Measuring the Effectiveness of the Designated Hitter in Major League Baseball

Rick Bump Ben Harwood Xavier Noel Kevin Schaeffer

2019-12-09 20:27:54

Abstract

Major League Baseball tasked our team with studying the effectiveness of the designated hitter in the American League and making a recommendation for or against the National League adopting the designated hitter as well. Using batting statistic data, we compare batting trends as well as look at how different statistics influenced runs over the last century or so. We ultimately see that while there are distinct increases in hitting statistics after the institution of the designated hitter, we find through linear regression that many of the stats' contributions toward runs are effectively equal between the leagues, and we find that after the designated hitter implementation in 1973, every stat not only influences the National League more than the American League but also has lower predictive inlfuence on the American Lague than it did prior to 1973. We ultimately conclude and recommend that the National League should adopt the desingated hitter.

Contents

1 Contribution Statement

Rick Bump: t-test development, editing, visualization development, model development

Ben Harwood: Model, visualization, and R function development, markdown file production

Xavier Noel: Visualization development, exploratory data analysis, model development

Kevin Schaeffer: Group 1 administrative assistant, model development, presentation creation, junior baseball historian.

2 Introduction

In 1887, Major League baseball adapted the rules to allow extra players to be assigned to a team's roster and be substituted into the game between innings. Changes like these, in the early days of baseball, generated a period of pitching dominance that ran into the late 1960's. Of Major League Baseball's two Leagues, the American League was lagging behind the National League in both scoring and attendance. In 1973, the

Designated Hitter was approved in the American League through rule 5.11. The rule allows for one player, or Designated Hitter, to bat in a fixed spot in the lineup as a substitute to the pitcher.

Forty-six years later, the American League (AL) still uses the DH, while the National League (NL) has continued to vote against adopting the change. In 1997, Major League Baseball introduced play between American and National League teams during the regular season. In those games, the rules of the home team are used. Major League Baseball (MLB) has taken steps over the last decade attempting to increase interest amongst younger fans through increased offense, a similar challenge the American League faced in the 1960s.

Major League Baseball tasked our team with determining if the NL should, in fact, adopt the DH. We have developed a data analysis and recommendation for Major League Baseball to determine whether or not the National League should adopt the Designated Hitter. We evaluated the trends in offensive statistics in both leagues, before and after 1973. Additionally, we used relationship based models to identify if there is a correlation between the use of the Designated Hitter and Offense, or runs scored.

3 Data

MLB has made available a collection of batting data going all the way back to 1871 (not long after the very first team, the Cincinnati Red Stockings, was founded in 1869). The data may also be found at this link: https://www.kaggle.com/open-source-sports/baseball-databank/#Batting.csv. We have four main objectives:

1. Evaluate the trends and changes in the distribution of hits (singles, doubles, triples, HR) over time, by team and league.
2. Evaluate the change in American League Batting after the switch to using a DH.
3. Compare AL and NL hitting from the time the DH was introduced.
4. Determine the impact of batting statistics on runs scored

The data set includes quite a bit of information:

playerID	yearID	stint	teamID	lgID	G	AB	R	Singles	Doubles	Triples	HR	RBI	SB	CS	BB	SO	IBB	HBP	SH	SF	GIDP
abercda01	1871	1	TRO	NA	1	4	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA	NA
addybo01	1871	1	RC1	NA	25	118	30	32	6	0	0	13	8	1	4	0	NA	NA	NA	NA	NA
allisar01	1871	1	CL1	NA	29	137	28	40	4	5	0	19	3	1	2	5	NA	NA	NA	NA	NA
allisdo01	1871	1	WS3	NA	27	133	28	44	10	2	2	27	1	1	0	2	NA	NA	NA	NA	NA
ansonca01	1871	1	RC1	NA	25	120	29	39	11	3	0	16	6	2	2	1	NA	NA	NA	NA	NA
armstbo01	1871	1	FW1	NA	12	49	9	11	2	1	0	5	0	1	0	1	NA	NA	NA	NA	NA

Most of this data we do not need. There are actually six different leagues represented in the data, as well as no league identification until 1876. Because we are focusing on the NL and AL, we will exclude anything prior to 1901 (the American League did not exist prior to 1901) as well as data from the other leagues.

The data has an extensive amount of missing values. This is to be expected, as any AL pitcher will not have batting data after 1973. Additionally, there are other instances of players with 0 at-bats (dedicated pinch-runners, for example), so we remove those as well.

3.1 Exploratory Data Analysis

As we are focused only on the various hitting statistics, we only really need five of the variables.

