

Negro Leagues

Paul A. Hodgetts

03 May 2021

Abstract

As Bradshaw left, the lawyer looked at his watch. “And now, Poole, let us get to ours,” he said; and taking the poker under his arm, led the way into the yard. The scud had banked over the moon, and it was now quite dark. The wind, which only broke in puffs and draughts into that deep well of building, tossed the light of the candle to and fro about their steps, until they came into the shelter of the theatre, where they sat down silently to wait. London hummed solemnly all around; but nearer at hand, the stillness was only broken by the sounds of a footfall moving to and fro along the cabinet floor. Code available on GitHub: <https://github.com/hodgettsp/nlmlb>

Introduction

On April 15th, 1947, Jackie Robinson made his debut for the Brooklyn Dodgers and broke the Major League Baseball (MLB) ‘Colour Line’ that had been established as an unwritten agreement amongst league owners preventing non-white players from playing in the MLB. Although not the first Black man to play in the major leagues, historians consider William Edward White the first Black player and Moses Fleetwood Walker as the first openly Black player, Robinson was the first to break the established segregation and usher in an era of integration for the MLB. Indeed, Moses Fleetwood Walker was the last Black player in the Majors and it was the refusal of Cap Anson to take the field against Fleetwood Walker that led to the segregation of baseball and establishment of the “Colour Line”.

Before joining the Brooklyn Dodgers organization in 1946, Robinson spent the 1945 baseball season playing for the Kansas City Monarchs of the Negro American League. While various iterations of the Negro Leagues had existed, it is generally agreed upon by baseball historians that the period of time that is referred to by use of the Negro Leagues is between the 1920 and the establishment of the Negro National League and the early 1960’s with the folding of the Negro American League. In a time during which Black players were not allowed to play in the MLB, the Negro Leagues provided a place for Black players to play and be paid for their efforts. Over the seasons, the Negro Leagues saw such stars as Cool Papa Bell, Satchel Paige, Josh Gibson, Ernie Banks, Willie Mays, and Henry Aaron to name a few. Like Robinson, some of these players saw time in the MLB. However, the integration of the MLB did not eliminate the racism these players experienced, which continued from the fans and other players. Moreover, many players from the Negro Leagues were never given a fair opportunity to play in the MLB, while those who retired from baseball before the 1947 season were never even given a chance.

While the integration of baseball has been considered necessary, the effects that the implementation of integration had on the Negro Leagues and Black-baseball in the United States in general have been questioned. Integration itself was also not an immediate process. For instance, while Jackie Robinson debuted in 1947 for the Brooklyn Dodgers of the National League and Larry Doby debuted in the same year shortly after for the Cleveland baseball team of the American League, it was not until 1959, twelve years after Jackie Robinson and Larry Doby first played and three years after Robinson retired and the same year that Doby retired, that the entire Major Leagues was integrated when Pumpsie Green debuted for the Boston Red Sox. Additionally, the difficulty of making the Major Leagues has been noted by previous Negro League players. Due to both the competition and that teams would often only carry two to three Black players on the roster at a given time.

This paper explores the following: the accessibility of Negro League data as compared to Major League data of the same period as discussed within the larger context of racially based historical data, the process of integration and the number of Black players (those who played in the Negro Leagues and those who did not) carried by teams in the Integration Era, and the effects of integration on the Negro Leagues. The Integration Era of baseball is often classified as the beginning of the 1947 season to the end of the 1960 season. After which, the Major Leagues expanded to include two new teams in the American League at the beginning of the 1961 season. While an official date for the end of the Negro Leagues is not officially agreed upon, Negro League All-Star games were played in 1962. For this reason, this paper will examine integration as from the beginning of the 1947 season to the end of the 1962 season. This is done to include as much of a period of the officially recognized Negro Leagues as possible; however, it does mean including two years in which the Major Leagues expanded (1961 and 1962). As such, a particular focus will be between the 1947 and 1960 seasons as the introduction of the four new teams would change the demographics of the player population. Additionally, the league was fully integrated in 1959, so use of 1960 provides a useful cut-off point to mark the end of initial integration.

The Negro Leagues

In 1920, the Negro National League was founded in large part due to the efforts of Andrew “Rube” Foster. While previous Black teams and leagues had existed in some form, the formulation of this league marked what is commonly considered the beginning of the modern Negro “Major” Leagues. For 11 years the Negro National League was considered the highest level of play amongst Negro league teams, but financial troubles brought on by the Great Depression led to the folding of the league in 1931. With the folding of the Negro National League in 1931, the remaining Negro Southern League, hosting such teams as the Nashville Elite Giants and Birmingham Black Barons, was elevated as the de facto major league for the 1932 season. However, this designation only lasted for the 1932 season as the Negro National League was re-established in 1933 and was again considered the major league for the Negro Leagues. Four years later the Negro National League would be joined at the major league level by the establishment of the Negro American League in 1937. These leagues continued operation until between the 1948 and 1949 seasons, a year after Jackie Robinson debuted for the Brooklyn Dodgers and Larry Doby debuted for the Cleveland baseball team, when the Negro National League folded. The Negro American League would continue operation until between the 1962 and 1963 seasons when the final All-Star game was held. This point is generally marked as the official end of the Negro

League. Other teams, like the Indianapolis Clowns existed and continued operating, but often only for exhibition games and not in a competitive sense.

