Baseball Sabermetrics

A Closer Look at Advanced Baseball Statistics and Their Change Over Time

Nathan Shaver
109410328
University of Colorado Boulder
CSCI 4502
nash4637@colorado.edu

Thomas Hawley
109706364
University of Colorado Boulder
CSCI 4502
thomas.hawley@colorado.edu

ABSTRACT

Sabermetrics is the empirical analysis of baseball, especially baseball statistics that measure in-game activity. These activities measure pitching, batting, offensive and defensive metrics. Analysis of these statistics has resulted in more advanced statistical measurements that coaches and team front offices use to analyze players that simpler statistics wouldn't offer. At its foundation, sabermetrics raises questions about how baseball is played and the most effective strategies to achieve, then attempts to answer them empirically. We used pybaseball, a Python library for storing baseball data, to analyze various relationships between different baseball statistics and extrapolate that relationship by stressing the data with multiple variables. As such, our general null hypothesis is as follows: various events and/or rule changes have no impact on the overall statistical trends throughout Major League Baseball.

INTRODUCTION

To start, let's take a look at several classic pitching statistics. These metrics have been used in the sport for over a century.

- ERA (Earned Run Avg)
- IP (Innings Pitched)
- BB (Base on Balls)
- K (Strikeouts)
- H (Hits Allowed)

These statistics follow a fairly consistent pattern. How many walks did the pitcher give up in a season? How many batters did the pitcher strike out in a season? They are all cumulative metrics. Now let's take a look at several advanced sabermetric statistics that have appeared in the last 20 years.

- WHIP (Walks plus hits per Innings Pitched)
- BABIP (Batting average on balls in play)
- FiP (Fielding Independent Pitching)
- SIERA (Skill-interactive Earned Run Avg)
- BQR (Bequeathed Runners Scored)

These metrics take the classic statistics and convert them into measurable data. For example, WHIP takes into account a pitcher's ability to keep batters off the basepaths. It combines walks and hits allowed per inning pitched.

Through the use of these metrics and tools described in the background, we examined various relationships between different baseball statistics and various rule changes. We graphed and analyzed the relationships on various plots. From there we drew conclusions on the results we found.

RELATED WORK

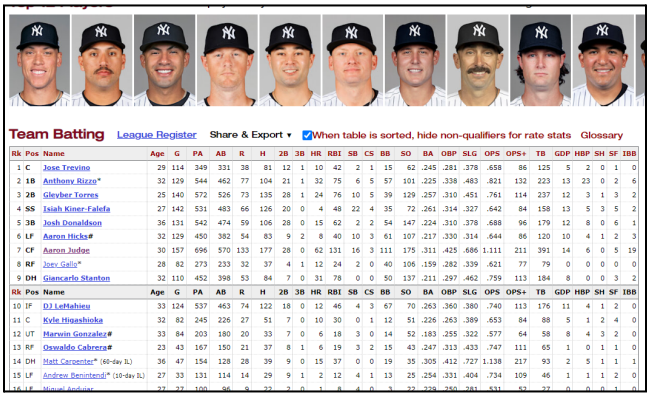
Sports teams would originally evaluate athletes based on gut instincts and arbitrary attributes such as fitness level, motivation, appearance, and body type. However, data scientists and statisticians have entered the sport with advanced measures over the previous twenty years. The 2002 Oakland Athletics were the first and most successful team to employ this strategy, taking a statistical approach to their season by using data analytics to draft, trade, and manage its players. Steinberg from [1] stated that the 2002 Oakland Athletics "sabermetrics system did not

require that players fit the height, weight, speed or body composition prototypes that dictated the moves of other clubs". This resulted in a high efficiency, low out system. Additionally the authors in [2] stated, "Early sabermetricians combined batters' and teams' on-base percentage and slugging percentage into a statistic called OPS (On base Plus Slugging.) However, even this measure doesn't fully describe how effective a team is at scoring runs, and has increasingly been replaced by sabermetric measures which weight on-base percentage and slugging percentage". As a result of this 2002 Oakland Athletics teams employing sabermetrics, other baseball teams and sports caught on to this new strategy.

Although our work is on a smaller scale than that of professional baseball teams, we looked at and analyzed similar relationships between baseball statistics. After performing some relatively basic analysis, we dove into deeper tests that pull from large datasets such as how rule changes affect baseball statistics over time.

METHODOLOGY/BACKGROUND

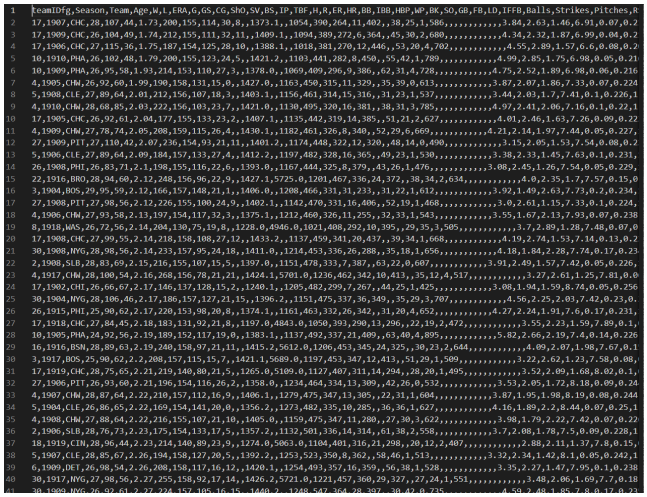
We used Baseball Reference and pyBaseball as our primary source of data. They contain a wealth of data dating back to as early as 1900, when the modern era of professional baseball began. Baseball reference contains seasonal data on teams and players. Pictured below is Baseball Reference's page on the 2022 New York Yankees. It displays their best players, classic baseball statistics as well as sabermetric statistics.