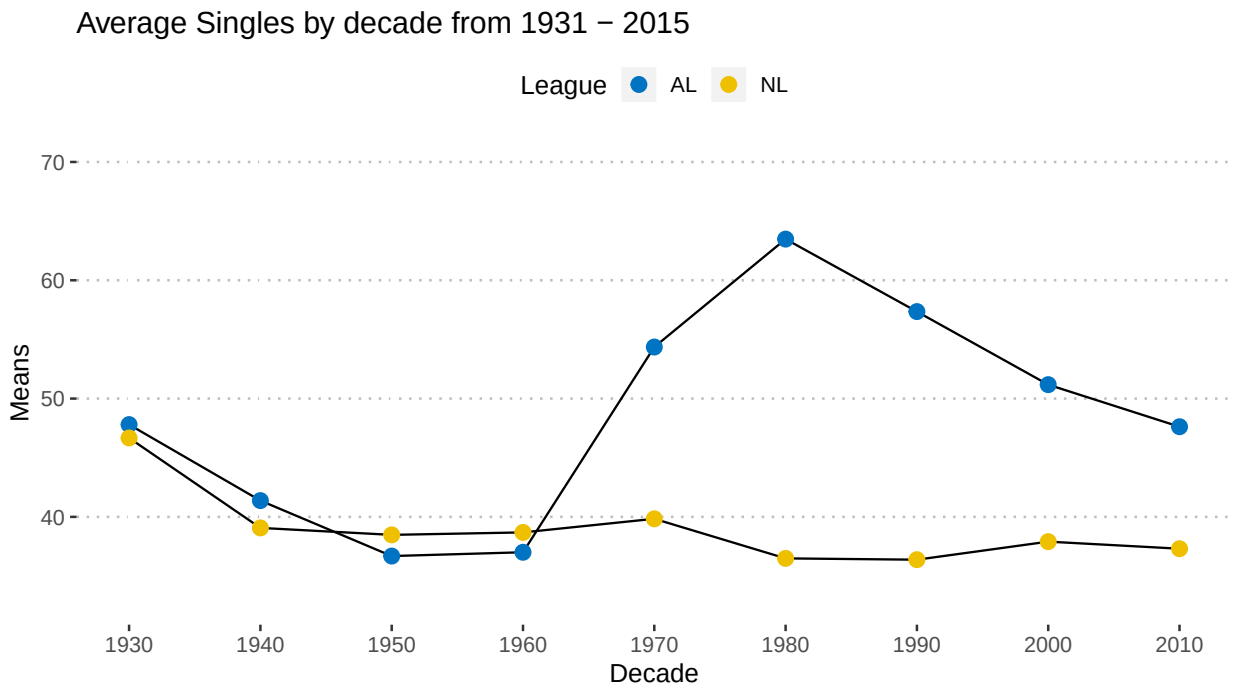
Table 1: National League

	min	max	range	median	mean	std.dev
Singles	0	237	237	10	38	53
Doubles	0	64	64	2	7	10
Triples	0	21	21	0	1	2
HR	0	73	73	0	4	7
RBI	0	160	160	4	18	26