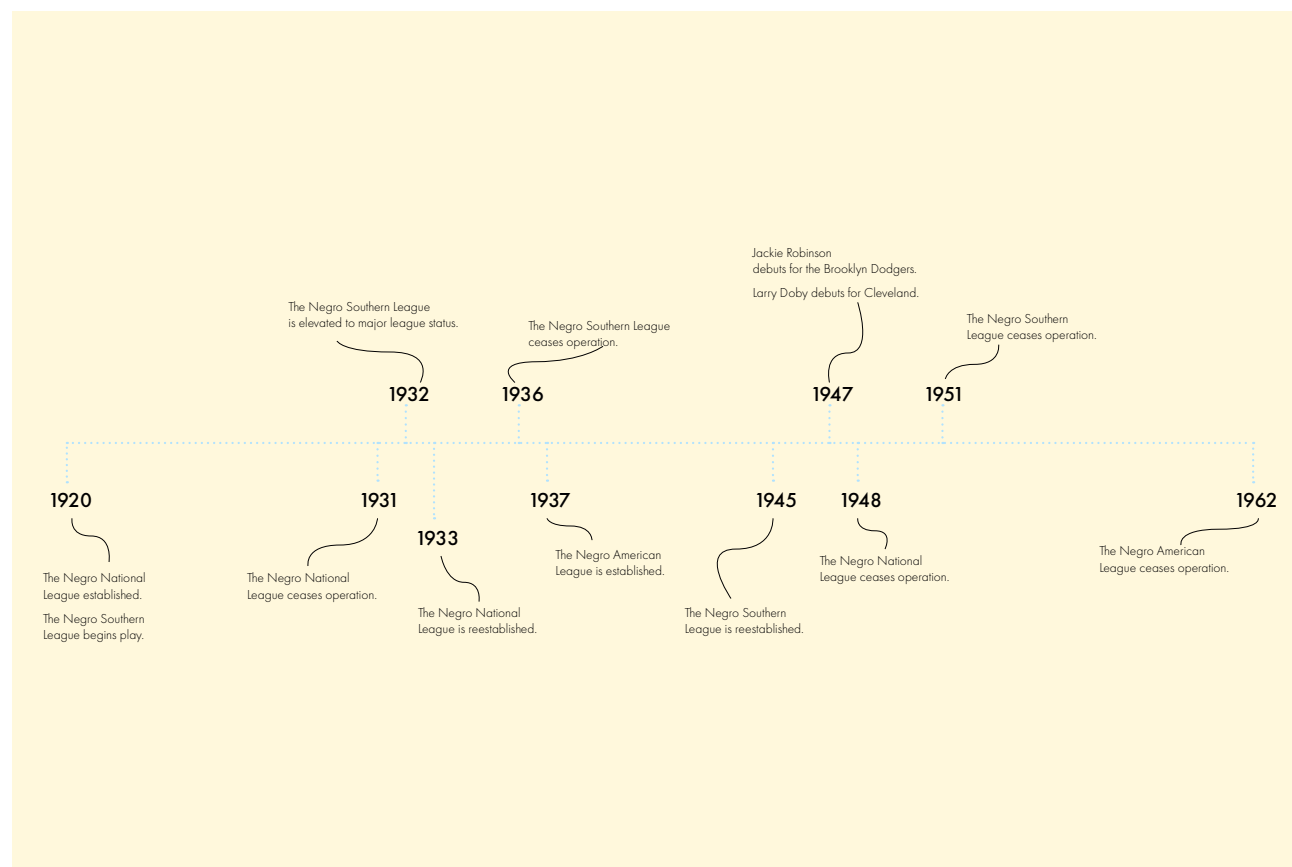


Figure 1: A general timeline of the Negro Leagues.

Data

Access

Access to Negro League data is limited when compared to the accessibility of Major League Baseball data for the same period of time (1920-1962). For example, the *Lahman* [4] R package provides access to MLB player, batting, pitching, and fielding data dating back to 1871. Whereas, no equivalent R package exists for Negro League data. Additionally, batting and fielding data is currently only available for 1920 to 1948 through the ‘Seamheads’ Negro Leagues database website [5]. Additionally, this database only covers games between teams that the managers of the database have deemed as Black major league teams. This means games against white semipro and amateur teams, lower-level white minor league teams, college teams, and teams defined as Black minor league teams. Prior to the 1920 season, these games comprised a large part of Black teams’ schedules, which means statistics from these games are largely unavailable. This is in comparison to the accessibility of MLB data from websites such as Baseball Reference, Fangraphs, and the MLB website. Additionally, in comparison to access to

Negro League data, this MLB data is readily accessible through databases like the Lahman database and Retrosheets database. Additionally, packages such as `Lahman` [4], `retrosheet` [3], and `baseballr` [10] make accessing data from these sites easier, while no such equivalent package exists for Seamheads or Negro League data.

Collection

Due to the lack of immediately downloadable data, data for this project was compiled from scraping PDF files made available by the Center for Negro League Baseball Research (CNLBR) [2]. Additional Negro League player lists were scraped from the Wikipedia entries that alphabetically list Negro League players [12]. Code from the Wikipedia pages was scraped using the `rvest` package [11]. In a comparison to Negro League player lists provided by the CNLBR, it was noted that the Wikipedia list of Negro League players was incomplete. To complete this list, these players were manually entered into the Wikipedia scraped list. However, these players were only found because they also played in the MLB, meaning that the list of Negro League players used in this paper may be incomplete. There are indeed discrepancies between both lists, with the Wikipedia list containing players who do not appear on the Center for Negro League Baseball Research list and vice versa. Meaning, the current list of Negro League players is incomplete. Highlighting another issue of the quality and accessibility of Negro League data. The data specifically collected from Wikipedia is a list of Negro League players.

Data collected from the Center For Negro League Baseball Research includes a list of Negro League players who also played in Major League Baseball[7]. This was downloaded as a PDF, from which the data was scraped using the `pdftools` package [9]. This list was chosen as it provides the most complete list of players who played in both the Negro Leagues and MLB. For instance, the list of players provided by Wikipedia only goes up to the 1950 season and then only includes a select few players who debuted in either league after that point. Other lists of players, such as that provided by the Negro League Players Association, only provide the names of players who played in both leagues and do not include the dates at which the players played their final game in the Negro Leagues and debuted in the MLB. These dates are necessary to better understand the timeline of the Negro Leagues and MLB, which is why the Wikipedia list and CNLBR lists were chosen.

To also understand the population of the MLB, the `Lahman` [4] R package was used to access player information for all MLB players. This package provides direct access to Sean Lahman's baseball archive, a free relational database of individual and team statistics that covers games back to 1871, for R users. For non-R users, the database is also available for download in a comma-delimited version (csv), and through MS Access. Seasons previous to the 2020 season are also available through SQL. In R, this database is divided into the sections `People` (player names, dates of birth, death, and other biographical info), `Batting` (batting statistics), `Pitching` (pitching statistics), and `Fielding` (fielding statistics). Information, when available, dates back to 1871. For the purpose of this paper, only the `People` section of data was used.

Analysis

Comparison of Data Sources

Player Debut Season vs. Player Final Season

In addition to exploring the accessibility of Negro League data in comparison to MLB data, a cursory analysis of the effects of integration on the Negro Leagues was also performed. Previous analyses have performed a statistical comparison between Negro League players and MLB players. For instance, a piece from 538 [8], provides a comparison of Negro League players, those who did and did not play in MLB, to current and historical MLB players. Other analyses, such as this report by Armour and Levitt [6], provide an analysis of racial demographics of baseball from the start of integration in 1947 to 2016. This report by Pavitt [1] examines player participation and fielding position in MLB by player ethnicity from 1946-2018. However, while there has been much written about the end of the Negro Leagues in relation to integration, there is little in regards to a statistical analysis of the process. Particularly as a focus on the Negro Leagues and the players. Two things in particular are examined in this analysis, the count of players in the Negro Leagues pre- and post-integration, and the count of Negro League players and Black players in MLB in the Integration era of MLB (1947 to 1960).

Figure 1 shows the distribution of players who debuted and played their final game for a given season in the Negro Leagues from 1873 to 1975. However, while numbers pre-1920 are available these records are spotty and may not truly represent the actual number of players in the Negro Leagues prior to the 1920 season. The founding of the Negro National League in 1920 can be seen in Figure 1 with the increased number of players making their debut that year and the following season. This could be for two reasons, firstly that there was actually that many players making their debut or that record keeping began for that season and many players who had previously been playing were marked as having debuted for that year. The first collapse of the Negro National League in 1931 can also be seen with close to 70 players their final game. The increase in players making their debut in 1937 can also be attributed to the founding of the Negro American League. Of particular note is the dramatic increase in players who played their final Negro League game in 1948 with 138 players playing their final game. This marks the year after integration formally began with MLB, but also marks the year in which the Negro National League ceased operation. The spike of final games played in 1951 also marks when the Negro Southern League ceased operation.

The Negro American League folded around the 1963 season, and as can be seen in Figure 1 there were very few players to have debuted or played their final game by that point. Indicating the dire state of the league.