Team Batting League Register Share & Export ▾ ☑When table is sorted, hide non-qualifiers for rate stats Glossary

Rk	Pos	Name	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB
1	C	Jose Trevino	29	114	349	331	38	81	12	1	10	42	2	1	15	62	.245	.281	.378	.658	86	125	5	2	0	1	0
2	1B	Anthony Rizzo*	32	129	544	462	77	104	21	1	32	75	6	5	57	101	.225	.338	.483	.821	132	223	13	23	0	2	6
3	2B	Gleyber Torres	25	140	572	526	73	135	28	1	24	76	10	5	39	129	.257	.310	.451	.761	114	237	12	3	1	3	2
4	SS	Isiah Kiner-Falefa	27	142	531	483	66	126	20	0	4	48	22	4	35	72	.261	.314	.327	.642	84	158	13	5	3	5	2
5	3B	Josh Donaldson	36	131	542	474	59	106	28	0	15	62	2	2	54	147	.224	.310	.378	.688	96	179	12	8	0	6	1
6	LF	Aaron Hicks#	32	129	450	382	54	83	9	2	8	40	10	3	61	107	.217	.330	.314	.644	86	120	10	4	1	2	3
7	CF	Aaron Judge	30	157	696	570	133	177	28	0	62	131	16	3	111	175	.311	.425	.686	1.111	211	391	14	6	0	5	19
8	RF	Joey Gallo*	28	82	273	233	32	37	4	1	12	24	2	0	40	106	.159	.282	.339	.621	77	79	0	0	0	0	0
9	DH	Giancarlo Stanton	32	110	452	398	53	84	7	0	31	78	0	0	50	137	.211	.297	.462	.759	113	184	8	0	0	3	2
Rk	Pos	Name	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB
10	IF	DJ LeMahieu	33	124	537	463	74	122	18	0	12	46	4	3	67	70	.263	.360	.380	.740	113	176	11	4	1	2	0
11	C	Kyle Higashioka	32	82	245	226	27	51	7	0	10	30	0	1	12	51	.226	.263	.389	.653	84	88	5	1	2	4	0
12	UT	Marwin Gonzalez#	33	84	203	180	20	33	7	0	6	18	3	0	14	52	.183	.255	.322	.577	64	58	8	4	3	2	0
13	RF	Oswaldo Cabrera#	23	43	167	150	21	37	8	1	6	19	3	2	15	43	.247	.313	.433	.747	111	65	1	0	1	1	0
14	DH	Matt Carpenter* (60-day IL)	36	47	154	128	28	39	9	0	15	37	0	0	19	35	.305	.412	.727	1.138	217	93	2	5	1	1	1
15	LF	Andrew Benintendi* (10-day IL)	27	33	131	114	14	29	9	1	2	12	4	1	13	25	.254	.331	.404	.734	109	46	1	1	1	2	0
16	LF	Miguel Andujar	27	37	100	86	9	22	2	0	1	8	4	0	3	22	.229	.250	.281	.531	52	27	0	0	0	1	0

