

Who's the Best Home Run Hitter of All time?

Trey Cox, 2014

Standards addressed:

CCSS.MATH.CONTENT.6.SP.B.4

Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

CCSS.MATH.CONTENT.6.SP.B.5

Summarize numerical data sets in relation to their context, such as by:

CCSS.MATH.CONTENT.6.SP.B.5.A

Reporting the number of observations.

CCSS.MATH.CONTENT.6.SP.B.5.B

Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

CCSS.MATH.CONTENT.6.SP.B.5.C

Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

CCSS.MATH.CONTENT.6.SP.B.5.D

Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Activity:

Baseball is known as a sport that is infatuated with statistics. One of the most interesting and sought after titles in major league baseball is "home run king." There have been many claims to the throne. Roger Maris played Major League Baseball from 1957 – 1968, and in 1961 he hit a record 61 home runs for the New York Yankees in a single season. Maris broke the record that Babe Ruth, another New York Yankee, had previously set. In 1974, Hank Aaron of the Milwaukee and Atlanta Braves became the all-time home run hitter, surpassing Babe Ruth. In the 1990s, two players, Mark McGwire of the St. Louis Cardinals and Barry Bonds of the San Francisco Giants, had incredible seasons and both broke Maris' former single-season record. In 2007, amid controversy, Barry Bonds set the record for the most home runs hit in a career (762), passing Hank Aaron. All of these players have retired from the sport, so the controversy rages – which one of them is the greatest home run hitter of all time? It is now your job to settle the matter!

Cox, Frank E. 4/8/14 9:59 AM

Comment [1]: It can be fun to have the students get into the context of the problem by having them dress as baseball players on the day of this lesson. There is no direct educational benefit, but some students enjoy doing something out of the ordinary.

In the table below, the number of home runs hit by each athlete during each season of their careers is listed.

Roger Maris		Mark McGwire		Barry Bonds		Babe Ruth		Henry Aaron	
Season	Home runs	Season	Home runs	Season	Home runs	Season	Home runs	Season	Home runs
1957	14	1986	3	1986	16	1914	0	1954	13
1958	28	1987	49	1987	25	1915	4	1955	27
1959	16	1988	32	1988	24	1916	3	1956	26
1960	39	1989	33	1989	19	1917	2	1957	44
1961	61	1990	39	1990	33	1918	11	1958	30
1962	33	1991	22	1991	25	1919	29	1959	39
1963	23	1992	42	1992	34	1920	54	1960	40
1964	26	1993	9	1993	46	1921	59	1961	34
1965	8	1994	9	1994	37	1922	35	1962	45
1966	13	1995	39	1995	33	1923	41	1963	44
1967	9	1996	52	1996	41	1924	46	1964	24
1968	5	1997	58	1997	40	1925	25	1965	32
		1998	70	1998	37	1926	47	1966	44
		1999	65	1999	34	1927	60	1967	39
		2000	32	2000	49	1928	54	1968	29
		2001	29	2001	73	1929	46	1969	44
				2002	46	1930	49	1970	38
				2003	45	1931	46	1971	47
				2004	45	1932	41	1972	34
				2005	5	1933	34	1973	40
				2006	26	1934	22	1974	20
				2007	28	1935	6	1975	12
								1976	10

Cox, Frank E. 4/6/14 2:45 PM

Comment [2]: Ask the students if it makes sense to, during analysis, keep the seasons in which the players only hit a few home runs and why or why not.

Cox, Frank E. 4/6/14 2:47 PM

Comment [3]: Ask the students whether we can still compare the players fairly even though they played different numbers of seasons.

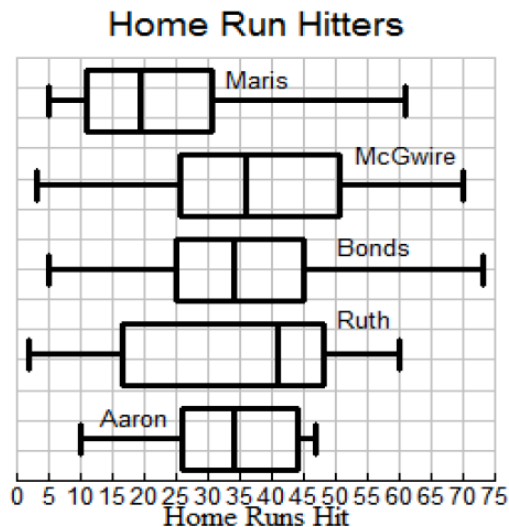
a) For each player, compute the five-number summary of the number of season home runs.