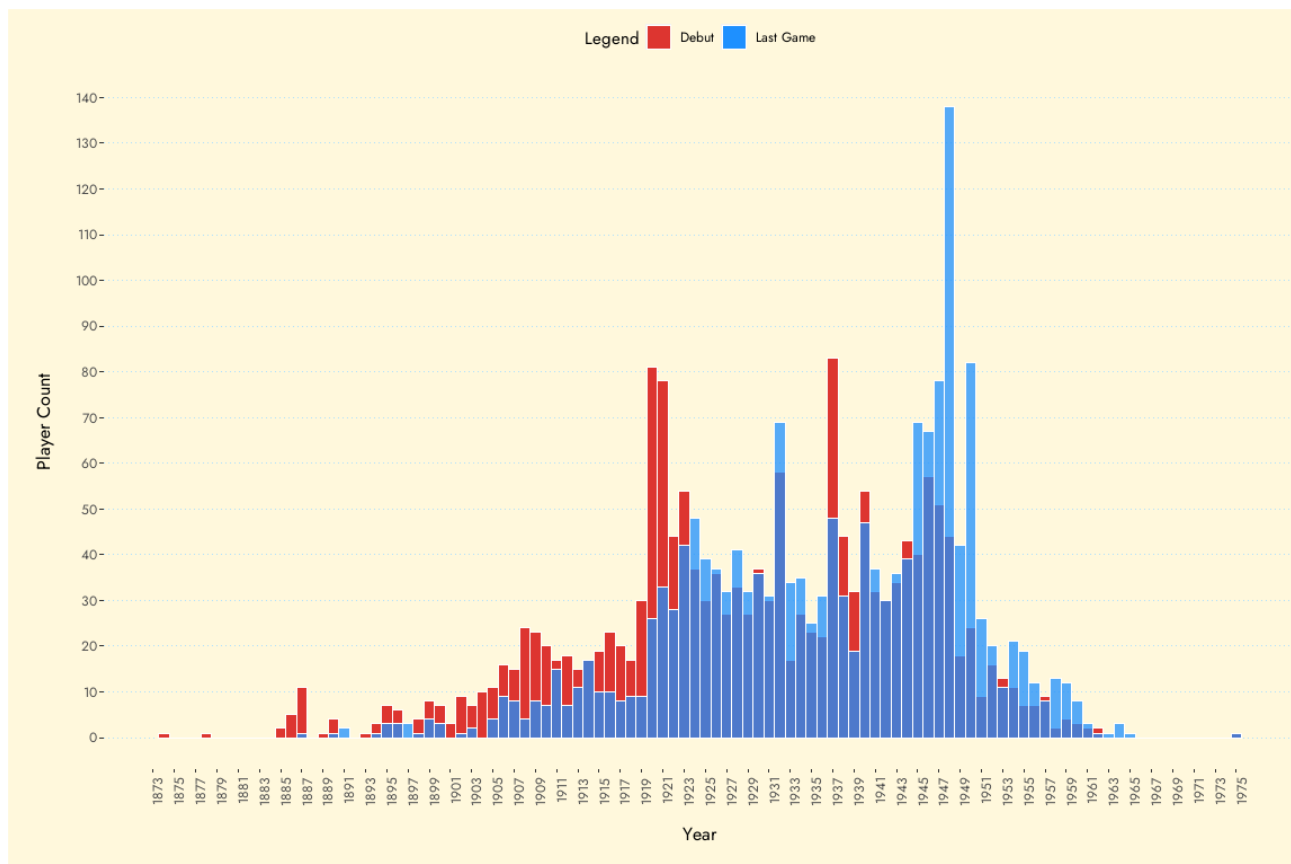


Figure 2: A distribution of the number of players debuting and playing their final game in a given season in the Negro Leagues, 1873 to 1975.

In comparison, Figure 2 shows the distribution of players making their debut or playing their final game according to Seamheads. As can be seen, the distribution is fairly similar as to the distribution collected from Wikipedia; however, the counts from Seamheads are higher, which may be in part to some players not being included on the Wikipedia list due to insufficient data that can be used to confirm the player's identity.

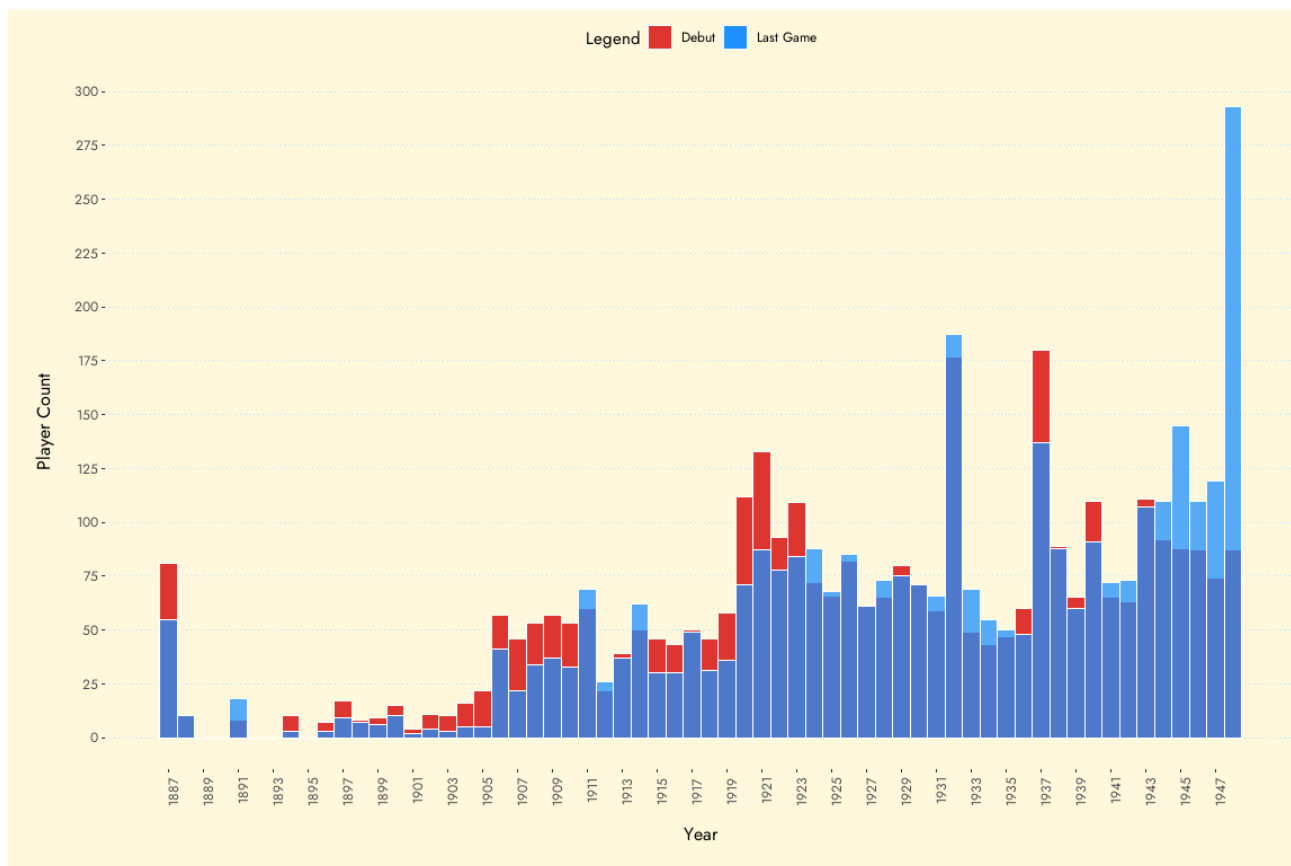


Figure 3: A distribution of the number of players debuting and playing their final game in a given season in the Negro Leagues, 1887 to 1948

Figure 3 again shows the distribution of the Wikipedia data, however, it has been filtered to only include the same range of years as is in the Seamheads data. This provides a more direct comparison between the two datasets. As can be seen between Figure 2 and Figure 3, the counts are indeed higher in the Seamheads data and the slightly different pattern is more evident between the two with Seamheads (Figure 2) showing a greater number of debuts in 1921 than 1920 whereas the Wikipedia data (Figure 3) shows a greater number of players making their debut in 1920 than 1921.