In addition we used pybaseball, which is a Python library for analyzing baseball data. The library pulls data from Baseball Reference, Baseball Savant, and FanGraphs, official websites that store tremendous amounts of information as previously discussed. The software gets statcast data, pitching statistics, hitting statistics, division standings/team records, awards data, and other information. Data is accessible at the individual pitching/batting level as well as aggregated at the season and time period level.



```
1  teamIDfg,Season,Team,Age,W,L,ERA,G,GS,CG,ShO,SV,BS,IP,TBF,H,R,ER,HR,BB,IBB,HBP,WP,BK,SO,GB,FB,LD,IFFB,Balls,Strikes,Pitches,R
2  17,1907,CHC,28,107,44,1.73,200,155,114,30,8,,1373.1,,1054,390,264,11,402,,38,25,1,586,,,,,,,,,,3.84,2.63,1.46,6.91,0.07,0.2
3  17,1909,CHC,26,104,49,1.74,212,155,111,32,11,,1409.1,,1094,389,272,6,364,,45,30,2,680,,,,,,,,,,4.34,2.32,1.87,6.99,0.04,0.2
4  17,1906,CHC,27,115,36,1.75,187,154,125,28,10,,1388.1,,1018,381,270,12,446,,53,20,4,702,,,,,,,,,,4.55,2.89,1.57,6.6,0.08,0.2
5  10,1910,PHA,26,102,48,1.79,200,155,123,24,5,,1421.2,,1103,441,282,8,450,,55,42,1,789,,,,,,,,,,4.99,2.85,1.75,6.98,0.05,0.21
6  10,1909,PHA,26,95,58,1.93,214,153,110,27,3,,1378.0,,1069,409,296,9,386,,62,31,4,728,,,,,,,,,,4.75,2.52,1.89,6.98,0.06,0.216
7  4,1905,CHW,26,92,60,1.99,190,158,131,15,8,,1427.0,,1163,450,315,11,329,,35,39,0,613,,,,,,,,,,3.87,2.07,1.86,7.33,0.07,0.224
8  5,1908,CLE,27,89,64,2.01,212,156,107,18,3,,1403.1,,1156,461,314,15,316,,31,21,1,537,,,,,,,,,,3.44,2.03,1.7,7.41,0.1,0.226,1
9  4,1910,CHW,28,68,85,2.03,222,156,103,23,7,,1421.0,,1130,495,320,16,381,,38,31,3,785,,,,,,,,,,4.97,2.41,2.06,7.16,0.1,0.22,1
10 17,1905,CHC,26,92,61,2.04,177,155,133,23,2,,1407.1,,1135,442,319,14,385,,51,21,2,627,,,,,,,,,,4.01,2.46,1.63,7.26,0.09,0.22
11 4,1909,CLE,27,78,74,2.05,208,159,115,26,4,,1430.1,,1182,461,326,8,340,,52,29,6,669,,,,,,,,,,4.21,2.14,1.97,7.44,0.05,0.227,
12 27,1909,PIT,27,110,42,2.07,236,154,93,21,11,,1401.2,,1174,448,322,12,320,,48,14,0,490,,,,,,,,,,3.15,2.05,1.53,7.54,0.08,0.2
13 5,1906,CLE,27,89,64,2.09,184,157,133,27,4,,1412.2,,1197,482,328,16,365,,49,23,1,530,,,,,,,,,,3.38,2.33,1.45,7.63,0.1,0.231,
14 26,1908,PHI,26,83,71,2.1,198,155,116,22,6,,1393.0,,1167,444,325,8,379,,43,26,1,476,,,,,,,,,,3.08,2.45,1.26,7.54,0.05,0.229,
15 22,1916,BRO,28,94,60,2.12,248,156,96,22,9,,1427.1,,1275,467,336,24,372,,38,34,2,634,,,,,,,,,,4.0,2.35,1.7,7.57,0.15,0
16 3,1904,BOS,29,95,59,2.12,166,157,148,21,1,,1406.0,,1208,466,331,31,233,,31,22,1,612,,,,,,,,,,3.92,1.49,2.63,7.73,0.2,0.234,
17 27,1908,PIT,27,98,56,2.12,226,155,100,24,9,,1402.1,,1142,470,331,16,406,,52,19,1,468,,,,,,,,,,3.0,2.61,1.15,7.33,0.1,0.224,
18 4,1906,CHW,27,93,58,2.13,197,154,117,32,3,,1375.1,,1212,460,326,11,255,,32,33,1,543,,,,,,,,,,3.55,1.67,2.13,7.93,0.07,0.238
19 8,1918,WAS,26,72,56,2.14,204,130,75,19,8,,1228.0,4946.0,1021,408,292,10,395,,29,35,3,505,,,,,,,,,,3.7,2.89,1.28,7.48,0.07,0
20 17,1908,CHC,27,99,55,2.14,218,158,108,27,12,,1433.2,,1137,459,341,20,437,,39,34,1,668,,,,,,,,,,4.19,2.74,1.53,7.14,0.13,0.2
21 30,1908,NYG,28,98,56,2.14,233,157,95,24,18,,1411.0,,1214,453,336,26,288,,35,18,1,656,,,,,,,,,,4.18,1.84,2.28,7.74,0.17,0.23
22 2,1908,SLB,28,83,69,2.15,216,155,107,15,5,,1392.0,,1151,478,333,7,387,,63,22,0,607,,,,,,,,,,3.91,2.49,1.57,7.42,0.05,0.226,
23 4,1917,CHW,28,100,54,2.16,268,156,78,21,21,,1424.1,5701.0,1236,462,342,10,413,,35,12,4,517,,,,,,,,,,3.27,2.61,1.25,7.81,0.0
24 17,1902,CHI,26,66,67,2.17,146,137,128,15,2,,1240.1,,1205,482,299,7,267,,44,25,1,425,,,,,,,,,,3.08,1.94,1.59,8.74,0.05,0.256
25 30,1904,NYG,28,106,46,2.17,186,157,127,21,15,,1396.2,,1151,475,337,36,349,,35,29,3,707,,,,,,,,,,4.56,2.25,2.03,7.42,0.23,0.
26 26,1915,PHI,25,90,62,2.17,220,153,98,20,8,,1374.1,,1161,463,332,26,342,,31,20,4,652,,,,,,,,,,4.27,2.24,1.91,7.6,0.17,0.231,
27 17,1918,CHC,27,84,45,2.18,183,131,92,21,8,,1197.0,4843.0,1050,393,290,13,296,,22,19,2,472,,,,,,,,,,3.55,2.23,1.59,7.89,0.1,
28 10,1905,PHA,24,92,56,2.19,189,152,117,19,0,,1383.1,,1137,492,337,21,409,,63,40,4,895,,,,,,,,,,5.82,2.66,2.19,7.4,0.14,0.226
29 16,1916,BSN,28,89,63,2.19,240,158,97,21,11,,1415.2,5612.0,1206,453,345,24,335,,30,23,2,664,,,,,,,,,,4.09,2.07,1.98,7.67,0.1,
30 3,1917,BOS,25,90,62,2.2,208,157,115,15,7,,1421.1,5689.0,1197,453,347,12,413,,51,29,1,500,,,,,,,,,,3.22,2.62,1.23,7.58,0.08,
31 17,1919,CHC,28,75,65,2.21,219,140,80,21,5,,1265.0,5109.0,1127,407,311,14,294,,28,20,1,495,,,,,,,,,,3.52,2.09,1.68,8.02,0.1,
32 27,1906,PIT,26,93,60,2.21,196,154,116,26,2,,1358.0,,1234,464,334,13,309,,42,26,0,532,,,,,,,,,,3.53,2.05,1.72,8.18,0.09,0.24
33 4,1907,CHW,28,87,64,2.22,210,157,112,16,9,,1406.1,,1279,475,347,13,305,,22,31,1,604,,,,,,,,,,3.87,1.95,1.98,8.19,0.08,0.244
34 5,1904,CLE,26,86,65,2.22,169,154,141,20,0,,1356.2,,1273,482,335,10,285,,36,36,1,627,,,,,,,,,,4.16,1.89,2.2,8.44,0.07,0.25,1
35 4,1908,CHW,27,88,64,2.22,216,155,107,21,10,,1405.0,,1159,475,347,11,280,,27,30,3,622,,,,,,,,,,3.98,1.79,2.22,7.42,0.07,0.22
36 2,1906,SLB,28,76,73,2.23,175,154,133,17,5,,1352.2,,1132,501,336,14,314,,61,38,2,598,,,,,,,,,,3.7,2.08,1.78,7.5,0.09,0.226,1
37 18,1919,CIN,28,96,44,2.23,214,140,89,23,9,,1274.0,5063.0,1104,401,316,21,298,,20,12,2,487,,,,,,,,,,2.88,2.11,1.37,7.8,0.15,
38 5,1907,CLE,28,85,67,2.26,194,158,127,20,5,,1392.2,,1253,523,350,8,362,,58,46,1,513,,,,,,,,,,3.32,2.34,1.42,8.1,0.05,0.242,1
39 6,1909,DET,26,98,54,2.26,208,158,117,16,12,,1420.1,,1254,493,357,16,359,,56,38,1,528,,,,,,,,,,3.35,2.27,1.47,7.95,0.1,0.238
40 30,1917,NYG,27,98,56,2.27,255,158,92,17,14,,1426.2,5721.0,1221,457,360,29,327,,27,24,1,551,,,,,,,,,,3.48,2.06,1.69,7.7,0.18
41 30,1903,NYG,26,92,61,2.27,224,157,105,16,15,,1440.2,,1248,542,364,28,297,,30,47,0,735,,,,,,,,,,4.50,2.48,1.85,7.8,0.17,0.21
```

Python is an incredible tool for analyzing and compiling tests on data and having these tools allowed us to access the database in a clean, direct manner.

