

The (Batting) Order of Canada

Introduction

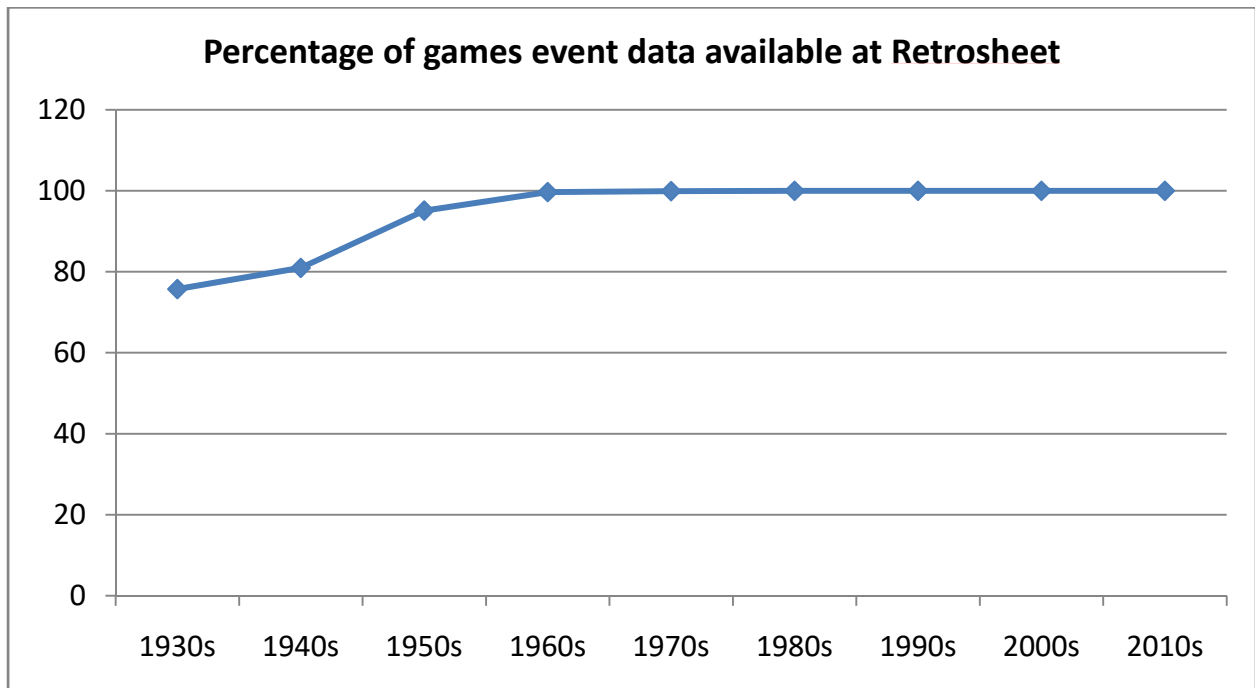
From 1930 through 2017, a total of 118 Canadian-born players have taken a spot in the batting order at the start of a game (that is, have been one of the first nine batters) in MLB on 20071 occasions.

This presentation examines how the distribution of these players throughout the nine spots in the order at the start of the game has changed over the years: from batting cleanup only twice in the 1950s to 1468 times in the 2000s ... and everything in between.

Methodology

Event files were downloaded from Retrosheet, unpacked, and stored in MySQL database tables. A query was constructed to find the first nine batters for each team in a particular season. This query was then ported to PERL to enable the query to be run for all seasons - generating a results table for each year.

The master.csv file from the Lahman Baseball Database was used as the source of birthplace information to select Canadian-born players. Results were queried and grouped by decade.



Canadian-born players by the numbers (Part I)

The number of Canadian-born players has fluctuated over the decades, and the distribution within the order has changed as well. In the 1930s, these players were found mostly in the middle of the order, but in the 1940s, most were found near the top of the order or in the traditional pitcher's spot of ninth.

In the 1950s, there was a substantial drop in the overall number of Canadian-born players. In the 1960s, numbers began to recover and both decades have a noticeable number of players batting ninth, and the 1960s have a substantial number of batters in the middle of the order once again.