Other differences between the two include Seamheads show a greater number of debut seasons than final seasons in 1929, whereas Wikipedia shows the opposite. Wikipedia shows a greater number of final season than debut season for 1936, whereas Seamheads shows the opposite. Other similar patterns can be seen throughout the various seasons between the datasets. This is of particular interests as the Wikipedia list is reliant upon the Seamheads database for the information contained within its list of players.

An issue that also was found in the process of this analysis was that in some cases Wikipedia had the birth-date of the player for the player's debut date in the Negro Leagues. This discrepancy may partially explain the inclusion of dates earlier than 1887 in the Wikipedia list of Negro League players.

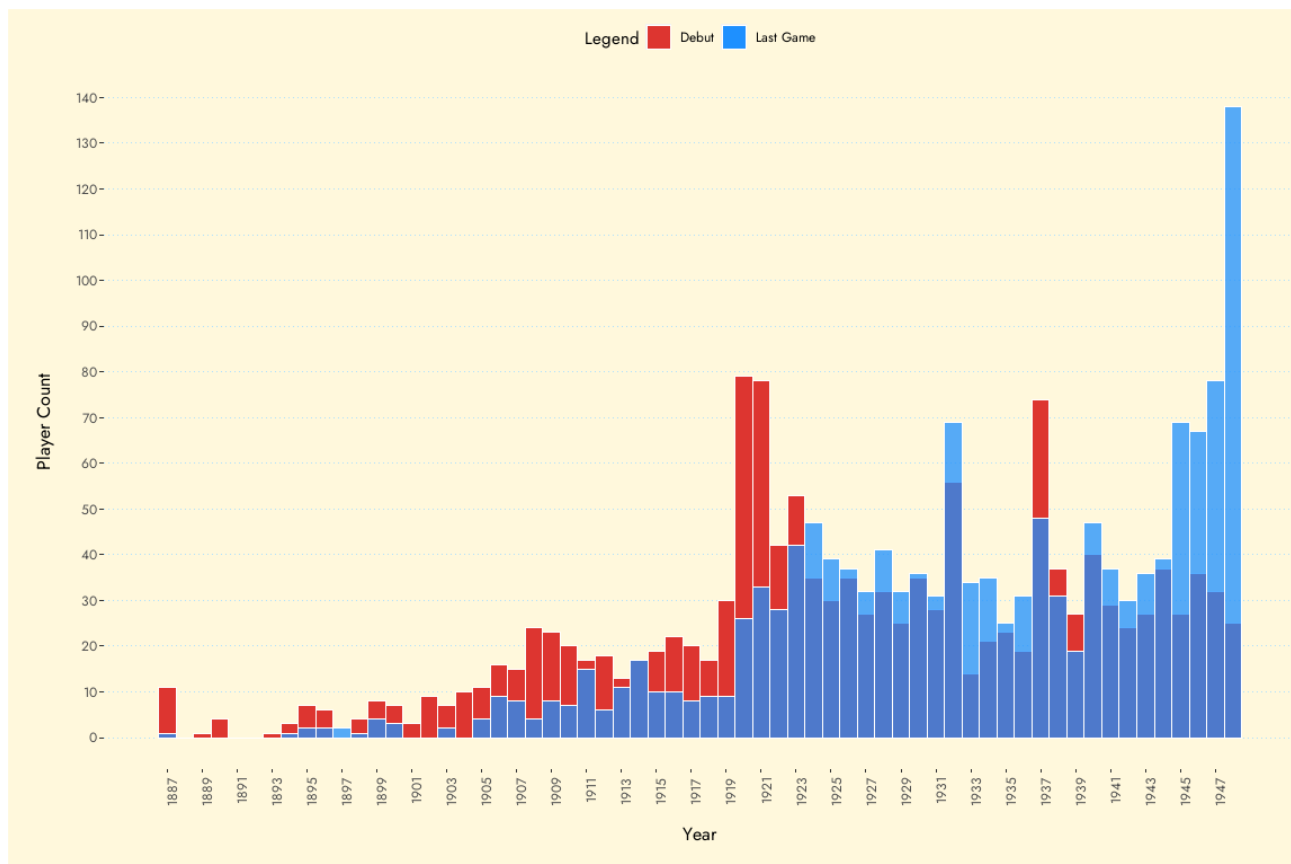


Figure 4: A distribution of the number of players debuting and playing their final game in a given season in the Negro Leagues, 1887 to 1948

Player Counts

While Figures 1 through 3 show the distribution of players who debuted or played their final game for a given season in the Negro Leagues, it is also necessary to understand the total number of players in the Negro League for a given season. Figure 4 shows the distribution of total number of players for the Negro Leagues from 1873 to 1975. This distribution shows a peak of total number of players in the Negro Leagues in the 1923 season with 383 players, shortly after the Negro National League was established and the Negro Southern League began play. Additionally, the decrease of players in the Negro Leagues due to the folding of leagues is evident towards the middle of the distribution with a resurgence when leagues were re-established or new leagues established. Interestingly, this decrease is greater than the one found during the United States' involvement in World War II (1941 to 1946). Moreover, the steep decline in players in the Negro Leagues that is evident in Figure 1 beginning between 1948 and 1949 is also evident here.

Given that the data available only provided the number of players who debuted or played their final game for a season, the total number of players was not available. In the following formula let d stand for the debut games for a season, p stand for the player count for the previous season, and f stand for the count of final games played in the previous season. As such, the count for the number of players for a given season was calculated as:

$$(d + p) - f$$

The collected data from Wikipedia contained missing values for both debut and final game counts for various seasons. These missing values were replaced with a value of 0. This was done to keep those seasons included in the count, as the complete-case approach would remove these seasons. These missing values were created when there was no data for either debuts or final games for a season. However, this approach of imputing missing values with 0 while allowing for keeping these seasons does mean the distribution for players may be off or less than the actual count. There were no missing values for the Seamheads data.