METHODOLOGY

Our main tasks included initially developing several key functions that mine data from pybaseball. These include: pitching statistical comparisons, graphical standings by season and a least squares residual calculator.

Following the initial work, we were tasked to mine data on a larger scale. We addressed our major topic being the general null hypothesis: various events and/or rule changes have no impact on the overall statistical trends throughout Major League Baseball. The experiments included these rule changes and events:

In 1968: MLB lowered the height of the pitching mound from 15 in to 10 inches.

Mid 1970s: Various MLB teams moved to larger stadiums with longer distances from home plate to the outfield fence.

In 2003: MLB began testing players for steroids.

After graphing the data between 1900-2022, it was necessary to begin data cleaning. We didn't want extraneous factors impacting our results. As such, we decided to clean the data to only include the year of the rule change plus or minus 5-7 years, depending on the experiment. We then analyzed the null hypothesis through a difference of means t-test. Our decision on whether to reject or to fail to reject the null hypothesis comes down to if our t-statistic is inside or outside the t-critical value bounds. From there, we looked at the key findings and implications of our results.

PRELIMINARY WORK

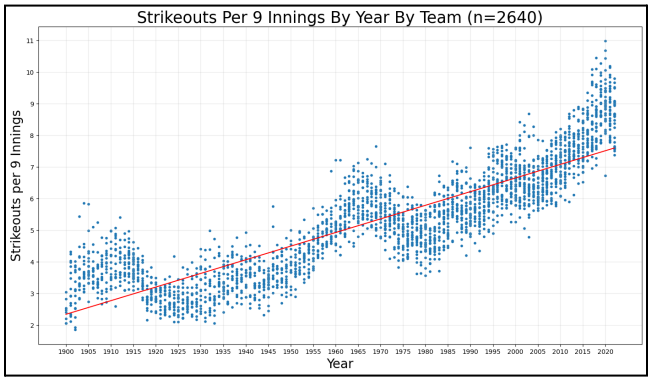
For our initial setup, we drew up code in Python to analyze statistics across a broad range (1900-2022). We wanted to analyze specific statistics that we thought could be impacted by the rule changes outlined in our methodology. As such, we plotted and analyzed K/9 (strikeouts per nine innings), SB (stolen bases), and HR (home runs). Each data point was averaged by team per year. For example, the 2022 Texas Rangers stole 128 bases as a team. That data is represented as a single point. The summary data collected for each team and each year are as follows.

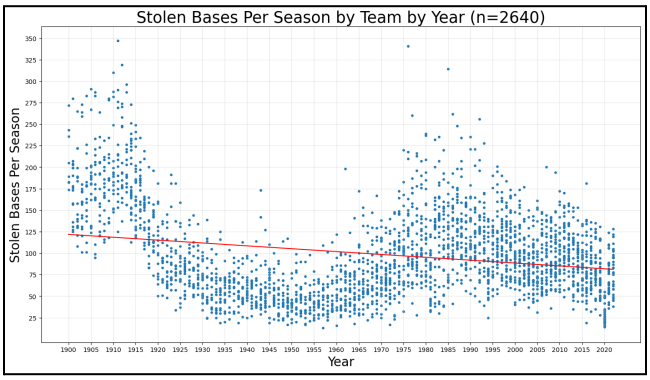
K/9 (1900-2022)	
n	2640
mean	5.3500
standard deviation	1.7571
min	1.8500
max	10.9800
variance	3.0873

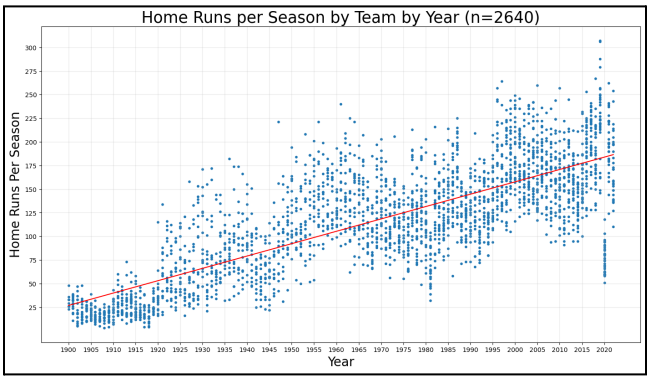
SB (1900-2022)	
n	2640
mean	98.5341
standard deviation	51.3285
min	13
max	347
variance	2634.6241

HR (1900-2022)	
n	2640
mean	118.3057
standard deviation	59.4055
min	3
max	307
variance	3529.0081

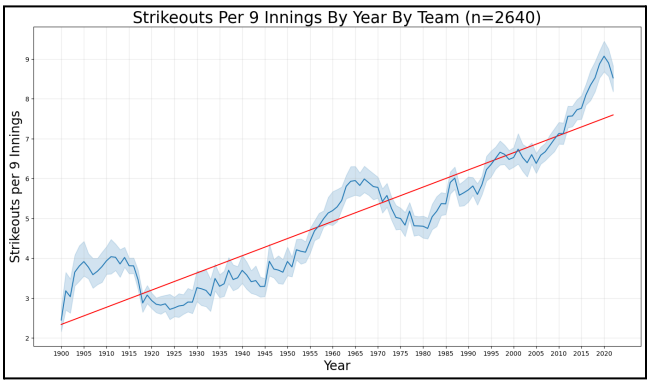
The scatterplots of each statistic vs year are as follows. Blue dots represent a single data point, the red line represents the best fit line of the entire dataset.

