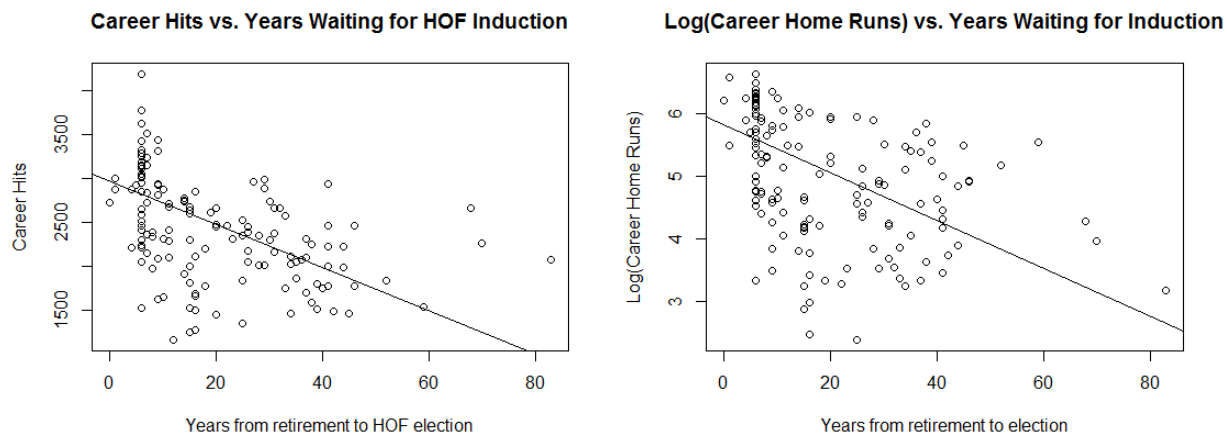


Figure 5: On the left is plotted the number of career hits vs. the years spent waiting for election to the HOF. The best fitting robust-resistant line has a slope of -24.53 , meaning that with each year passing from retirement to election, a player gets about 24 fewer career hits. On the right is plotted $\log(\text{career home runs})$ vs. time waiting for induction. Again, players who wait longer for induction tend to have less impressive career statistics.

To see how career hits and home runs vary both in baseball era and in time from retirement to induction, Appendix D gives a median polish of the data. Players elected in their first year of eligibility (6 years after retirement) tend to get about 1040 more career hits than players elected 20 or more years after retirement. The difference in number of career hits when batching players by the mid-points of their careers (1900-1924, 1925-1949, 1950-1974, 1975-2000) is much less drastic.

Combining the results of Figures 3, 4, and 5 and Appendix D, we see that players who are elected long after their retirement tend to have lower career statistics than players elected as soon as they reach eligibility. It seems likely that more second-tier HOFers will eventually be elected, boosting the number of HOF players who had mid-points of their careers after 1970. These

second-tier HOFers will pull down the average performance of recent HOFers. The performance difference over time suggested in Figures 1 and 2 may disappear.

It is possible to speculate on second-tier players who may eventually be elected to the HOF. When the Baseball Writers' Association of America votes on potential HOFers, a player must receive favorable votes from 75% of the votes to be elected. Any player that receives at least 10% favorable votes is at least a borderline HOF candidate. We can examine the players who have not yet been inducted into the HOF but have received more than 10% of the vote in a single year to find potential future second-tier HOF selections. Figure 6 plots the performance of these potential future HOF candidates against the performance of elected members. These potential future HOF candidates are listed in Appendix E.