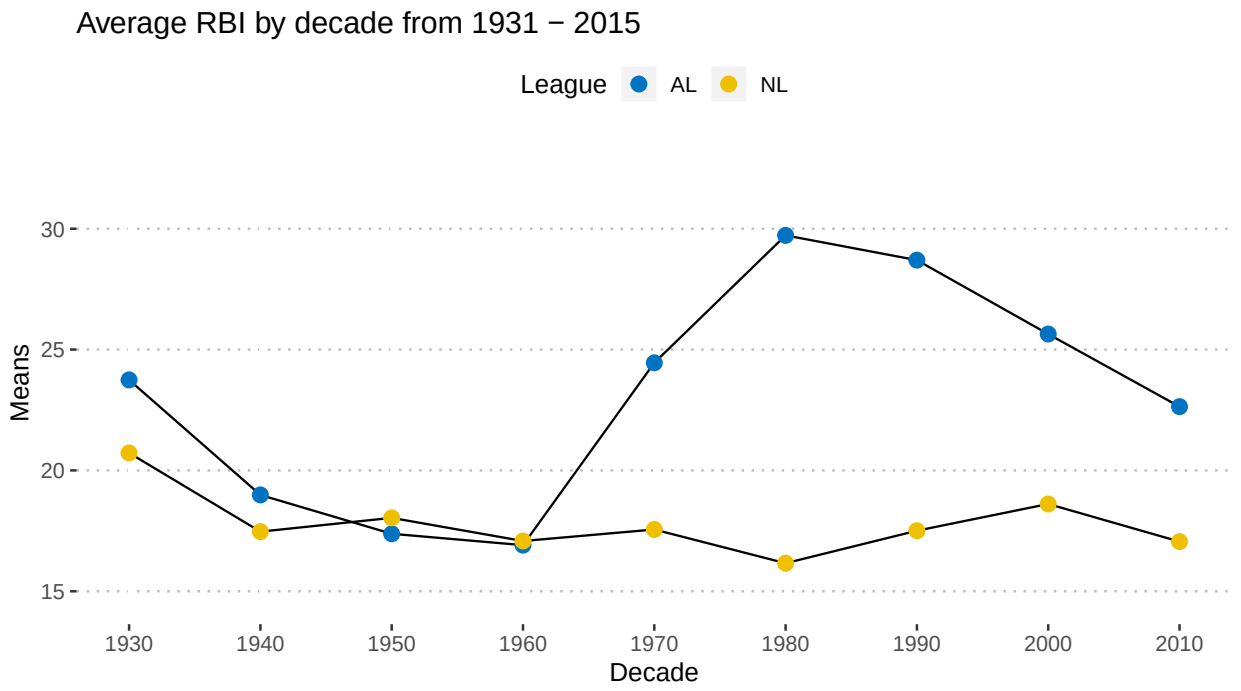
Table 2: American League

	min	max	range	median	mean	std.dev
Singles	0	262	262	21	49	56
Doubles	0	67	67	3	9	11
Triples	0	23	23	0	1	2
HR	0	61	61	1	5	8
RBI	0	184	184	9	23	30

The descriptive statistics above are, obviously, for each of the four hit categories (and RBI's while we're at it). This is reflective of data from 1931 to 2015, which puts our year of interest right in the middle. Nothing especially enlightening, so let's look at singles and RBI's.

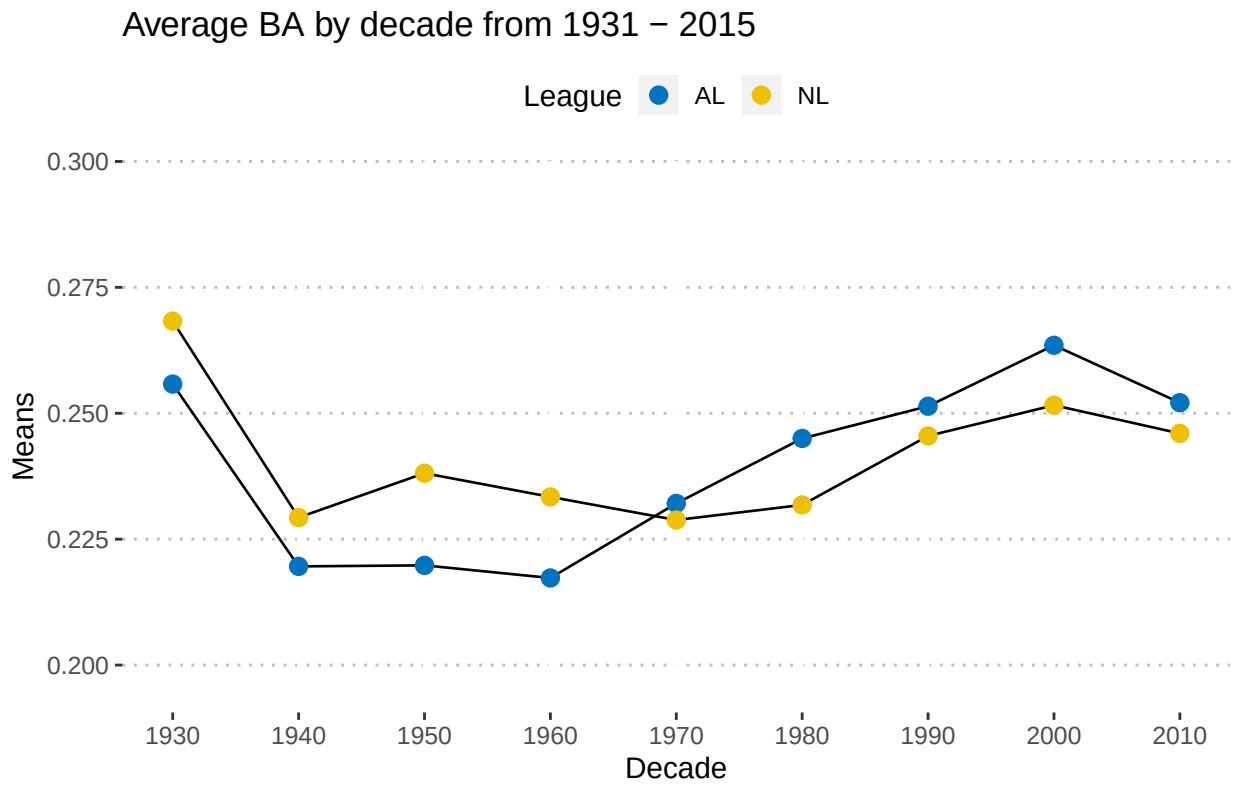


It should come as no surprise to see the sharp increase on the AL side in the early 1970s given that the DH came to being in 1973. This is, perhaps unsurprisingly, by the RBI data.