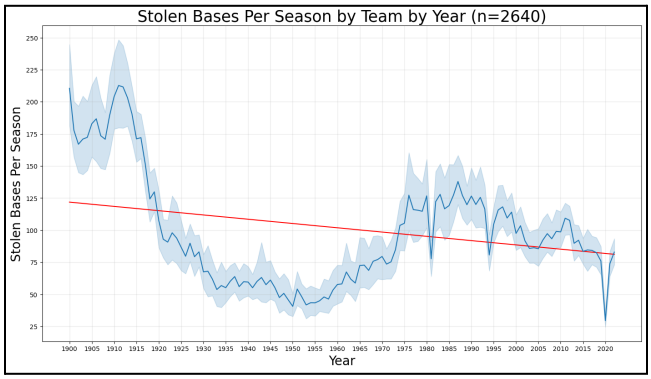


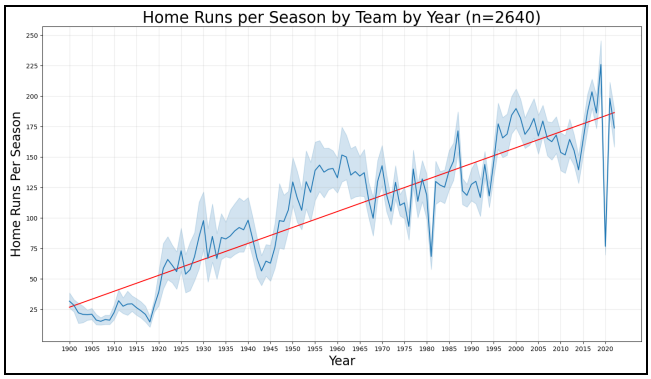




For visual purposes, we wanted to clean the data and contain the line plot to a 95% confidence level. Below are the plots for each statistic. The dark blue line represents the yearly mean, the shaded blue area represents the 95% confidence interval of all data points for each year.







Now that our preliminary data was collected and displayed, it was time to move on to the evaluation part of our three experiments.

EVALUATION

With the established framework for the project, we move on to the major topic. We formed the general null hypothesis: various events and/or rule changes have no impact on the overall statistical trends throughout Major League Baseball. The experiments tested are the following rule changes: MLB lowered the height of the pitching mound from 15 in to 10 in in 1968, various MLB teams moved to larger stadiums with longer distances from home plate to the outfield fence in the 1970s and MLB began testing players for steroids in 2003. Diving into the rule changes, we start with MLB lowering the height of the pitching mound.

Experiment 1

Null hypothesis: MLB lowering the pitching mound in 1968 from 15 inches to 10 inches had no impact on team strikeouts per 9 innings ratio.

Null hypothesis: $\mu_1 - \mu_2 = 0$

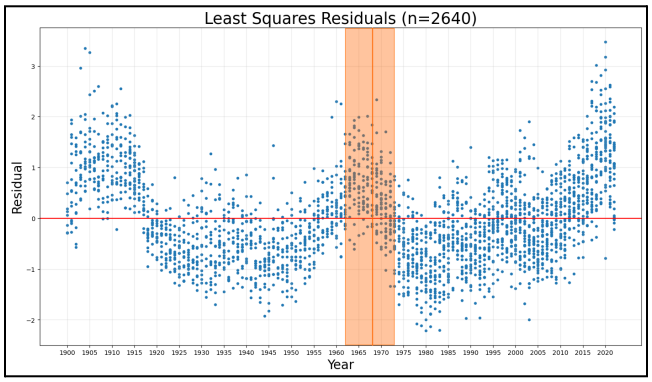
Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

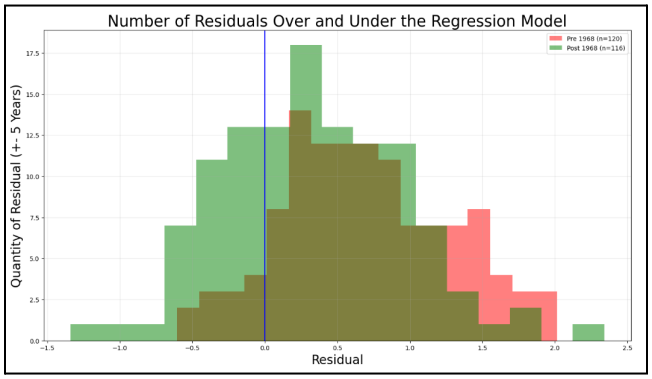
As written above, we used pyBaseball to extract team pitching data from 1900 to 2022. We then mined the K/9 column, averaged each team by year, and plotted them by year. However, we wanted to clean the data further by only analyzing the years surrounding 1968, when the rule was put into effect. This way, no extraneous variables in the game would impact our results. The summary data collected for each team and each year are as follows.

K/9 (1962-1967)	
n	120
mean	5.8236
standard deviation	0.5961
min	4.4900
max	7.2400
variance	0.3553

K/9 (1968-1973)	
n	116
mean	5.6853
standard deviation	0.6026
min	4.100
max	7.6500
variance	0.3631

We then graphed the least squares residuals for this rule change. The rule went into effect in 1968, so we wanted to look at what percentage of residuals were over/under the lifelong trendline, which is highlighted in orange on the histogram. Each histogram was separated into years before and after the rule went into effect.





Of the 120 residual points between 1962-1967 (red), 90.00% of them were at or above the lifelong trendline. Of the 116 residual points between 1968-1973 (green), 68.96% of them were at or above the lifelong trendline.

To test our null hypothesis, we used a difference of means hypothesis test (two tailed, α<.05). This decision comes down to the t-statistic value. If it is outside the critical values from the t-table below, we reject the null. If it is inside, we fail to reject the null hypothesis. Using the equation: t = $[(\overline{x_1} - \overline{x_2})/SE$, we found the test statistic to be t = 1.7723. Using a T-table, $P(t \leq 1.7723) = 0.96118$. Therefore, $P(t > 1.7723) = 0.0388$. The two tailed P value equals 0.776. Since $\alpha = .05$, this P value is not considered to

be significantly significant, and we fail to reject the null hypothesis.

Experiment 2

Null hypothesis: Several baseball teams moving into stadiums with larger fields in the 70s had no impact on team successful stolen bases.