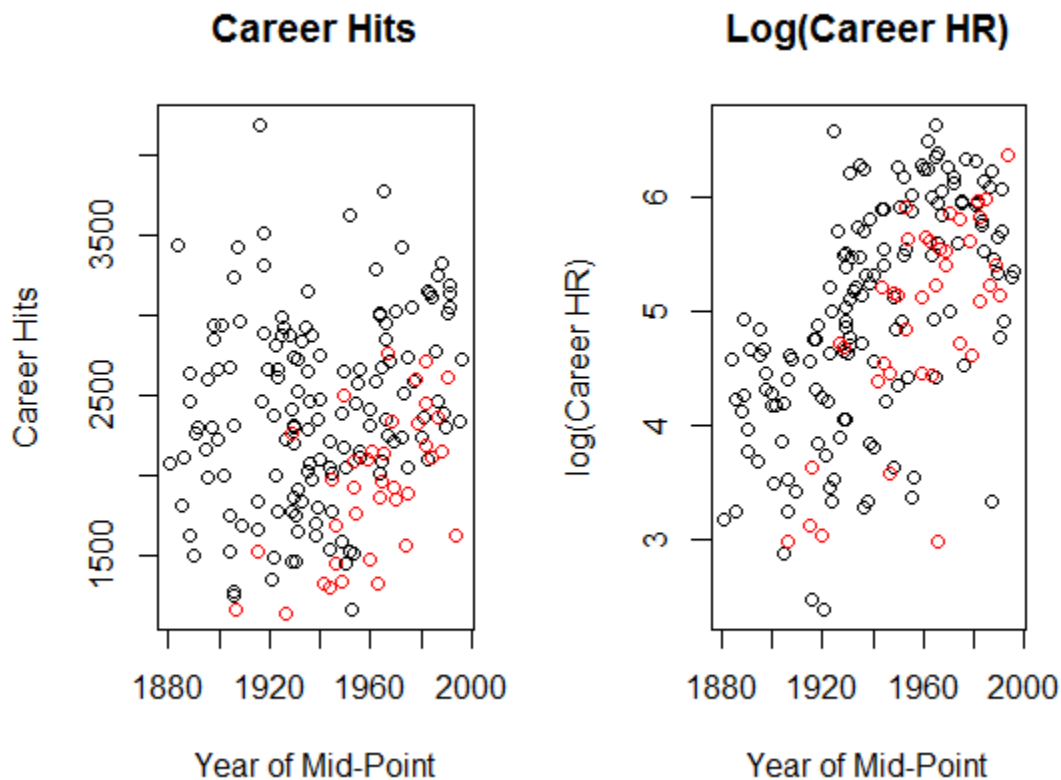


Figure 6: On the left is shown the career hits of HOFers (in black) and potential future second-tier HOF inductees (in red) against the year of the mid-point of their careers. On the right, log(career home runs) of the same players are shown. Second-tier HOF candidates tend to be average or worse at career hits. They tend to be average or worse, for their era, at career home runs.

Once potential second-tier HOFers have been added to Figure 6, it appears that career hits will have a minimal slope over time and $\log(\text{career home runs})$ will still have a large positive slope. The slope of 4.1 for hits that was seen in Figure 2B will largely evaporate when the red points in Figure 6 are included. The best fitting robust-resistant line over all the points on the plot on the left of Figure 6 has a slope of just 2.3. Now, there is no guarantee that all of the players in red in Figure 6 will be inducted eventually. But note the minimal number of players prior to 1950 in the plots of Figure 6. Almost every player whose career mid-point was before 1950 that had a year with 10% or more of the HOF vote has been elected to the HOF. If history is any judge, there is a good chance that most of the players in red in Figure 6 will eventually find their way into the HOF. Once that happens, the trends over time in WAR, hits, and home runs that motivated this analysis will be minimized. Only the considerable trend of home runs increasing over time will remain, and that is largely a product of the league-average home runs increasing over time.

Conclusion

When viewed as a trend, it appears that the quality of the players inducted into the Baseball Hall of Fame (HOF) has increased over time. They have more career hits and home runs (Figure 2). However, many of the second-tier HOF players from recent decades have not yet been elected to the HOF. Over time, second-tier HOF players from before the 1970's have been elected (Figure 4). The second-tier HOFers tend to have slightly worse career statistics than the first-tier HOFers that are elected to the HOF soon after they become eligible (Figure 5). Once many of the potential second-tier HOF candidates have gained admission to the HOF, the trend of increased career hits over time largely disappears. Once all eventual HOF candidates have been elected to the HOF, the caliber of HOF players is only slightly increasing over time. This slight increase over time may be explained by looking at the increased use of statistical analysis over time that tends to keep more of the sub-par candidates out of the HOF as time goes on.

Bibliography

1. National Baseball Hall of Fame and Museum. <http://baseballhall.org/>
2. Baseball-Reference.com. <http://www.baseball-reference.com/>

Appendix A: Statistics for Hall of Famers

Career statistics for 146 MLB non-pitcher HOFers are shown below. Willard Brown and Monte Irvin are not shown below because they spent a significant portion of their careers in the Negro Leagues and thus are outliers with regard to career hits and home runs. 'Inducted' represents the year the player was inducted into the HOF. 'From' represents the first year of the player's MLB playing career and 'To' represents the last year. 'WAR' is career Wins Above Replacement Player, taken from Baseball-Reference.com. 'H' is career hits. 'HR' is career home runs.

Name	Inducted	From	To	WAR	H	HR
Hank Aaron	1982	1954	1976	142.39	3771	755
Roberto Alomar	2011	1988	2004	66.78	2724	210
Cap Anson	1939	1871	1897	93.95	3435	97
Luis Aparicio	1984	1956	1973	55.69	2677	83
Luke Appling	1964	1930	1950	74.46	2749	45
Richie Ashburn	1995	1948	1962	63.40	2574	29
Earl Averill	1975	1929	1941	48.06	2019	238
Home Run Baker	1955	1908	1922	62.65	1838	96
Dave Bancroft	1971	1915	1930	48.51	2004	32
Ernie Banks	1977	1953	1971	67.59	2583	512
Jake Beckley	1971	1888	1907	61.36	2934	87
Johnny Bench	1989	1967	1983	75.16	2048	389
Yogi Berra	1972	1946	1965	59.27	2150	358
Wade Boggs	2005	1982	1999	90.91	3010	118
Jim Bottomley	1974	1922	1937	35.35	2313	219
Lou Boudreau	1970	1938	1952	63.00	1779	68
Roger Bresnahan	1945	1897	1915	40.98	1252	26
George Brett	1999	1973	1993	88.39	3154	317
Lou Brock	1985	1961	1979	45.24	3023	149
Dan Brouthers	1945	1879	1904	79.37	2296	106
Jesse Burkett	1946	1890	1905	62.84	2850	75
Roy Campanella	1969	1948	1957	34.16	1161	242
Rod Carew	1991	1967	1985	81.17	3053	92
Max Carey	1961	1910	1929	53.89	2665	70
Gary Carter	2003	1974	1992	69.84	2092	324
Orlando Cepeda	1999	1958	1974	50.14	2351	379
Frank Chance	1946	1898	1914	45.76	1274	20
Fred Clarke	1945	1894	1915	67.73	2678	67
Roberto Clemente	1973	1955	1972	94.26	3000	240
Ty Cobb	1936	1905	1928	151.08	4189	117
Mickey Cochrane	1947	1925	1937	52.08	1652	119