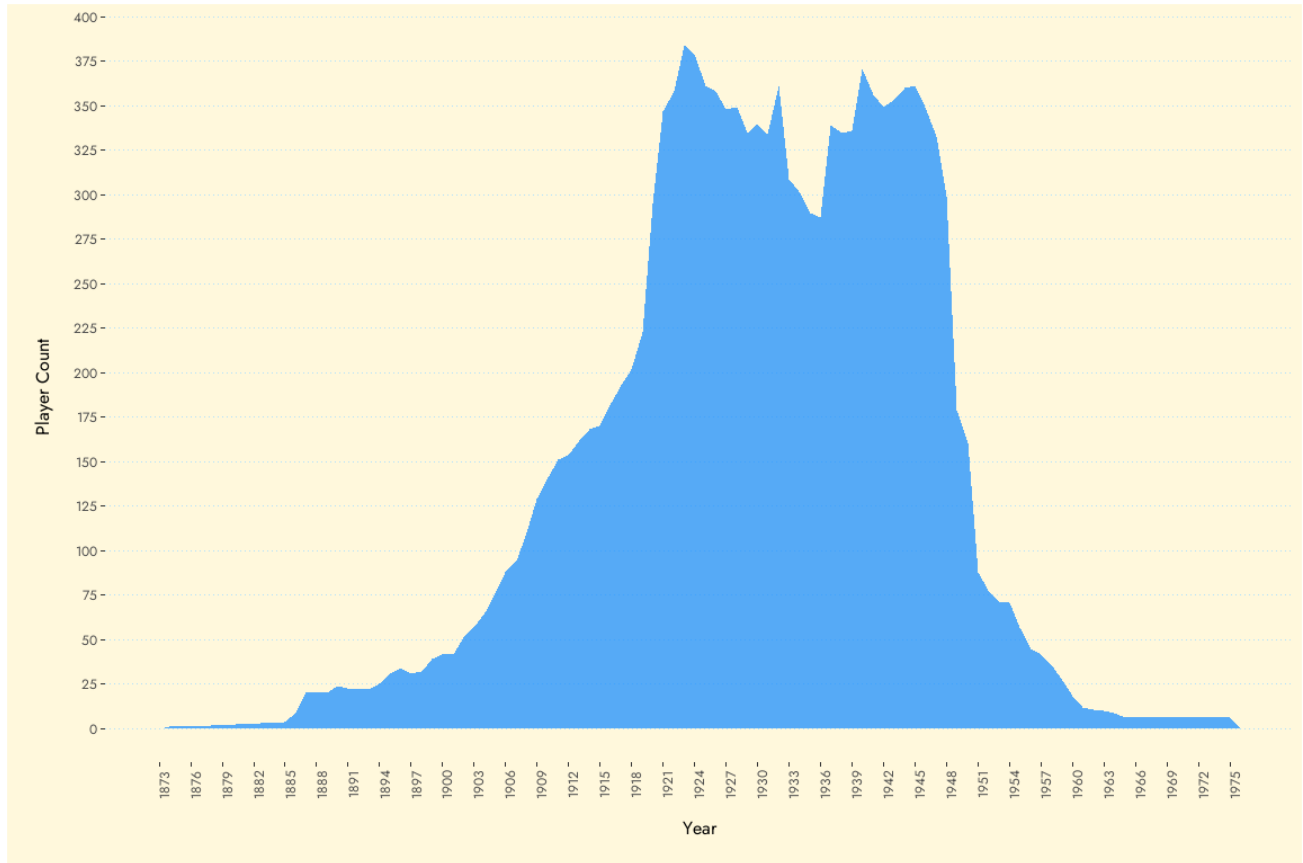


Figure 5: The estimated distribution of players in the Negro Leagues, 1873 to 1975: Wikipedia data.

In comparison, Figure 5 shows the estimated player distribution for the Negro Leagues from 1886 to 1948 according to the data collected from Seamheads. Figure 6, likewise, shows the estimated player distribution for the Negro Leagues according to the Wikipedia data but filtered to only include dates from the same range as the Seamheads data (1886 to 1948). A notable difference between Figure 5 and Figure 6 is that the peak for players in the Negro Leagues comes in 1932 for the Seamheads data ($N = 504$) while it comes in 1923 for the Wikipedia data ($N = 384$). Additionally, these total counts differ by a difference of 120. Again, the discrepancies between the data may be due to Wikipedia contributors excluding players, players with only one name in the Seamheads database being counted twice, or the imputation of 0 to replace NA values skews the Wikipedia distribution.

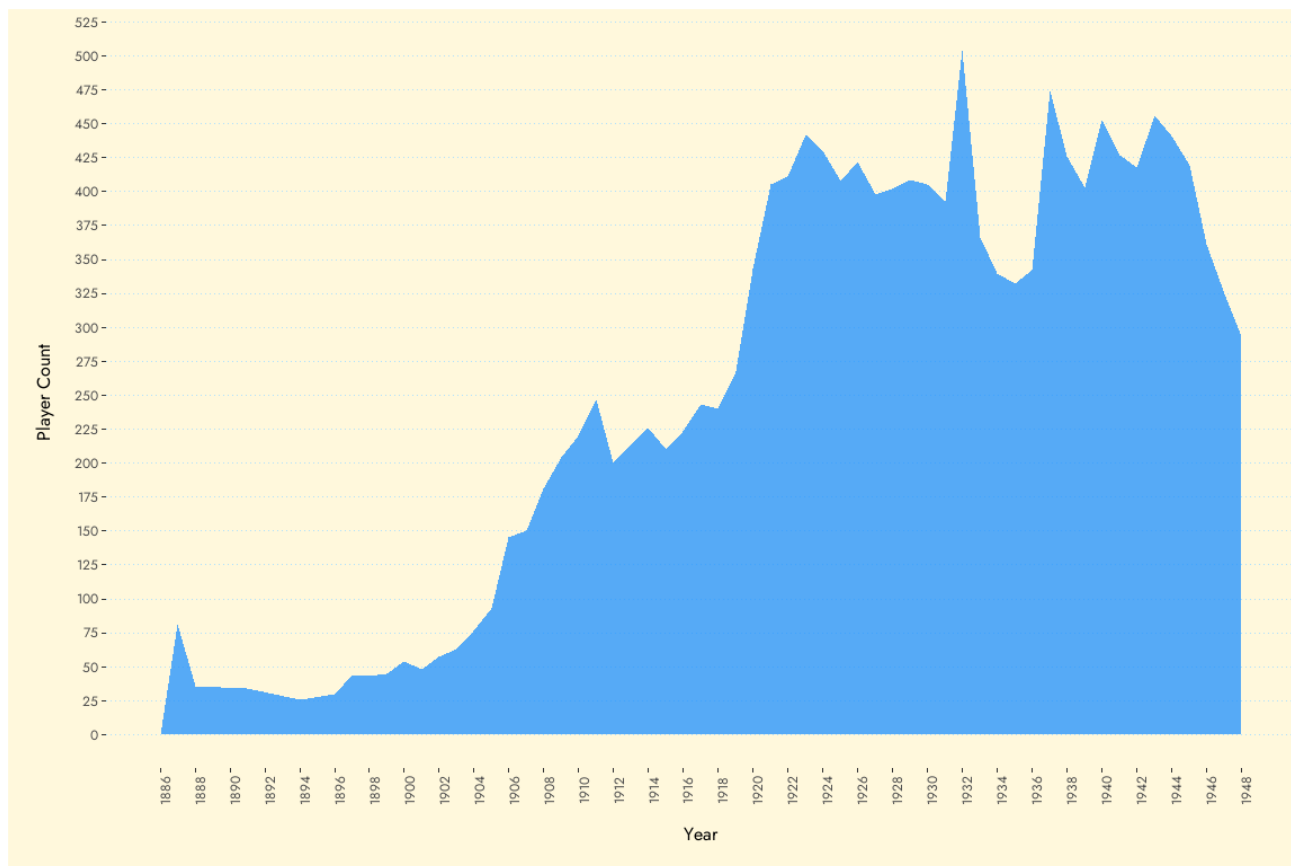


Figure 6: The estimated distribution of players in the Negro Leagues, 1886 to 1948: Seamheads data.