Null hypothesis: $\mu_1 - \mu_2 = 0$

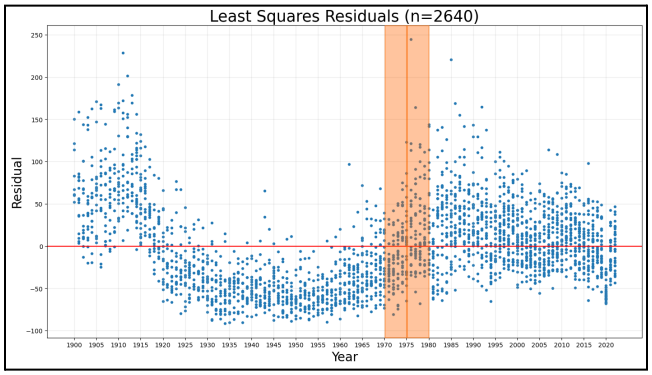
Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

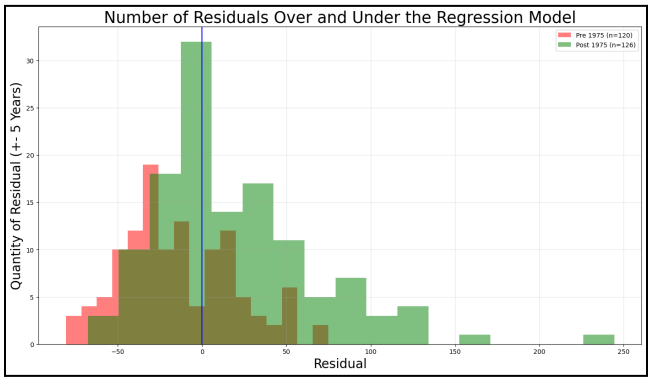
As written above, we used pyBaseball to extract team batting data from 1900 to 2022. We then mined the SB column, averaged each team by year, and plotted by year. However, we wanted to clean the data further by only analyzing the years surrounding 1975. Unlike before, when there was a rule put into effect in a single year, teams moving into larger stadiums happened over the course of several years. We decided 1975 would be a good year to separate and clean our data. This was after the Phillies, Reds and Pirates switched stadiums. This way, no extraneous variables in the game would impact our results. The summary data collected for each team and each year are as follows.

SB (1970-1974)	
n	120
mean	83.3583
standard deviation	33.0716
min	17.0000
max	172.0000
variance	1093.7277

SB (1975-1979)	
n	126
mean	115.7540
standard deviation	48.6203
min	28.0000
max	341.000
variance	2363.9310

We then graphed the least squares residuals. The stadium change happened roughly in 1975, so we wanted to look at what percentage of residuals were over/under the lifelong trendline, which is highlighted in orange on the histogram. Each histogram was separated into years before and after 1975.





Of the 120 residual points between 1970-1974 (red), 33.33% of them were at or above the lifelong trendline. Of the 126 residual points between

1975-1979 (green), 61.91% of them were at or above the lifelong trendline.

To test our null hypothesis, we used a difference of means hypothesis test (two tailed, α<.05). This decision comes down to the t-statistic value. If it is outside the critical values from the t-table below, we reject the null. If it is inside, we fail to reject the null hypothesis. Using the equation: t = $[(\overline{x_1} - \overline{x_2})/SE$, we found the test statistic to be t = -6.1358. Using a T-table, P(t≤$-6.1358) \approx 0$. Therefore, P(t $>-6.1358) = 99.999$. The two tailed P value is <0.0001. Since $\alpha = .05$, this P value is considered to be significantly significant, and we reject the null hypothesis.

Experiment 3

Null hypothesis: MLB testing players for steroids in 2003 onward had no impact on team home runs throughout a season

Null hypothesis: $\mu_1 - \mu_2 = 0$

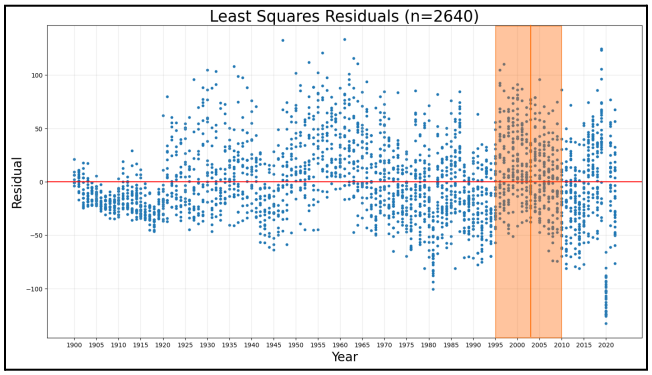
Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

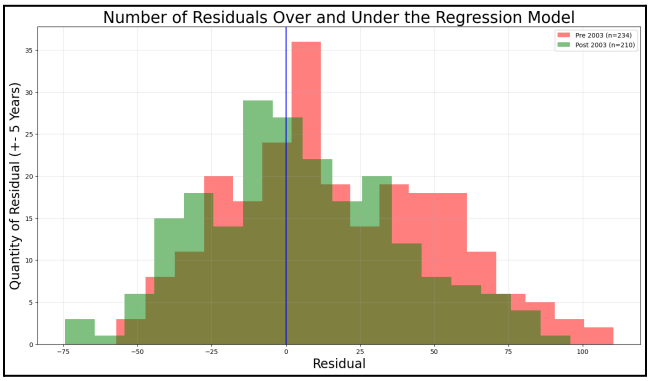
As written above, we used pyBaseball to extract team batting data from 1900 to 2022. We then mined the HR column, averaged each team by year, and plotted them by year. However, we wanted to clean the data further by only analyzing the years surrounding 2003, when the rule was put into effect. This way, no extraneous variables in the game would impact our results. The summary data collected for each team and each year are as follows.

HR (1995-2002)	
n	234
mean	173.0128
standard deviation	35.3272
min	94.0000
max	264.0000
variance	1248.0127

HR (2003-2009)	
n	210
mean	171.1333
standard deviation	32.7115
min	94.0000
max	260.0000
variance	1070.0396

We then graphed the least squares residuals for this rule change. The rule went into effect in 2003, so we wanted to look at what percentage of residuals were over/under the lifelong trendline, which is highlighted in orange on the histogram. Each histogram was separated into years before and after the rule went into effect.





Of the 234 residual points between 1995-2002 (red), 66.24% of them were at or above the lifelong trendline. Of the 210 residual points between 2003-2009 (green), 54.76% of them were at or above the lifelong trendline.