Name	Inducted	From	To	WAR	H	HR
Eddie Collins	1939	1906	1930	123.99	3315	47
Jimmy Collins	1945	1895	1908	53.27	1999	65
Earle Combs	1970	1924	1935	42.55	1866	58
Roger Connor	1976	1880	1897	84.16	2467	138
Sam Crawford	1957	1899	1917	75.15	2961	97
Joe Cronin	1956	1926	1945	66.38	2285	170
Kiki Cuyler	1968	1921	1938	46.69	2299	128
George Davis	1998	1890	1909	84.77	2665	73
Andre Dawson	2010	1976	1996	64.41	2774	438
Ed Delahanty	1945	1888	1903	69.49	2597	101
Bill Dickey	1954	1928	1946	55.88	1969	202
Joe DiMaggio	1955	1936	1951	78.25	2214	361
Larry Doby	1998	1947	1959	49.37	1515	253
Bobby Doerr	1986	1937	1951	51.26	2042	223
Hugh Duffy	1945	1888	1906	42.90	2293	106
Johnny Evers	1946	1902	1929	47.80	1659	12
Buck Ewing	1939	1880	1897	47.74	1625	71
Rick Ferrell	1984	1929	1947	29.72	1692	28
Carlton Fisk	2000	1969	1993	68.34	2356	376
Elmer Flick	1963	1898	1910	53.12	1752	48
Nellie Fox	1997	1947	1965	49.04	2663	35
Jimmie Foxx	1951	1925	1945	96.46	2646	534
Frankie Frisch	1947	1919	1937	70.38	2880	105
Lou Gehrig	1939	1923	1939	112.47	2721	493
Charlie Gehringer	1949	1924	1942	80.74	2839	184
Joe Gordon	2009	1938	1950	57.11	1530	253
Goose Goslin	1968	1921	1938	66.14	2735	248
Hank Greenberg	1956	1930	1947	57.60	1628	331
Tony Gwynn	2007	1982	2001	68.88	3141	135
Chick Hafey	1971	1924	1937	30.13	1466	164
Billy Hamilton	1961	1888	1901	63.10	2164	40
Gabby Hartnett	1955	1922	1941	53.28	1912	236
Harry Heilmann	1952	1914	1932	72.16	2660	183
Rickey Henderson	2009	1979	2003	110.69	3055	297
Billy Herman	1975	1931	1947	54.54	2345	47
Harry Hooper	1971	1909	1925	53.47	2466	75
Rogers Hornsby	1942	1915	1937	126.97	2930	301
Reggie Jackson	1993	1967	1987	74.00	2584	563
Travis Jackson	1982	1922	1936	44.08	1768	135
Hughie Jennings	1945	1891	1918	42.39	1526	18
Al Kaline	1980	1953	1974	92.65	3007	399
Willie Keeler	1939	1892	1910	54.08	2932	33