Here's one that's really interesting: batting average. Batting average(BA) is not included in the initial data set, but we can add it using the following formula (BB stands for walks)

$$BA = \frac{H + 2B + 3B + HR - BB}{AB}$$

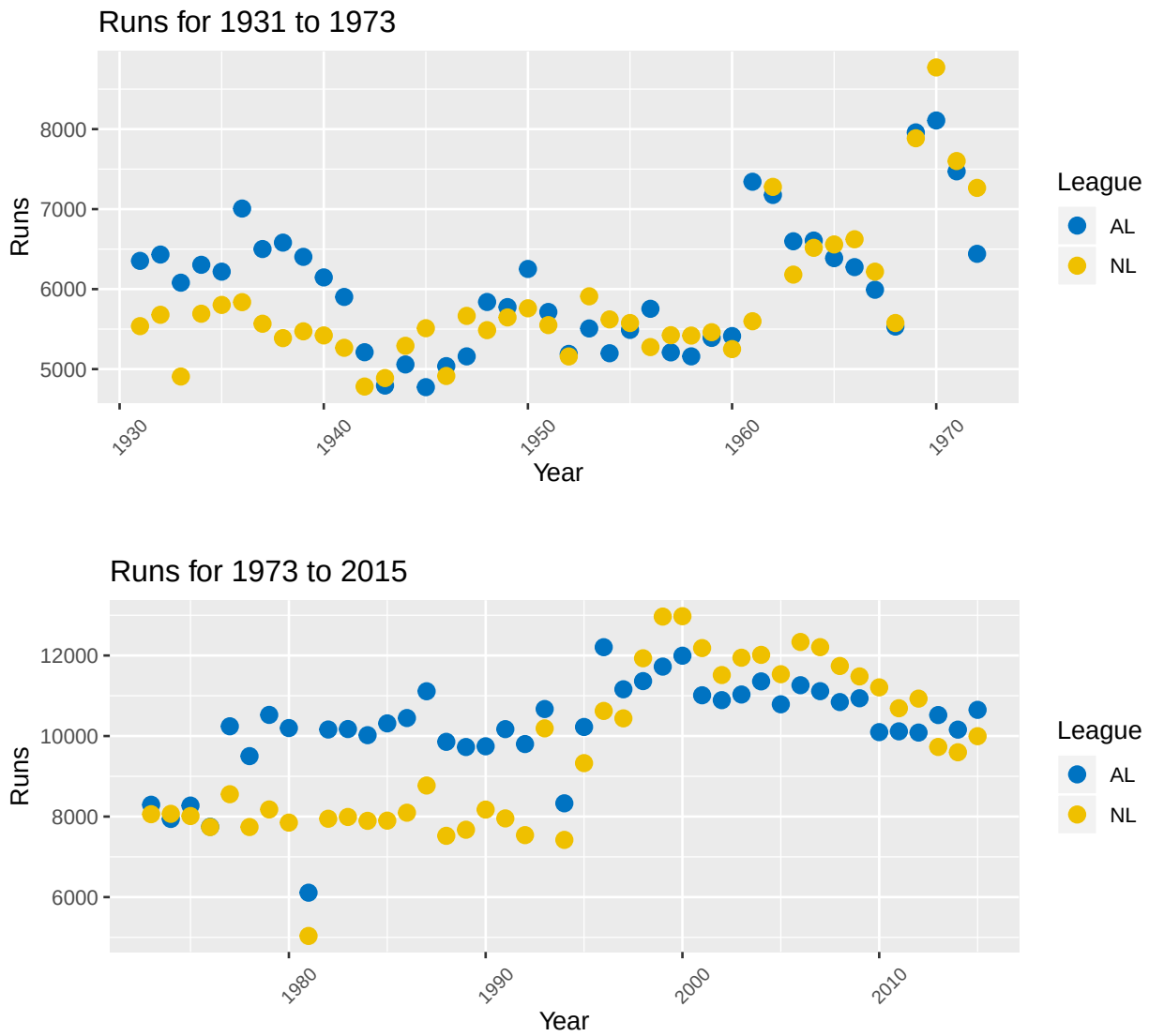


Here again we see the shift in leadership between the two leagues in the 1970's. If our goal is to generate more exciting games through offense, a strong case in favor of the DH is beginning to take shape.