To test our null hypothesis, we used a difference of means hypothesis test (two tailed, α<.05). This decision comes down to the t-statistic value. If it is outside the critical values from the t-table below, we reject the null. If it is inside, we fail to reject the null hypothesis. Using the equation: t = $[(\overline{x_1} - \overline{x_2})/SE$, we found the test statistic to be t = 0.5820 Using a T-table, P(t≤ $0.5820) = 0.71957$. The two tailed P value equals 0.5625. Since $\alpha = .05$, this P value is not considered to be significantly significant, and we fail to reject the null hypothesis.

DISCUSSION

One critical challenge we encountered when analyzing sabermetrics is that many have not been recorded for more than seven years. The 'Statcast-Era' began in 2015 when Major League Baseball began tracking several advanced analytics. These include pitching velocities, baserunning speeds, fielding percentages and much more. This made it difficult to conduct a proper analysis of the data when there is little to begin with.

Additionally, pybaseball limits the amount of pull requests by source for a certain time period. This posed a minor problem for making rapid adjustments or quick changes. Often, we had to wait for our time period to reset and we could pull more data. It may be helpful to store the data locally and then run tests. This can only work if the dataset is small enough to be reasonably stored locally.

CONCLUSION

The three experiments had the following results:

Experiment 1:

Fail to reject the null hypothesis that MLB lowering the pitching mound in 1968 from 15 inches to 10 inches had no impact on team strikeouts per 9 innings ratio.

Experiment 2:

Reject the null hypothesis that several baseball teams moving into stadiums with larger fields in the 70s had no impact on team successful stolen bases.

Experiment 3:

Fail to reject null hypothesis that MLB testing players for steroids in 2003 onward had no impact on team home runs throughout a season

In Experiment 1's case, there are several reasons why we failed to reject the null hypothesis. Our leading theory is that strikeouts are not an absolute performance measure when it comes to pitching. Pitchers can have very good strikeout rates and still perform poorly by giving up home runs and hits. There are a variety of other possible statistics that could have resulted in a rejection of the null hypothesis. In a future experiment, we would like to look at other sabermetrics, such as BABIP (batted balls in play) or OBA (opponent batting average). These metrics could offer a more absolute indication of pitching performance and result in a rejection of the null hypothesis. However, experimenting on these metrics would require an alteration in the experiment itself, as these metrics haven't been recorded as far back as 1900.

In Experiment 2's case, it remains fairly clear through both the graphs and analysis that there was a significant difference in the means. However, we remained curious about these results. It's possible that teams moving into larger stadiums with bigger fields were not the direct cause of the increase in stolen bases. We decided to do some follow up research on if this was partly the case. According to John McMurray at SABR, it appears that it was. "With

the opening of several large stadiums in the early 1970s…it is easy to see why teams emphasized speed over home run hitting power. Further, the attraction of scoring runners from second base on singles hit on Astroturf surely influenced teams to place an additional premium on fast players." [3]. The switch from grass to astroturf was not something we had considered in our research, and we feel confident saying that it could have been another significant factor.

For Experiment 3, we failed to reject the null hypothesis and have a leading theory as to why. The steroid era was a time filled with infamous players playing with performance enhancing drugs. However, the number of people taking these PEDs remained relatively low. We believe that when steroid testing began in 2003, only a handful of players were impacted. On a team based scale, like the one we measured, there wasn't a significant decrease in home runs because most people weren't on PEDs. However, it will be impossible to prove this as the number of players who were on PEDs has yet to be proven. All we can rely on is a speculating theory. If a statistic like home runs per player existed in the pyBaseball library, it would certainly be a focus for a future experiment.

Baseball remains unique in the fact that it has one of the most documented histories in sports. Through the past 122 years, almost every player and team has recorded statistics. Trends in these statistics have fluctuated and been influenced by many events and rule changes as the sport evolves. Often, these rule changes are applied to have a direct impact on trends of certain statistics. However, as we have proven in our experiments, it's not always the case that a statistic is impacted. This will have to be something MLB takes into consideration as they continue to modify and change the sport.

APPENDIX

Division of Work:

Python Setup - Nathan
Project Proposal Report - Nathan
Project Proposal Slides - Nathan
Coding Experiments - Nathan
Project Checkpoint Report - Nathan/Thomas
Project Checkpoint Slides - Nathan/Thomas
Coding Evaluation and Plots - Nathan
Project Final Slides - Nathan/Thomas
Project Final Report - Nathan/Thomas
Code Architecture Submission - Nathan

On my honor, as a University of Colorado Boulder student, I have neither given nor received unauthorized assistance.

See images below for full sized renders.