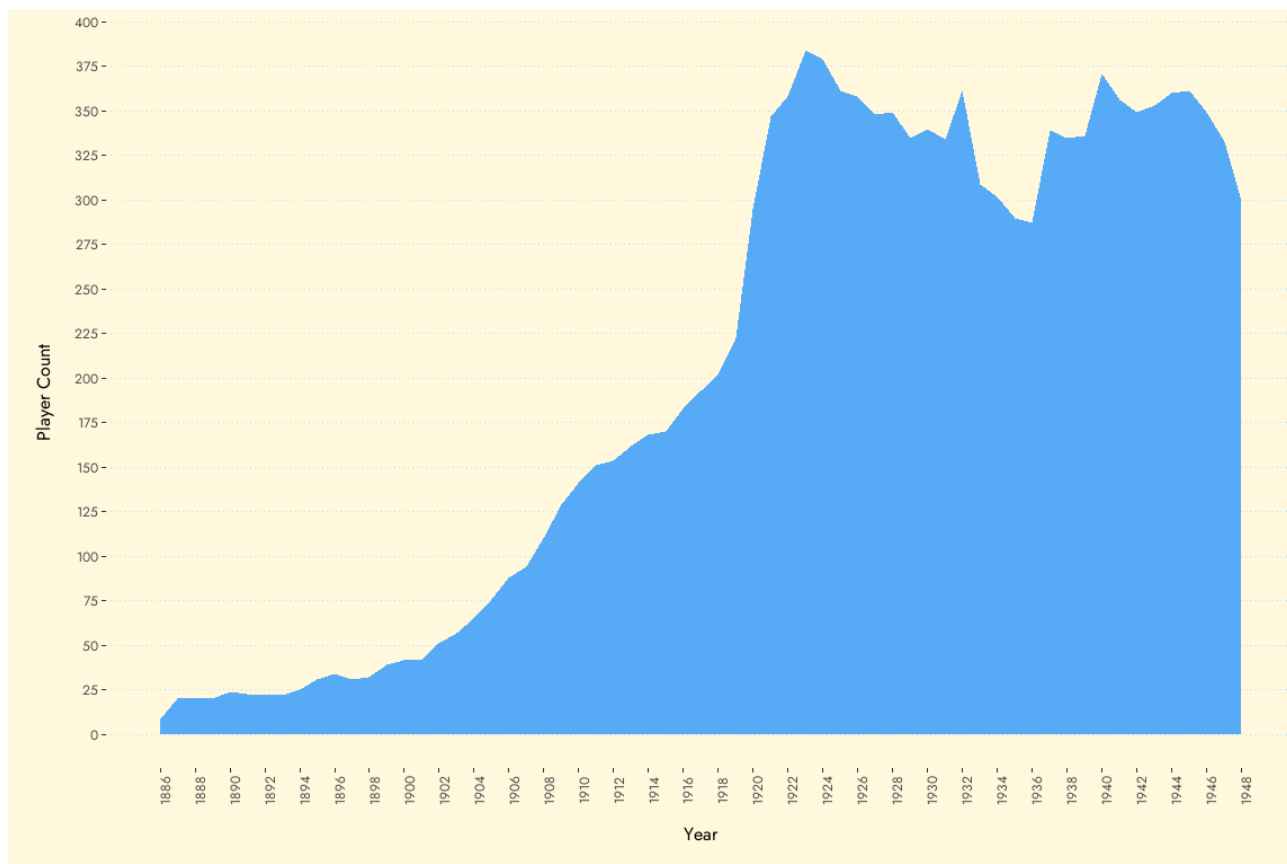


Figure 7: The estimated distribution of players in the Negro Leagues, 1886 to 1948: Wikipedia data.

Negro League Players in Major League Baseball

While Jackie Robinson was the first player from the Negro Leagues to cross the color line when he debuted for the Brooklyn Dodgers in 1947, he was not the only player to do so. Four others joined Robinson that year: Larry Doby for Cleveland (becoming the first Black player for the American Leagues), Hank Thompson and Willard Brown for the St. Louis Browns (team moved in 1953 to Baltimore and became the Baltimore Orioles), and Dan Bankhead also for the Brooklyn Dodgers (team moved in 1957 to Los Angeles and became the Los Angeles Dodgers). Over the years, the number of Negro League players who debuted for MLB teams fluctuated until 1978 when Harry Chappas became the last Negro League player to debut for an MLB team (Harry Chappas is actually white but played for the Indianapolis Clowns who are considered to be a Negro League team). In total, 87 players who played in the Negro Leagues also played for MLB teams. Figure 8 shows the years at which Negro League players debuted in MLB and the total count of players debuting that season. The greatest number of Negro League players to debut in a MLB season was 11 in 1954.

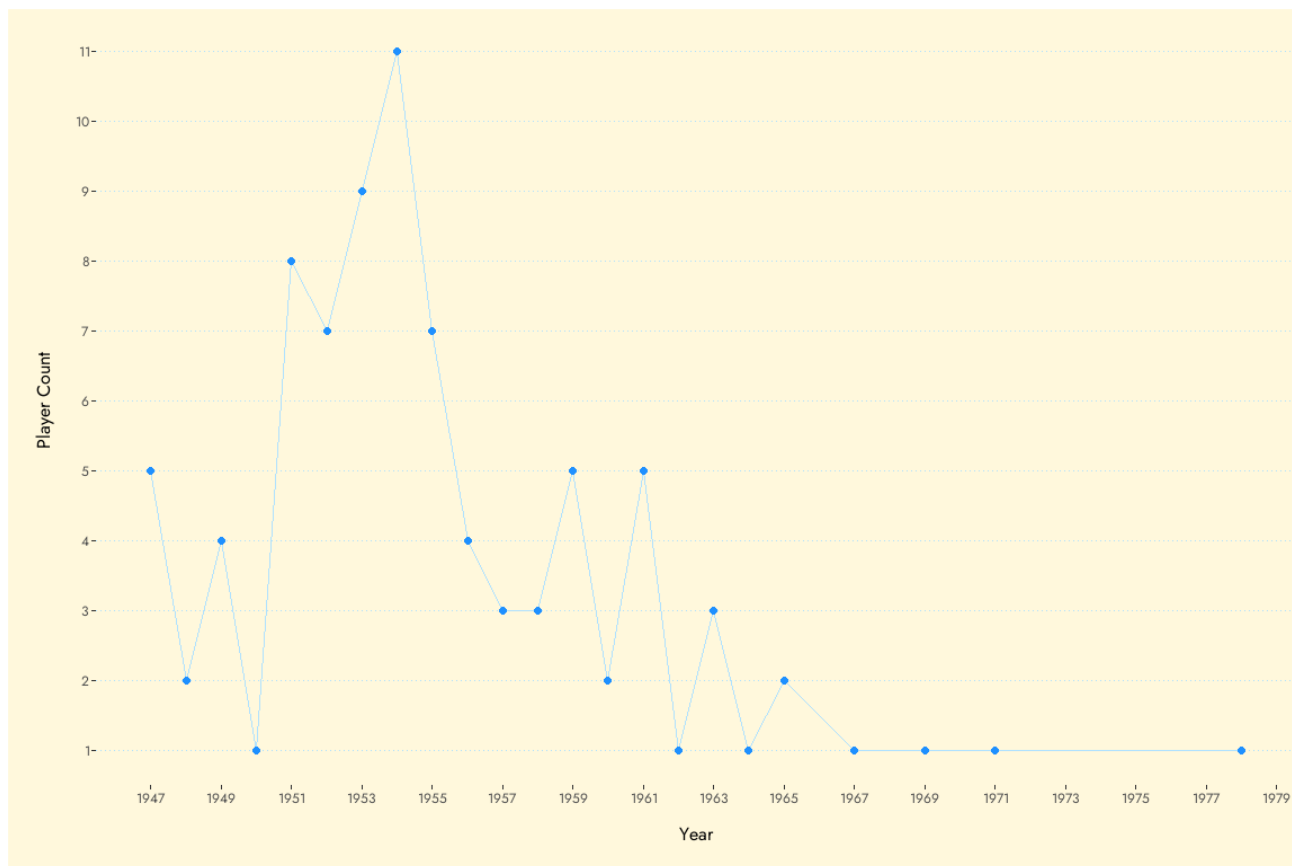


Figure 8: The number of players debuting for a given season in Major League Baseball who also played in the Negro Leagues.