Maris	Min: 5	Q1: 11	Med: 19.5	Q3: 30.5	Max: 61
McGwire	Min: 3	Q1: 25.5	Med: 36	Q3: 50.5	Max: 70
Bonds	Min: 5	Q1: 25	Med: 34	Q3: 45	Max: 73
Ruth	Min: 2	Q1: 16.5	Med: 41	Q3: 48	Max: 60
Aaron	Min: 10	Q1: 26	Med: 34	Q3: 44	Max: 47

Cox, Frank E. 4/4/14 12:14 PM

Comment [4]: This question could be an introduction to the five-number summary or practice activity that students do after they have learned about it. If it is introductory, this question will take longer, as you will need to model how students calculate the summary and/or how to use technology, such as graphing calculators, to do so. You may want to share this website with your students for a "how to" create a five number summary:
<https://www.youtube.com/watch?v=TRnVn3ZxLDs>

- b) Make 5 side-by-side boxplots of the home run distributions. Make sure to label each graph with the player's name.



Cox, Frank E. 4/4/14 12:14 PM

Comment [5]: You may choose to have students do this on their calculators, but the advantage to having them draw the plots by hand is that they're more likely to understand why the boxes and whiskers are drawn the way they are from the five-number summary.

- c) Using the analysis of data above, make a case statistically for who you think is the best home run hitter of all time. Be sure to back up your claim with relevant statistics.

There is a case to be made for more than one player (Bonds, Ruth, Aaron, or McGwire) but it appears that from this data that Mark McGwire is the best home run hitter if we define that by more consistently hitting more home runs per year versus more consistently hitting the same number (in which case it would be Hank Aaron). McGwire has the highest or second highest number of home runs for quartile one, the median, quartile three, and the maximum. His box plot is also skewed toward the higher number of home runs.

Bonds could be considered the best of all time if one focuses on the fact that his 4th quartile is similar to McGwire's but has a higher maximum. Ruth's 2nd quartile encompasses all of the others' medians, meaning he was more consistent with higher home run totals, which could be an argument for why he is the best. Finally, Aaron played the most seasons and his home run totals had a smaller overall range and smaller quartile ranges, meaning he was consistent throughout his career.

Cox, Frank E. 4/4/14 12:16 PM

Comment [6]: A convincing argument would be one that is backed up with data and not simply a statement of personal feelings and thoughts. The conclusion should be understandable by anyone looking at the same data.

Cox, Frank E. 4/4/14 12:17 PM

Comment [7]: There isn't only one possible answer. The student would be correct if they are able to compare and contrast the player they chose against the others based on the center, shape, spread, and distribution.

- d) Consider Barry Bonds' home run counts. Estimate the existence any outliers and provide reasonable explanations (at least 3) for their existence. Be sure to include the numerical value and a potential contextual explanation for any outlier(s).

For Barry Bonds it appears that when 5 home runs and 73 home runs were hit they would be outliers. The low number of 5 most likely is because of a season filled with injury(ies) and the very high number of 73 occurred at the peak of the baseball era of steroid usage.

Note: If you choose to use the standard of $1.5 \times IQR$ there are no outliers for Bonds.

Cox, Frank E. 4/4/14 12:18 PM

Comment [8]: The reasons may not be apparent to students that do not participate in athletics, so be prepared to give hints about injuries, trades to another team, family issues, or cheating (steroids, corked bats, etc.).

- e) If you had to describe one of the five players as the most consistent power hitter of all time, based on the above graphs and the summaries that you calculated, which player would you say would fit that description?

Hank Aaron is probably the most consistent home run hitter of the five because home run totals had smaller overall range and smaller quartile ranges meaning he was consistent throughout his career. If one does not consider Roger Maris' season in which he hit 61 home runs (likely an outlier) then he would also be a consistent home run hitter.

Cox, Frank E. 4/6/14 2:46 PM

Comment [9]: This could very well be a different answer than they gave for the greatest home run hitter. In fact, it would be important to state that to the students and make them aware that consistency and greatest are not the same thing necessarily. Consistency pertains to hitting similar numbers of home runs year after year and the greatest could mean the most home runs, or the largest numbers of home runs, etc.

- f) What type of career would a future player have to have in order for you to say that he has dethroned the player you chose in (c)? Be specific in terms of the number of seasons played and the numbers of home runs hit.

He would have to play as many seasons or more than these players, and also have more seasons with home runs in the 45 – 60 range. He would not necessarily have to hit a greater maximum number of home runs, but would have to have fewer seasons with low totals and more with higher totals.

Cox, Frank E. 3/29/14 10:10 AM

Comment [10]: This is a great question to see if students can take the synthesis of information that they have gathered from the data and project what it would take to surpass those statistical summaries.