The 1970s and 1980s saw a shift away from the middle of the order and a noticeable decline in players found batting ninth.

The 1990s and subsequent decades have seen a return to the middle of the order for Canadian-born players.

Canadian-born players by the numbers

decade	first	second	third	fourth	fifth	sixth	seventh	eighth	ninth	
1930s	125	55	61	164	133	220	189	27	22	996
1940s	80	307	179	514	397	96	64	8	397	2042
1950s	3	29	28	2	17	22	7	9	106	223
1960s	4	25	187	388	218	78	63	3	323	1289
1970s	253	25	75	16	8	46	134	135	349	1041
1980s	170	279	129	4	82	111	256	146	122	1299
1990s	42	62	456	747	290	186	167	116	279	2345
2000s	174	231	1121	1468	1025	744	415	324	507	6009
2010s	164	310	1183	485	700	624	593	433	335	4827
	1015	1323	3419	3788	2870	2127	1888	1201	2440	20071

Canadian-born players by the numbers (Part II)

Individual players were selected to make note of their participation.

- 1930s – George Selkirk (Canadian Baseball Hall of Fame Inductee - 1983) took over from Babe Ruth in right field and wore Ruth's #3 for a time.
- 1940s – Jeff Heath (born in Canada but raised in the U.S.) was a three-time All-Star and was named one of the all-time top 100 Cleveland Indians. Dick Fowler (CBHOF 1985) threw the first no-hitter by a Canadian-born player.
- 1950s – Glen Gorbous had a short career with Cincinnati and Philadelphia, but holds the world record for longest baseball throw at 445 feet , 10 inches.
- 1960s – Pete Ward (another player born in Canada but raised in the US – his dad, Jimmy played for the Montreal Canadiens) was runner-up for AL ROY in 1963. Pete was inducted into the Canadian Baseball Hall of Fame in 1991.
- 1970s – Terry Puhl (CBHOF 1995) begins play for the Astros. Among other accomplishments, as of 2017, Terry has the 18th best fielding percentage among outfielders all-time.
- 1980s – Doug Frobel is being groomed by the Pittsburgh Pirates to take over from Dave Parker.
- 1990s – Paul Quantrill (CBHOF 2010) and Rheel Cormier (CBHOF 2012) are appearing for the Blue Jays and Expos (among other teams).

From the 1990s on, there has been a shift back to the middle of the order with the appearance of Larry Walker (CBHOF 2009), Matt Stairs (CBHOF 2015), Justin Morneau, Jason Bay, Corey Koskie (CBHOF 2015), Russell Martin, Brett Lawrie, Michael Saunders, and, of course, Joey Votto.

Canadian-born players by the numbers

decade	first	second	third	fourth	fifth	sixth	seventh	eighth	ninth	
1930s	125	55	61	164	133	220	189	27	22	996
1940s	80	307	179	514	397	96	64	8	397	2042
1950s	3	29	28	2	17	22	7	9	106	223
1960s	4	25	187	388	218	78	63	3	323	1289
1970s	253	25	75	16	8	46	134	135	349	1041
1980s	170	279	129	4	82	111	256	146	122	1299
1990s	42	62	456	747	290	186	167	116	279	2345
2000s	174	231	1121	1468	1025	744	415	324	507	6009
2010s	164	310	1183	485	700	624	593	433	335	4827
	1015	1323	3419	3788	2870	2127	1888	1201	2440	20071

Annotations in the table:

- Glen Gorbous: points to 29 (1950s second)
- Jeff Heath: points to 514 (1940s fourth)
- George Selkirk: points to 220 (1930s sixth)
- Dick Fowler: points to 397 (1940s ninth)
- Terry Puhl: points to 253 (1970s first)
- Pete Ward: points to 388 (1960s fourth)
- Doug Frobel: points to 256 (1980s seventh)
- Quantrill / Cormier: points to 279 (1990s ninth)

Canadian-born players by percent of order

The relative distribution of the players helps to show the changes over the decades. The higher number of Canadian-born players in the ninth spot from the 1940s through the 1970s is readily apparent. The shifts from the middle to the ends of the order and back again are also quite noticeable.