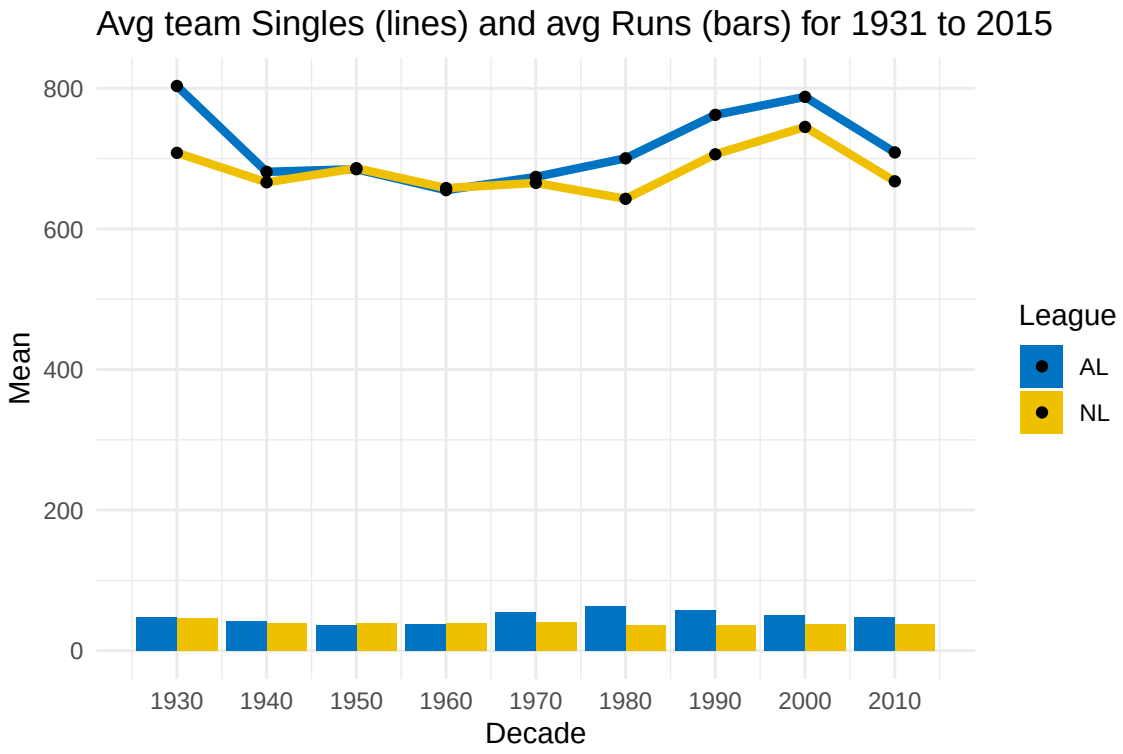
4 Deeper Analysis

Let's begin our deeper dive by looking at how many runs each league scored.

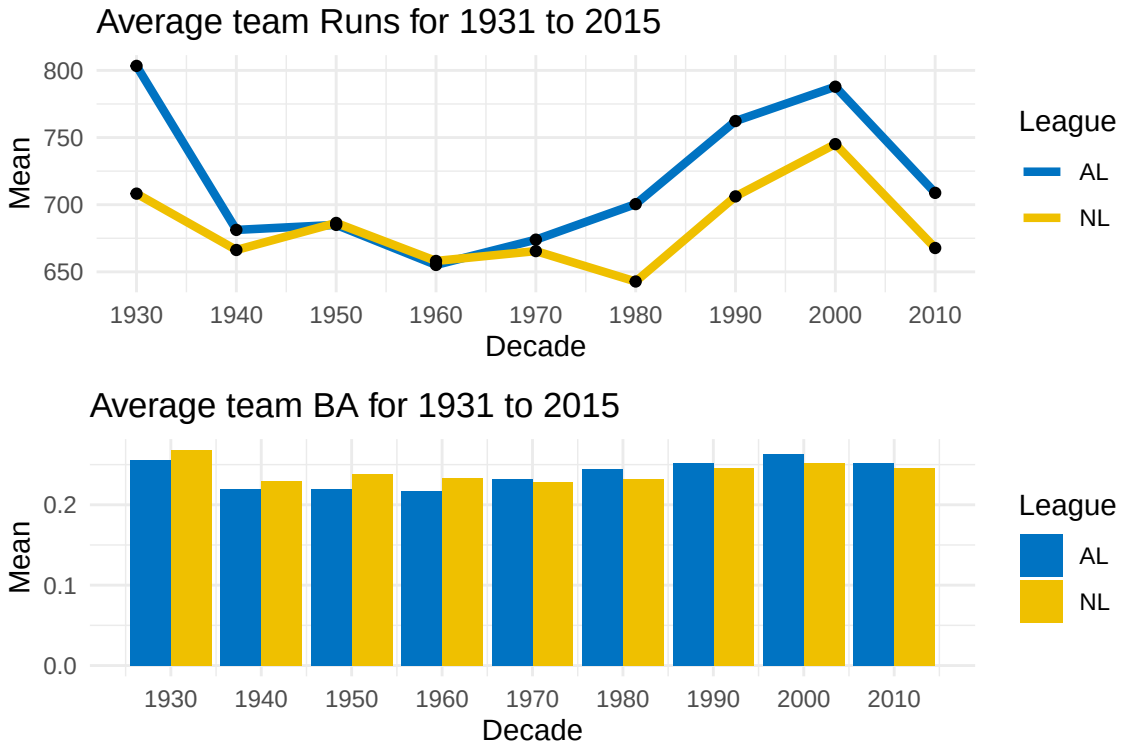
4.1 Runs



This is perhaps our first... "interesting" result, at least on the surface. Until 1976, the leagues were fairly well matched in terms of run counts. Then the AL run count jumped by over 1000 runs while the NL stayed the same. This is actually due to the AL adding two teams (Seattle and Toronto) while the NL stayed at twelve teams. Additionally, the sudden overtake by the NL in 1998 was due to further expansion, bringing the team counts to 16 and 14 for the NL and AL, respectively. Then in 2013, one team was moved from the NL to the AL to even the team distribution which closed the gap between the two leagues in their run counts. Let's see if we there is any relationship between runs and different batting stats.