Name	Inducted	From	To	WAR	H	HR
George Kell	1983	1943	1957	37.54	2054	78
Joe Kelley	1971	1891	1908	50.53	2220	65
High Pockets Kelly	1973	1915	1932	25.22	1778	148
King Kelly	1945	1878	1893	44.23	1813	69
Harmon Killebrew	1984	1954	1975	60.40	2086	573
Ralph Kiner	1975	1946	1955	49.43	1451	369
Chuck Klein	1980	1928	1944	43.53	2076	300
Nap Lajoie	1937	1896	1916	107.20	3243	82
Barry Larkin	2012	1986	2004	70.18	2340	198
Tony Lazzeri	1991	1926	1939	49.96	1840	178
Freddie Lindstrom	1976	1924	1936	28.38	1747	103
Ernie Lombardi	1986	1931	1947	45.94	1792	190
Mickey Mantle	1974	1951	1968	109.71	2415	536
Heinie Manush	1964	1923	1939	45.76	2524	110
Rabbit Maranville	1954	1912	1935	42.64	2605	28
Eddie Mathews	1978	1952	1968	96.19	2315	512
Willie Mays	1979	1951	1973	156.09	3283	660
Bill Mazeroski	2001	1956	1972	36.11	2016	138
Tommy McCarthy	1946	1884	1896	16.18	1493	44
Willie McCovey	1986	1959	1980	64.44	2211	521
Bid McPhee	2000	1882	1899	52.67	2258	53
Joe Medwick	1968	1932	1948	55.46	2471	205
Johnny Mize	1981	1936	1953	70.91	2011	359
Paul Molitor	2004	1978	1998	75.46	3319	234
Joe Morgan	1990	1963	1984	100.37	2517	268
Eddie Murray	2003	1977	1997	68.19	3255	504
Stan Musial	1969	1941	1963	128.13	3630	475
Jim O'Rourke	1945	1872	1904	51.53	2639	62
Mel Ott	1951	1926	1947	107.89	2876	511
Tony Perez	2000	1964	1986	53.95	2732	379
Kirby Puckett	2001	1984	1995	50.79	2304	207
Pee Wee Reese	1984	1940	1958	66.17	2170	126
Jim Rice	2009	1974	1989	47.21	2452	382
Sam Rice	1963	1915	1934	52.80	2987	34
Cal Ripken	2007	1981	2001	95.59	3184	431
Phil Rizzuto	1994	1941	1956	40.56	1588	38
Brooks Robinson	1983	1955	1977	78.36	2848	268
Frank Robinson	1982	1956	1976	107.11	2943	586
Jackie Robinson	1962	1947	1956	61.39	1518	137
Edd Roush	1962	1913	1931	45.55	2376	68
Babe Ruth	1936	1914	1935	163.20	2873	714
Ryne Sandberg	2005	1981	1997	67.58	2386	282

Name	Inducted	From	To	WAR	H	HR
Ron Santo	2012	1960	1974	70.57	2254	342
Ray Schalk	1955	1912	1929	28.63	1345	11
Mike Schmidt	1995	1972	1989	106.51	2234	548
Red Schoendienst	1989	1945	1963	42.16	2449	84
Joe Sewell	1977	1920	1933	53.75	2226	49
Al Simmons	1953	1924	1944	68.64	2927	307
George Sisler	1939	1915	1930	54.20	2812	102
Enos Slaughter	1985	1938	1959	55.12	2383	169
Ozzie Smith	2002	1978	1996	76.51	2460	28
Duke Snider	1980	1947	1964	66.53	2116	407
Tris Speaker	1937	1907	1928	133.94	3514	117
Willie Stargell	1988	1962	1982	57.29	2232	475
Bill Terry	1954	1923	1936	54.31	2193	154
Sam Thompson	1974	1885	1906	44.13	1988	126
Joe Tinker	1946	1902	1916	53.28	1690	31
Pie Traynor	1948	1920	1937	36.15	2416	58
Arky Vaughan	1985	1932	1948	72.96	2103	96
Honus Wagner	1936	1897	1917	130.59	3420	101
Bobby Wallace	1953	1894	1918	70.09	2309	34
Lloyd Waner	1967	1927	1945	24.11	2459	27
Paul Waner	1952	1926	1945	72.83	3152	113
Monte Ward	1964	1878	1894	35.75	2107	26
Zack Wheat	1959	1909	1927	60.00	2884	132
Deacon White	2013	1871	1890	45.44	2067	24
Billy Williams	1987	1959	1976	63.75	2711	426
Ted Williams	1966	1939	1960	123.16	2654	521
Hack Wilson	1979	1923	1934	38.78	1461	244
Dave Winfield	2001	1973	1995	64.04	3110	465
Carl Yastrzemski	1989	1961	1983	95.98	3419	452
Ross Youngs	1972	1917	1926	32.17	1491	42
Robin Yount	1999	1974	1993	77.12	3142	251

Appendix B: Need for Transformation of Career Home Runs Statistic

Number of career home runs for HOFers varies greatly. As shown in Figure 7, the distribution of career home runs for HOFers has changed greatly over time as well and tends to be right-skewed.