Canadian-born players by percent of order

decade	first	second	third	fourth	fifth	sixth	seventh	eighth	ninth
1930s	12	6	6	17	14	22	19	3	2
1940s	4	15	9	25	20	5	3	0	19
1950s	1	13	13	1	8	10	3	4	47
1960s	0	2	15	30	17	6	5	0	24
1970s	24	2	7	2	1	4	13	13	33
1980s	13	22	10	0	6	9	20	11	9
1990s	2	3	19	32	12	8	7	5	12
2000s	3	4	19	24	17	12	7	5	8
2010s	3	6	25	10	15	13	12	9	7

Percentage of Canadian-born players all-time

The percentage of Canadian-born players compared to all participants shows that Canadian-born players have really started to make their mark in the middle of the order in recent decades.

Percentage of Canadians All-Time

decade	first	second	third	fourth	fifth	sixth	seventh	eighth	ninth
1930s	0.66	0.30	0.33	0.88	0.72	1.18	0.99	0.15	0.11
1940s	0.41	1.60	0.92	2.67	2.06	0.50	0.33	0.04	2.02
1950s	0.01	0.12	0.12	0.01	0.07	0.09	0.03	0.04	0.45
1960s	0.01	0.08	0.59	1.22	0.69	0.25	0.20	0.01	0.99
1970s	0.64	0.06	0.19	0.04	0.02	0.12	0.34	0.34	0.88
1980s	0.42	0.69	0.32	0.01	0.19	0.27	0.63	0.36	0.30
1990s	0.10	0.14	1.05	1.73	0.67	0.43	0.39	0.26	0.64
2000s	0.36	0.48	2.30	3.02	2.11	1.54	0.86	0.67	1.05
2010s	0.42	0.80	3.04	1.25	1.80	1.60	1.52	1.12	0.86

Issues

A couple of issues popped up while doing the analysis.

Initial query results had different total numbers of batters in each of the nine spots. This indicated an error in the query as the number of batters in each of the nine spots must be the same. Some detective work revealed that the cause was the occurrence of a fielding error prolonging the first at-bat. The query was corrected and the results then added up correctly. Interestingly, this has occurred 159 times in the years from 1930 – 2017.

Another oddity was the discovery that Ferguson Jenkins (CBHOF 1987) had batted in a game in 1974 while playing in the American League for the Texas Rangers. Some more research (using Baseball Reference in particular) revealed that the Rangers did not use a DH in the last game of that season – instead choosing to let Jenkins hit.

This led me to look for other occurrences where teams have not used a DH. There have been six occasions where this has occurred. These games do not include occasions where another pitcher has been used as the DH in the game (for example, Rick Rhoden) and also does not include games where teams began with a DH but lost it during the game (due to injury or substitution).

Interestingly, one of these six games involved Ferguson Jenkins as the pitcher – pitching to Ken Brett who was hitting (without a DH being used).

Summary

The occurrence of Canadian-born players within the batting order has fluctuated over the decades – from a greater occurrence in the ninth spot early on - to the middle of the order in recent times. This seems to mirror the occurrence of many Canadians as pitchers - through the installation of the DH rule for the 1973 season – to current times where most Canadian-born players are occupying spots in the middle of the order.

I am interested in modelling batting order using simulation driven by the probability of transition-state changes. The batting order data is being used to determine orders for different teams over the years to compare actual production to the production suggested by simulation. I want to try to characterize what kind of batter has batted in each spot historically, and, if possible, to model what kind of batter should bat in each spot in the order.

Acknowledgements

The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at www.retrosheet.org. Anecdotal information about players was taken from Wikipedia.

Three books in particular have provided inspiration to me. They are: Baseball Hacks by Joseph Adler; The Book by Tom Tango, Mitchel Lichtman, and Andrew Dolphin; and Analyzing Baseball Data with R by Max Marchi and Jim Albert.

I also want to acknowledge all those people who post ideas, data, and code. None of this would have been possible without their generosity.

Finally, I would like to thank the Centre for Canadian Baseball Research (at <http://baseballresearch.ca/>) for the opportunity to make a presentation at the 3rd Canadian Baseball History Conference.