This is an interesting plot. It shows a clear discrepancy between the two leagues' singles counts, and the run counts mirror this behavior. This is our first real indicator of the effect of the insitution of the DH into the AL. More enlightening is batting average.



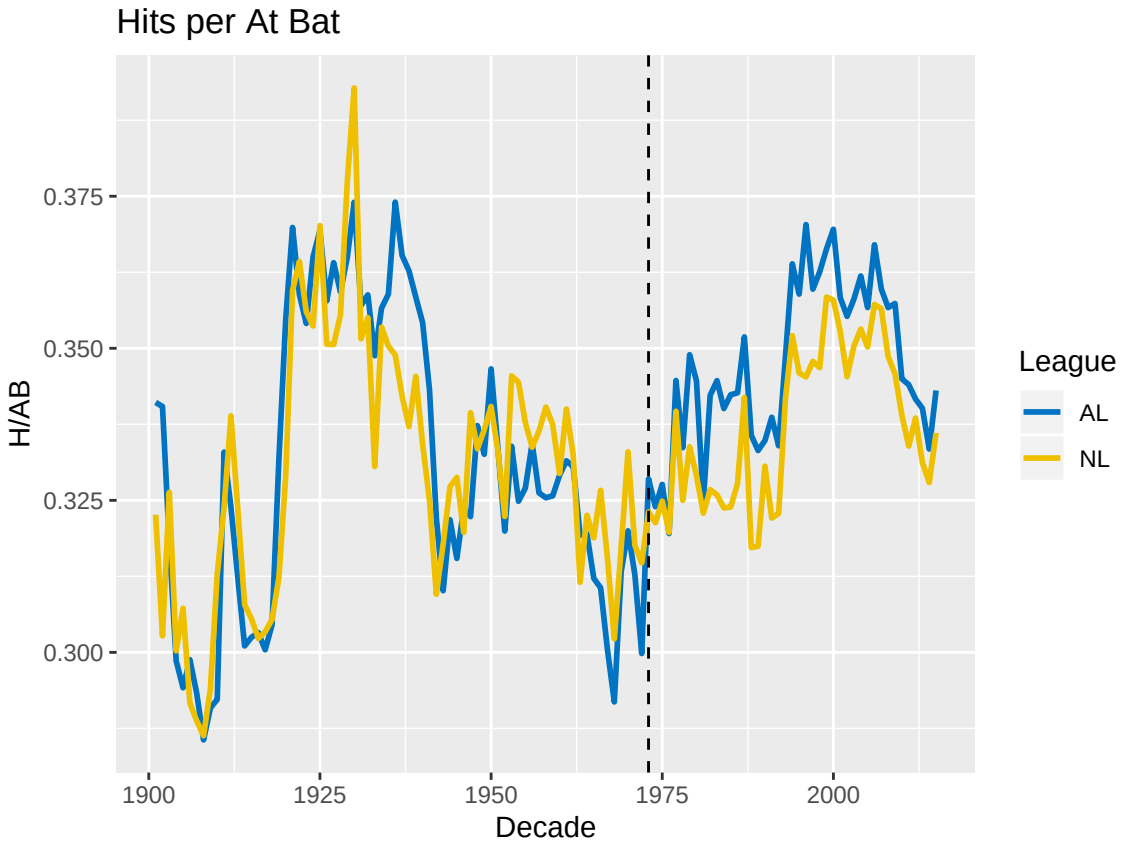
Notice the dramatic shift in power from the NL to the AL in the 70's in both batting average and runs scored.

The case in favor of adding the DH to NL is getting stronger.

4.2 Models

We became curious to see the hits per at-bat ratio differed between leagues before and after the DH introduction. Intuitively, one would suspect that because of the removal of the "weak hitters" that the AL would lead in this consideration after 1973, and indeed the plot reflects this. However, in considering the potential impact of the the use of the designated hitter for the generation of offense we used a single-sample t-test (testing for AL hits/at-bat being greater than NL hits/at-bat). We created a portable metric of hits/at-bats to remove any fluctuation due to games played, shortened seasons, etc. in the ratio of hits/at-bat for the AL against that of the NL for the period of 1973 through 2015.

The t-test is significant ($p = 2.1 \times 10^{-4}$), with a mean for the AL of 0.3478666 and a mean for the NL of 0.3375106. In other words, there is a 99.98% chance that there is an increase in offense generated by using the designated hitter.



One thing we could do, while we're here, is to exploit the power of statistics and perform a regression to see how well single base hits predict runs. We would like to do this for every batting stat, but we must run diagnostics on the data first before regression can be reliably interpreted. The following table shows the result of the Breusch Pagan test for heteroscedasticity, and whether the data passes the test. Anything marked "no" (based on significance of the test) will need to be transformed before regression is trustworthy.