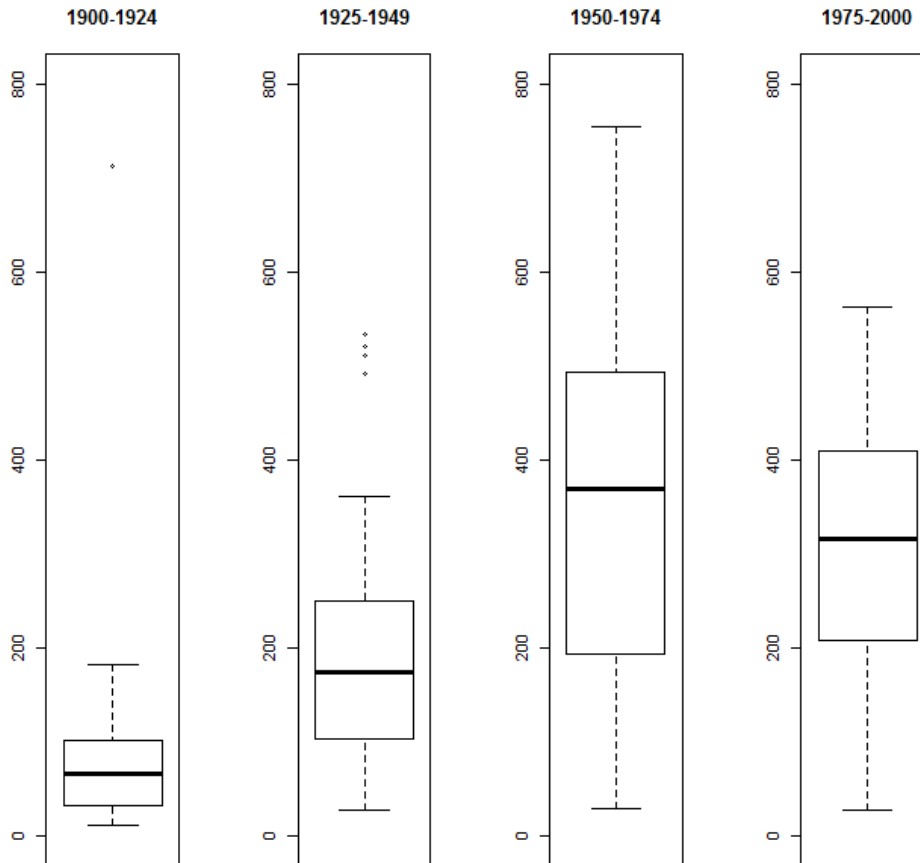


Figure 7: Boxplots of home runs for HOF players, separating players into batches based on the midpoints of their careers (shown above each boxplot). Home runs were relatively rare until the 1930's. Babe Ruth is the high outlier on the boxplot on the left.

The boxplots above suggest that the spread of the home run distribution varies greatly with the median of the batch. The boxplot on the left, for 1900-1924, is much tighter than the boxplots on the right, for 1950-1974 and 1975-2000. To check whether the spread of the home run distribution changes significantly over time, we will split players into 10 batches based on the mid-points of their careers, one for each decade of the 1900's. Shown in Figure 8 is a spread

vs. level plot of the medians and fourth-spreads of the home runs for HOF players in each decade.

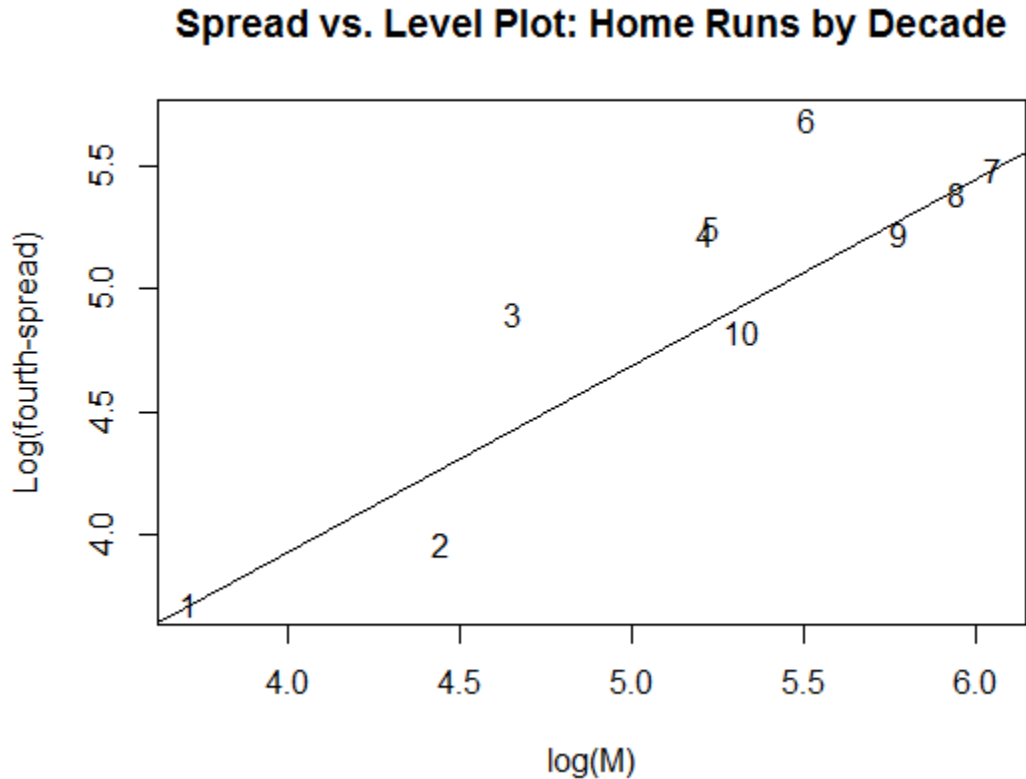


Figure 8: Spread Vs. Level plot of career home runs. Players are batched according to the mid-point of their careers. Those players whose midpoint was between 1900 and 1910 are in batch 1, as shown on the graph. Batches continue until batch 10 shows the players whose mid-points were between 1990 and 2000. The y-axis is the log of the fourth spread of career home runs and the x-axis is the log of the median of career home runs. Clearly, the spread of the home run distribution varies greatly with the level of the median. The best fitting robust-resistant slope of the plot is .761.