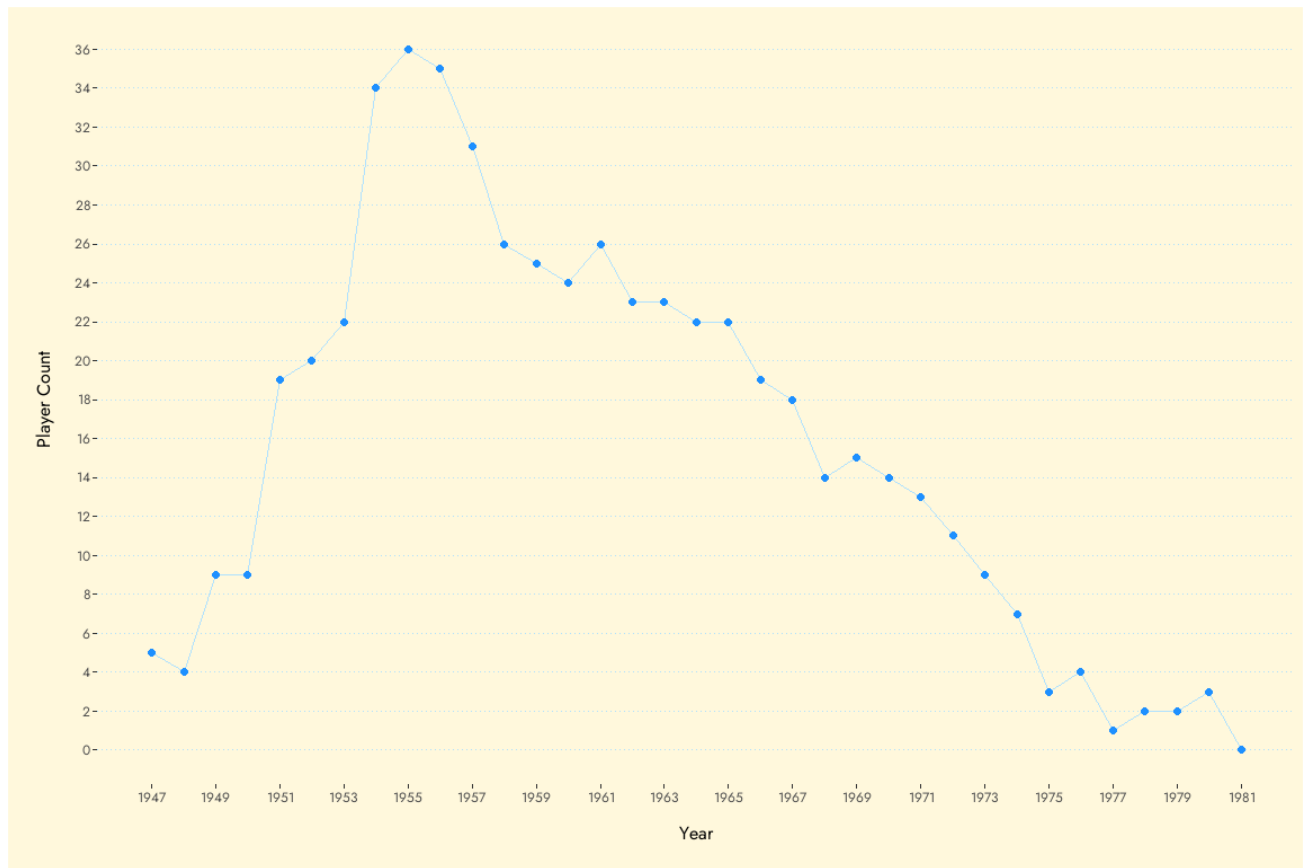


Figure 9: The number of players for a given season in Major League Baseball who also played in the Negro Leagues, 1947 to 1981.

Discussion

Appendix A: Tables

Table 1: Negro League Player Counts, 1887-1948: Seamheads Data.

Season	Player Debut Count	Player Final Game Count	Player Count
1887	81	55	81
1888	10	10	36
1891	8	18	34
1894	10	3	26
1896	7	3	30
1897	17	9	44
1898	8	7	43
1899	9	6	45
1900	15	10	54
1901	4	2	48
1902	11	4	57
1903	10	3	63
1904	16	5	76
1905	22	5	93
1906	57	41	145
1907	46	22	150
1908	53	34	181
1909	57	37	204
1910	53	33	220
1911	60	69	247
1912	22	26	200
1913	39	37	213
1914	50	62	226
1915	46	30	210
1916	43	30	223
1917	50	49	243
1918	46	31	240
1919	58	36	267
1920	112	71	343
1921	133	87	405
1922	93	78	411
1923	109	84	442
1924	72	88	430
1925	66	68	408
1926	82	85	422

Table 1: Negro League Player Counts, 1887-1948: Seamheads
Data. (*continued*)

Season	Player Debut Count	Player Final Game Count	Player Count
1927	61	61	398
1928	65	73	402
1929	80	75	409
1930	71	71	405
1931	59	66	393
1932	177	187	504
1933	49	69	366
1934	43	55	340
1935	47	50	332
1936	60	48	342
1937	180	137	474
1938	89	88	426
1939	65	60	403
1940	110	91	453
1941	65	72	427
1942	63	73	418
1943	111	107	456
1944	92	110	441
1945	88	145	419
1946	87	110	361
1947	74	119	325
1948	87	293	293

Table 2: Negro League Player Counts, 1874-1975: Wikipedia
Data.

Season	NL Player Debut Count	NL Player Final Game Count	NL Player Count
1874			1
1878	1		2
1885	2	0	4
1886	5		9
1887	11	1	
1889	1	0	20
1890	4	1	24
1891	0	2	23

Table 2: Negro League Player Counts, 1874-1975: Wikipedia
Data. *(continued)*

Season	NL Player Debut Count	NL Player Final Game Count	NL Player Count
1893	1	0	22
1894	3	1	25
1895	7		31
1896	6	3	34
1897	0		31
1898	4	1	32
1899	8	4	39
1900	7	3	42
1901	3	0	
1902	9	1	51
1903	7	2	57
1904	10	0	65
1905	11	4	76
1906	16	9	88
1907	15	8	94
1908	24	4	110
1909	23	8	129
1910	20	7	141
1911	17	15	151
1912	18	7	154
1913	15	11	162
1914	17	17	168
1915	19	10	170
1916	23		183
1917	20	8	193
1918	17	9	202
1919	30		223
1920	81	26	295
1921	78	33	347

Table 2: Negro League Player Counts, 1874-1975: Wikipedia
Data. *(continued)*

Season	NL Player Debut Count	NL Player Final Game Count	NL Player Count
1922	44	28	358
1923	54	42	384
1924	37	48	379
1925	30	39	361
1926	36	37	358
1927	27	32	348
1928	33	41	349
1929	27	32	335
1930	37	36	340
1931	30	31	334
1932	58	69	361
1933	17	34	309
1934	27	35	302
1935	23	25	290
1936	22	31	287
1937	83	48	339
1938	44	31	335
1939	32	19	336
1940	54	47	371
1941	32	37	356
1942	30	30	349
1943	34	36	353
1944	43	39	360
1945	40	69	361
1946	57	67	349
1947	51	78	333
1948	44	138	299
1949	18	42	179
1950	24	82	161
1951	9	26	88

Table 2: Negro League Player Counts, 1874-1975: Wikipedia
Data. (*continued*)

Season	NL Player Debut Count	NL Player Final Game Count	NL Player Count
1952	16	20	78
1953	13	11	71
1954	11	21	
1955	7	19	57
1956		12	45
1957	9	8	42
1958	2	13	36
1959	4	12	27
1960	3	8	18
1961	2	3	12
1962		1	11
1963	10		
1964	0	3	9
1965		1	6
1975	1		

Table 3: Negro League Players in MLB - Counts and Proportions: CNLBR and Retrosheet Data.