Singles, doubles, hits, batting average, and hits per at-bat must be transformed. Once we do that then we can proceed with our regression models. Similarly to the Breusch-Pagan test, we present a table with the R^2 and significance of each variable.

Table 3: Breusch Pagan Reults

	BP.PValue	Use
Singles	0.0003121	No
Doubles	0.0000101	No
Triples	0.9686000	Yes
Homers	0.5512000	Yes
Hits	0.0016990	No
BA	0.0000000	No
Hits_AB	0.0000000	No

Table 4: Regression Results

	P.Value	R.2
tSingles	0.0000000	0.9735736
tDoubles	0.0000000	0.9644682
Triples	0.3221802	0.0042987
Homers	0.0000000	0.9075614
tHits	0.0000000	0.9824581
tBA	0.0000000	0.1270526
tHits_AB	0.0000000	0.2503034

These results are very enlightening. They show that each of the batting statistics do contribute toward the run count, however they also show that some statistics contribute very little. We can see that overall hits account for the most variance (this shouldn't be a surprise, since hits are basically required for runs). Clearly, triples are not especially helpful due to the insignificant regression. This could be due to the small number of triples that are hit. But also notice batting average only accounting for 12.7% of the variance. This is surprising, as conventional wisdom would suggest that teams with strong batting averages should score more runs, while weaker batting average teams would score fewer runs. In other words, overall batting performance has very little influence on whether a team wins or loses. Let's now do the same thing for each league to see if there is any noticeable difference.

Table 5: National League

	P.Value	R.2
tSingles	0.0000000	0.9783859
tDoubles	0.0000000	0.9670805
Triples	0.4606865	0.0048258
Homers	0.0000000	0.9155626
tHits	0.0000000	0.9853820
tBA	0.0517293	0.0330798
tHits_AB	0.0000010	0.1922817

Table 6: American League

	P.Value	R.2
tSingles	0.0000000	0.9732759
tDoubles	0.0000000	0.9661138
Triples	0.0648142	0.0298531
Homers	0.0000000	0.9019361
tHits	0.0000000	0.9834478
tBA	0.0000000	0.2760241
tHits_AB	0.0000000	0.3015796

Notice here that five of the stats are effectively the same between the two leagues, except batting average. In the AL, we can account for 27.6% of the variance with batting average, compared to a meager 5.2% for the NL. Let's conclude this portion of the analysis by focusing on the two time windows we have considered previously.

Notice that for the 42 years prior to the DH implentation, only hits and maybe singles could reliably predict runs. Now, let's look at the 42 years after the DH implementation.

Table 7: National League, 31-73

	P.Value	R.2
tSingles	0.0000000	0.9084979
tDoubles	0.0000000	0.7629461
Triples	0.0002373	0.2835399
Homers	0.0000000	0.5796386
tHits	0.0000000	0.9410601
tBA	0.2131539	0.0375413
tHits_AB	0.1904005	0.0414453

Table 8: American League, 31-73

	P.Value	R.2
tSingles	0.0000000	0.8360352
tDoubles	0.0000000	0.8114981
Triples	0.1131985	0.0600623
Homers	0.0000002	0.4933695
tHits	0.0000000	0.8746687
tBA	0.1621778	0.0470911
tHits_AB	0.2846007	0.0278725

Table 9: National League, 1973-2015

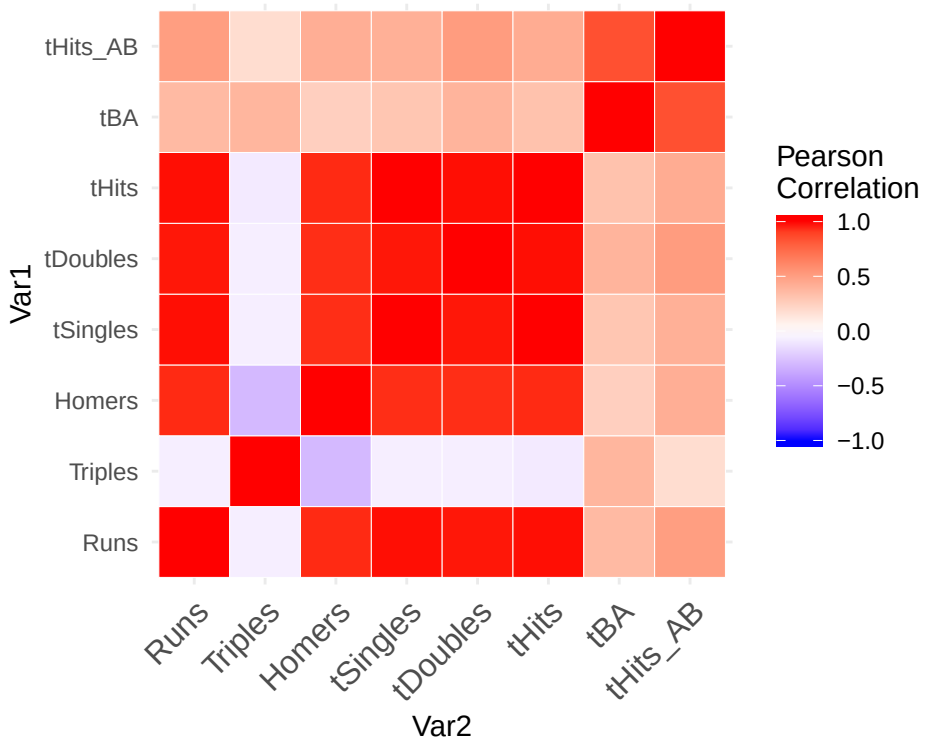
	P.Value	R.2
tSingles	0.0e+00	0.9670333
tDoubles	0.0e+00	0.9401378
Triples	5.9e-06	0.3974828
Homers	0.0e+00	0.9653813
tHits	0.0e+00	0.9748011
tBA	0.0e+00	0.5879834
tHits_AB	0.0e+00	0.7107554