Because the slope of the spread vs. level plot is .761, home runs could be transformed by either square roots or logs to lessen or eliminate the variation in spread. The log transformation seemed to have a better effect. In Figure 9, home runs have been transformed by logs and the spread vs. level plot is repeated. This time, there is no significant slope to the plot. Figure 10 recreates the boxplots of Figure 7, showing that the spread of the boxplots no longer varies

significantly with the level of the plots. Additionally, the right-skewness of the career home runs has been eliminated, as shown in Figure 10. For the rest of the analysis, the logarithm of career home runs will be used instead of unscaled career home runs.

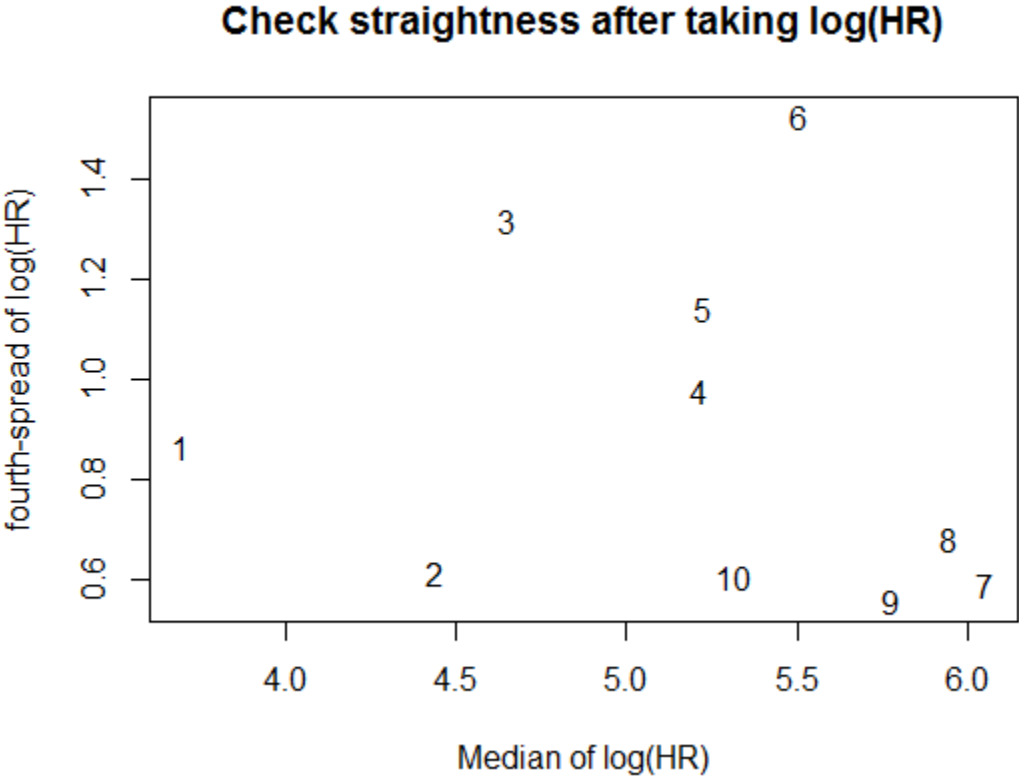


Figure 9: After transforming career home runs by logarithms, the spread vs. level plot shows that there is no significant relationship between the level of career home runs and the variability of career home runs. The numbers in the plot are the batches of HOF players, split by decade in the same manner as Figure 8.