Season	MLB Population	NL Player Count	NL Players per MLB Population	NL Players per Team
1947	544	5	0.01	0.31
1948	542	4	0.01	0.25
1949	532	9	0.02	0.56
1950	530	9	0.02	0.56
1951	543	19	0.03	1.19
1952	556	20	0.04	1.25
1953	534	22	0.04	1.38
1954	536	34	0.06	2.12
1955	599	36	0.06	2.25
1956	552	35	0.06	2.19
1957	563	31	0.06	1.94
1958	580	26	0.04	1.62
1959	571	25	0.04	1.56
1960	575	24	0.04	1.50
1961	620	25	0.04	1.39
1962	702	22	0.03	1.10
1963	695	22	0.03	1.10
1964	700	22	0.03	1.10
1965	702	22	0.03	1.10
1966	714	18	0.03	0.90
1967	718	18	0.03	0.90
1968	676	14	0.02	0.70
1969	849	15	0.02	0.62
1970	849	14	0.02	0.58
1971	828	13	0.02	0.54
1972	824	11	0.01	0.46
1973	821	9	0.01	0.38
1974	861	7	0.01	0.29
1975	844	3	0.00	0.12
1976	818	4	0.00	0.17
1977	900	1	0.00	0.04
1978	904	2	0.00	0.08
1979	899	2	0.00	0.08
1980	916	3	0.00	0.12

Appendix B: Charts

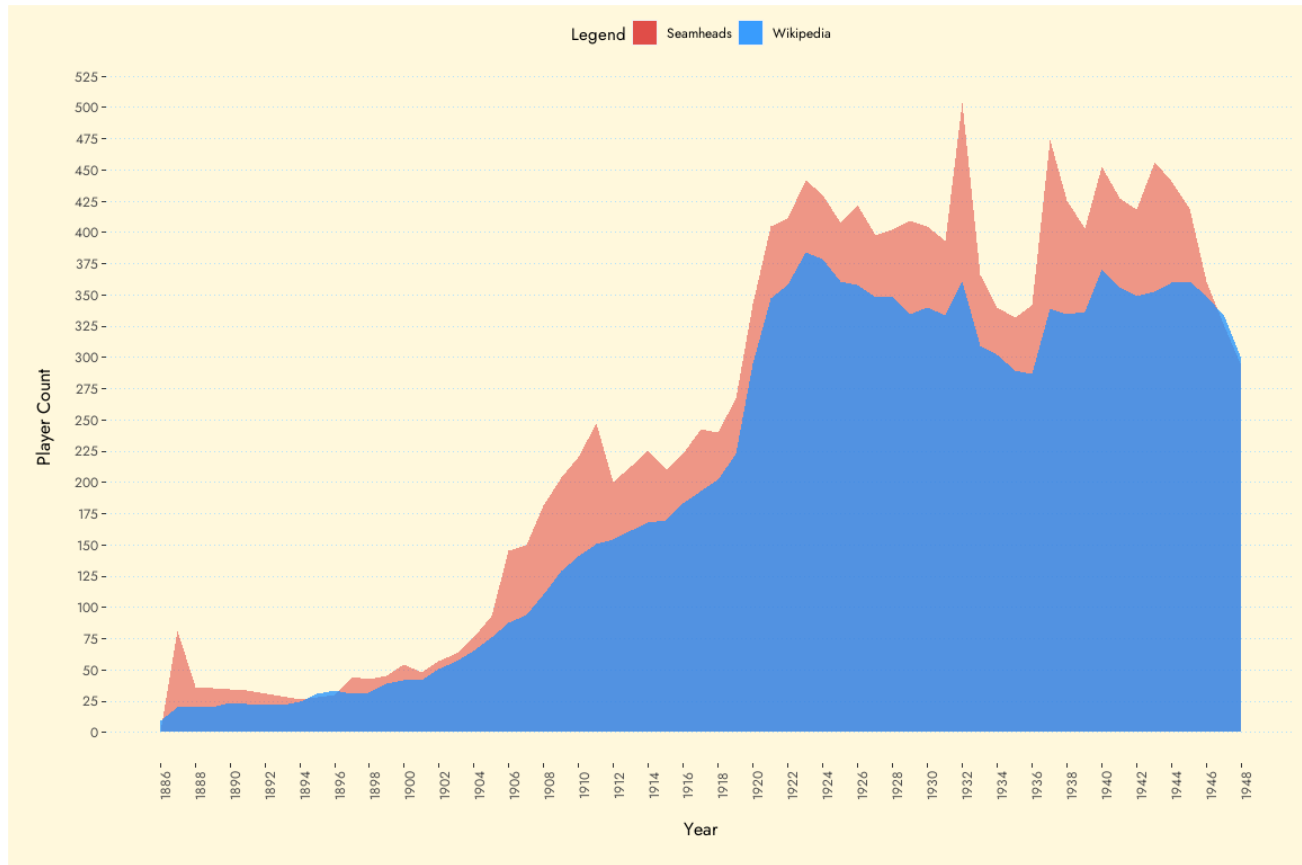


Figure 10: The estimated distribution of players in the Negro Leagues, 1886 to 1948: Wikipedia-Seamheads comparison.

References

- [1] Charlie Pavitt. *Major League Player Ethnicity, Participation, and Fielding Position, 1946-2018*. 2020. URL: <https://sabr.org/journal/article/major-league-player-ethnicity-participation-and-fielding-position-1946-2018/>.
- [2] CNLBR. *Center for Negro League Baseball Research*. 2021. URL: <http://www.cnlbr.org/Default.aspx>.
- [3] Colin Douglas and Richard Scriven. *retrosheet: Import Professional Baseball Data from 'Retrosheet'*. R package version 1.1.3. 2020. URL: <https://CRAN.R-project.org/package=retrosheet>.
- [4] Michael Friendly et al. *Lahman: Sean 'Lahman' Baseball Database*. R package version 9.0-0. 2021. URL: <https://CRAN.R-project.org/package=Lahman>.
- [5] Gary Ashwill and Patrick Rock and Scott Simkus and Larry Lester and Larry Lester. *Seamheads Negro League Database*. 2018. URL: <https://www.seamheads.com/NegroLgs/index.php>.

- [6] Mark Armour and Daniel R. Levitt. *Baseball Demographics, 1947-2016*. 2016. URL: <https://sabr.org/bioproj/topic/baseball-demographics-1947-2016/>.
- [7] Center for Negro League Baseball Research. *Negro League Players Who Made the Major Leagues*. URL: <http://www.cnlbr.org/Portals/0/RL/Negro%20League%20Players%20Who%20Played%20in%20the%20Major%20Leagues.pdf>.
- [8] Neil Paine and Anna Wiederkher and Jay Boice. *The Negro League stars that MLB kept out - and is finally recognizing*. 2021. URL: <https://projects.fivethirtyeight.com/negro-leagues-mlb/>.
- [9] Jeroen Ooms. *pdftools: Text Extraction, Rendering and Converting of PDF Documents*. R package version 2.3.1. 2020. URL: <https://CRAN.R-project.org/package=pdfutils>.
- [10] Bill Petti. *baseballr: A package written for R focused on baseball analysis*. R package version 0.8.4. 2021. URL: <http://billpetti.github.io/baseballr/>.
- [11] Hadley Wickham. *rvest: Easily Harvest (Scrape) Web Pages*. R package version 1.0.0. 2021. URL: <https://CRAN.R-project.org/package=rvest>.
- [12] Wikipedia contributors. *List of Negro league players — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/wiki/List_of_Negro_league_baseball_players (visited on 04/16/2021).