Table 10: American League, 1973-2015

	P.Value	R.2
tSingles	0.0000000	0.8619075
tDoubles	0.0000000	0.8467441
Triples	0.3237282	0.0237559
Homers	0.0000000	0.8338884
tHits	0.0000000	0.9272475
tBA	0.0000000	0.6170260
tHits_AB	0.0000000	0.6200726

We showed earlier that there was a 99.98% chance that using a designated hitter would lead to more offense. Here is evidence that having a designated hitter makes predicting runs scored less reliable. This does not counter the previous result, however.

A natural next step would be to test for variable interactions, but first we must check correlation:



Clearly, variable interaction is not something we can really consider given the immense level of correlation.

5 Conclusions

We recommend that Major League Baseball fully adopt the utilization of designated hitters. We evaluated the trends and changes in the distribution of hits in both leagues throughout the life of MLB and there is a significant change after the introduction of the designated hitter in the American League. After confirming there was a positive shift in offense generation, we compared the American and National Leagues to see how significant the shift was. Our analysis has revealed that the addition of the designated hitter in the American League has provided a lift in offense, RBIs and hits have increased and utilizing a single sample t-test we were able to confirm with over 95% confidence that the utilization of a designated hitter increases offense. We believe that adopting the designated hitter throughout the major leagues will increase attendance, viewership and profit.

We would also like to conduct further study of the following questions:

1. Does an increase in offense lead to increased ticket sales?
2. Does an increase in offense lead to an increase in consumer engagement?
3. Does an increase in offense broaden the potential market outside of traditional baseball fans?
4. Do fans have an increase in excitement when designated hitters come up to bat, potentially increasing enjoyment of the game?
5. Does the increase of offense lead to better or more exciting defensive plays that generate consumer engagement?
6. Would the complete adoption of a designated hitter marginalize baseball purists?

We believe answering these questions would provide a fuller picture and aid in the final decision of whether to institute the designated hitter for both leagues.

6 References

[Bro19] M. Brown. *MLB Proposes Bringing The DH To The National League*. 2019. URL: https://www.forbes.com/sites/maurybrown/2019/02/05/mlbpa-proposes-bringing-the-dh-to-the-national-league/#6c53a1093d75 (visited on 12/08/2019).

[Cro16] J. Cronin. *The Historical Evolution of the Designated Hitter Rule*. 2016. URL: https://sabr.org/research/historical-evolution-designated-hitter-rule (visited on 12/01/2019).

[Edi09] H. Editors. *American League adopts designated hitter rule*. 2009. URL: https://www.history.com/this-day-in-history/american-league-adopts-designated-hitter-rule (visited on 12/04/2019).

[Ker19] J. Keri. *MLB rule changes: The universal DH makes sense, and it's time for the National League to make the switch*. 2019. URL: https://www.cbssports.com/mlb/news/mlb-rule-changes-the-universal-dh-makes- sense-and-its-time-for-the-national-league-to-make-the-switch/ (visited on 11/01/2019).

[Unk19a] Unknown. *Designated Hitter/ExtraHitter*. Nov. 11, 2019. URL: https://www.umpirebible.com/index.php/rules-batting/designated-hitter-extra-hitter (visited on 11/11/2019).

[Unk19b] Unkown. *5.11 (6.10) Designated Hitter Rule*. Dec. 01, 2019. URL: https://baseballrulesacademy.com/user/official-rules/mlb/5-11-6-10-designated-hitter-rule/ (visited on 12/01/2019).

[Win19] H. Winnicki. *MLB commissioner Rob Manfred says don't expect the DH to come to the National League before 2021*. 2019. URL: https://www.newsday.com/sports/baseball/dh-national-league-rob-manfred-1.30552878 (visited on 12/08/2019).