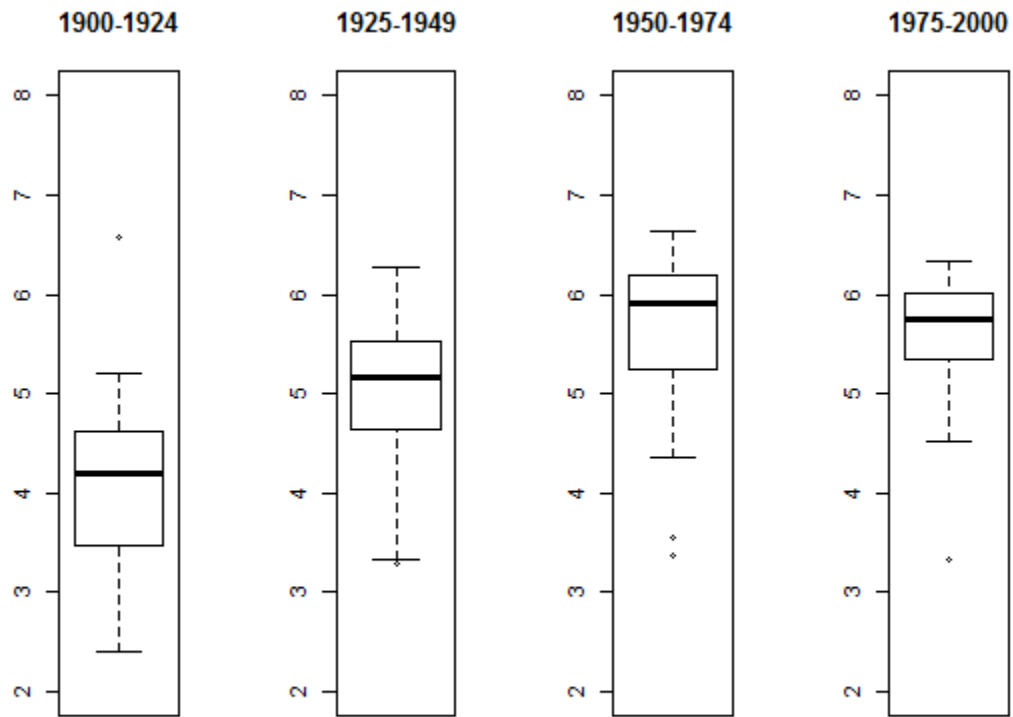


Figure 10: After transforming career home runs by logarithms, the boxplots of Figure 7 are re-created. Now, the spread of the boxplots does not vary significantly with the level of the transformed home runs. The right-skewness of the boxplots has also been eliminated.

Appendix C: No Need to Transform Career Hits

Career hits do not suffer from the same distributional problems as career home runs.

Figure 11 below repeats the analysis of spread vs. level shown in Figure 8 in Appendix B. The plot does not call for any transformation of career hits.

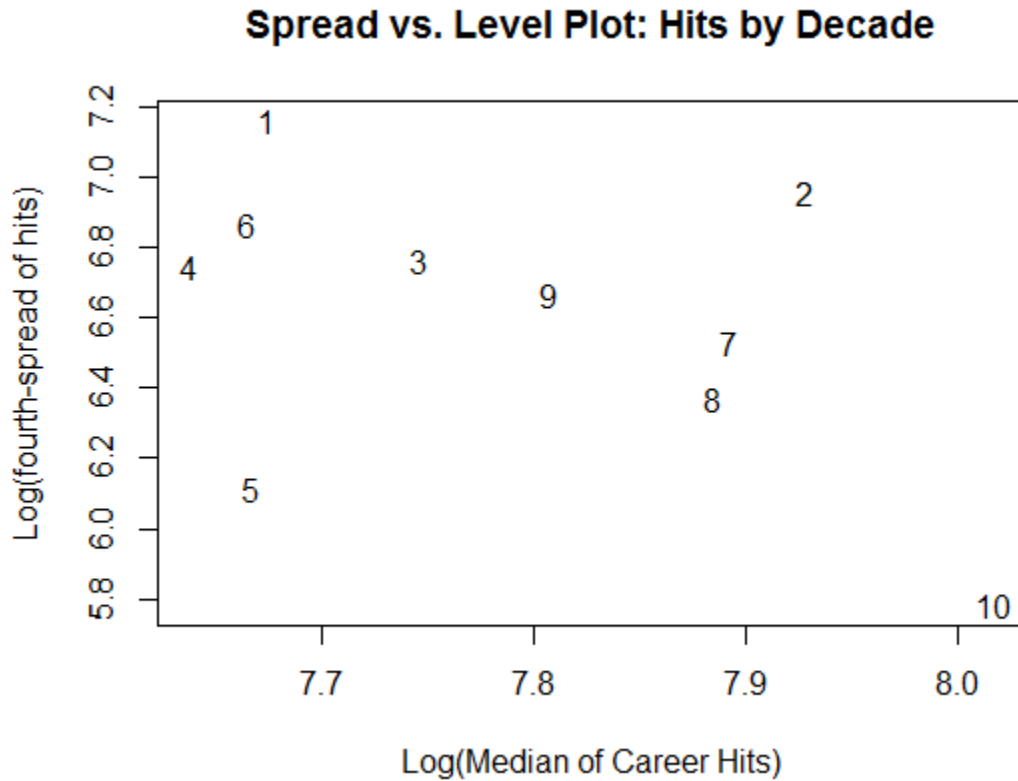


Figure 11: Spread Vs. Level plot of career hits. Players are batched according to the mid-point of their careers. Those players whose midpoint was between 1900 and 1910 are in batch 1, as shown on the graph. Batches continue until batch 10 shows the players whose mid-points were between 1990 and 2000. The y-axis is the log of the fourth spread of career hits and the x-axis is the log of the median of career hits. The spread of career hits does not seem to vary much with the level of hits by decade. Additionally, the level and spread do not vary much over time; note that the log of median(hits) for each decade is between 7.6 and 8.1.

Appendix D: Median Polish of Career Hits and Home Runs due to Era of Ballplayer and Years Until Induction into HOF

To examine the relative effects of the year of the mid-point of a player's career and the time between the player's retirement and his induction, it is possible to use a median polish. Median polish decomposes the player's career statistic into a sum of an overall value, an effect from the wait between retirement and induction, an effect from the era the player played in, and a residual error. The era the player played in is treated as the row effect in the median polish. For the rows, players are batched according to when the mid-point of their career was: 1900-1924, 1925-1949, 1950-1974, or 1975-2000. The time to induction is the column effect. For the columns, the players are batched according to how long they waited from retirement to induction: 6 years or less, 7-10 years, 11-20 years, or more than 20 years. Every combination of row and column has at least three HOF players in it except row 1975-2000, column wait >20 years, which has 0 players.

When using career hits in the median polish with the rows and columns as described above, the overall value is 2380. The era effects were as follows: 1900-1924 had an effect of +317, 1925-1949 had an effect of +10, 1950-1974 had an effect of -136, and 1975-2000 had an effect of -10. The years until induction effects were as follows: a wait of 6 years or less had an effect of +692, a wait of 7-10 years had an effect of +118, a wait of 11-20 years had an effect of -116, and a wait of more than 20 years had an effect of -359.

With effects ranging from +692 to -359, it is clear that the time to induction is a significant predictor of career hits. With effects from +318 to -136, occurring in a non-monotonic fashion, the effects of era were less significant in predicting total career hits.

When running a median polish with $\log(\text{career home runs})$ as the statistic, the overall effect was 5.3. The era effects were -0.9, -0.1, 0.6, and 0.1. The years until induction effects were 0.3, 0.1, -0.0, -0.1. Here, the era effects were more significant, as home runs have become more prevalent over time.

Figure 12 shows the Tukey Additivity Plots for each median polish. Neither plot suggests that non-additivity effects are significant here.

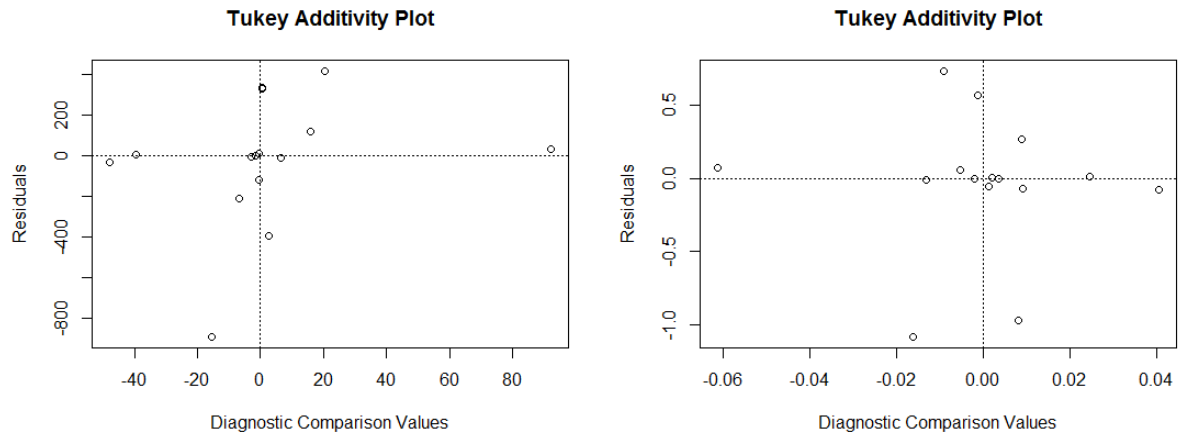


Figure 12: Tukey Additivity Plots for median polish described in Appendix D for career hits (left) and $\log(\text{career home runs})$ (right). Neither plot has a strong linear component that would suggest significant non-additivity effects.

Appendix E: Potential Future Second-Tier HOF Candidates

The players listed here have received at least 10% of the votes from the Baseball Writers' Association of America in a single year but have not yet been elected to the HOF. They represent potential second-tier candidates for future election by the Veteran's Committee. All of these players were voted upon before 2010 and have yet to gain election. Players who first appeared on the ballot after 2010 were not included because many of them are first-tier HOFers who will be elected on their second or third time on the ballot.

Catchers:

Johnny Kling
Hank Gowdy
Elston Howard
Walker Cooper
Thurman Munson
Joe Torre

First Basemen:

Phil Cavarretta
Gil Hodges
Mickey Vernon
Ted Kluszewski
Dick Allen
Steve Garvey
Keith Hernandez
Don Mattingly
Mark McGwire

Shortstops:

Marty Marion
Al Dark
Harvey Kuenn
Maury Wills
Dave Concepcion
Alan Trammell

Third Basemen:

Jimmy Dykes
Ken Boyer

Left Fielders:

Duffy Lewis
Lefty O-Doul
Minnie Minoso
Tim Lincecum

Center Fielders:

Terry Moore
Dom DiMaggio
Vada Pinson
Curt Flood
Dale Murphy

Right Fielders:

Smokey Joe Wood
Tommy Henrich
Roger Maris
Tony Oliva
Bobby Bonds
Dwight Evans
Dave Parker