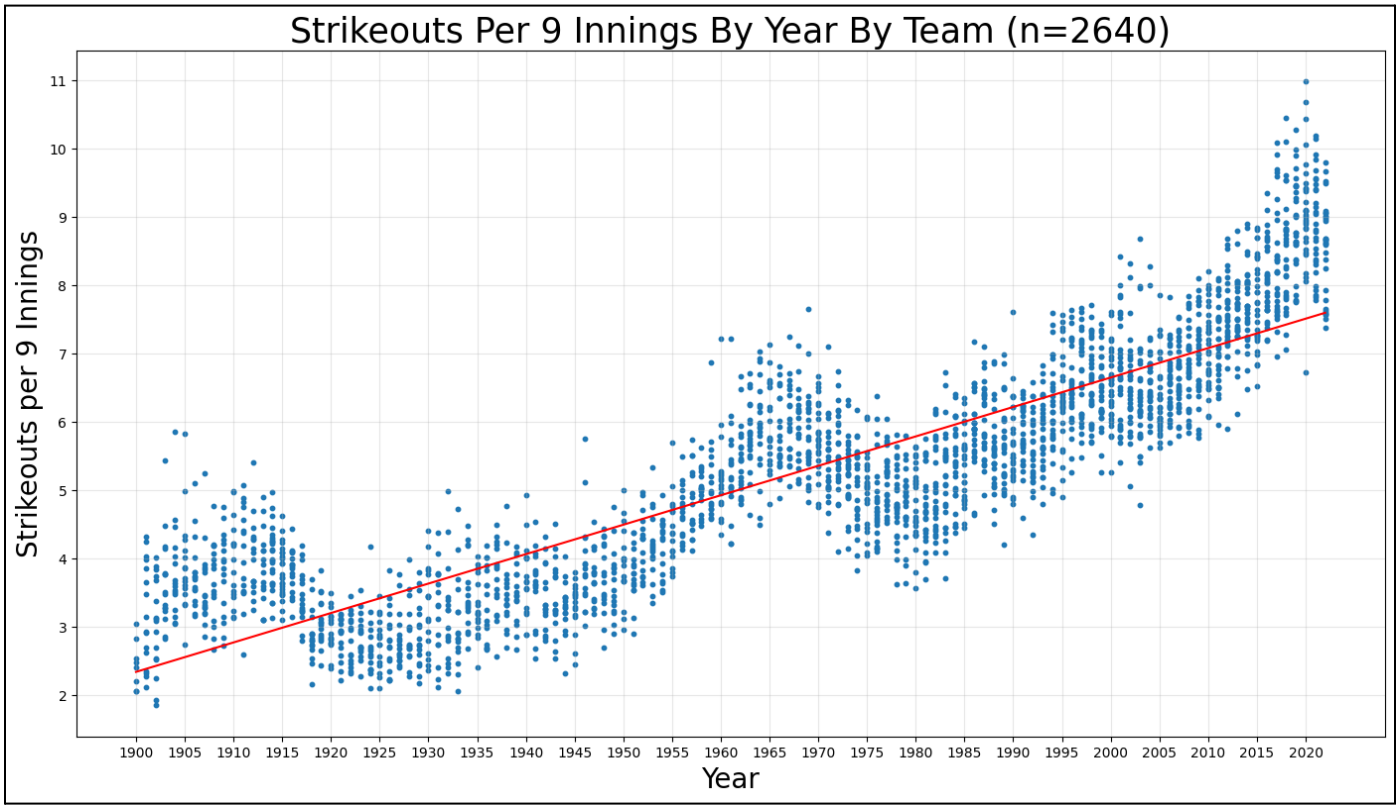
REFERENCES

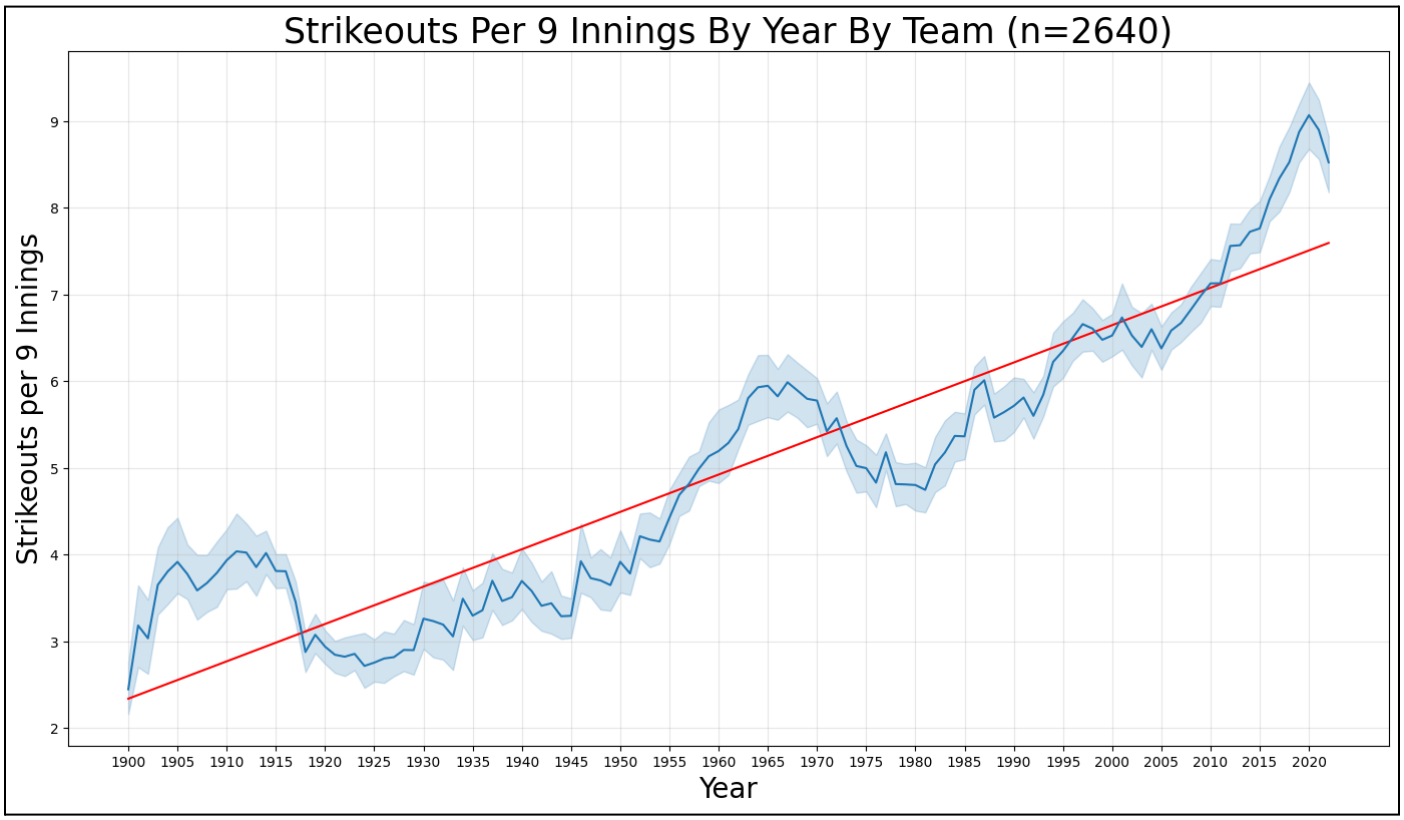
[1] Steinberg, Leigh. "Changing the Game: The Rise of Sports Analytics." Forbes, Forbes Magazine, 18 Aug. 2015,

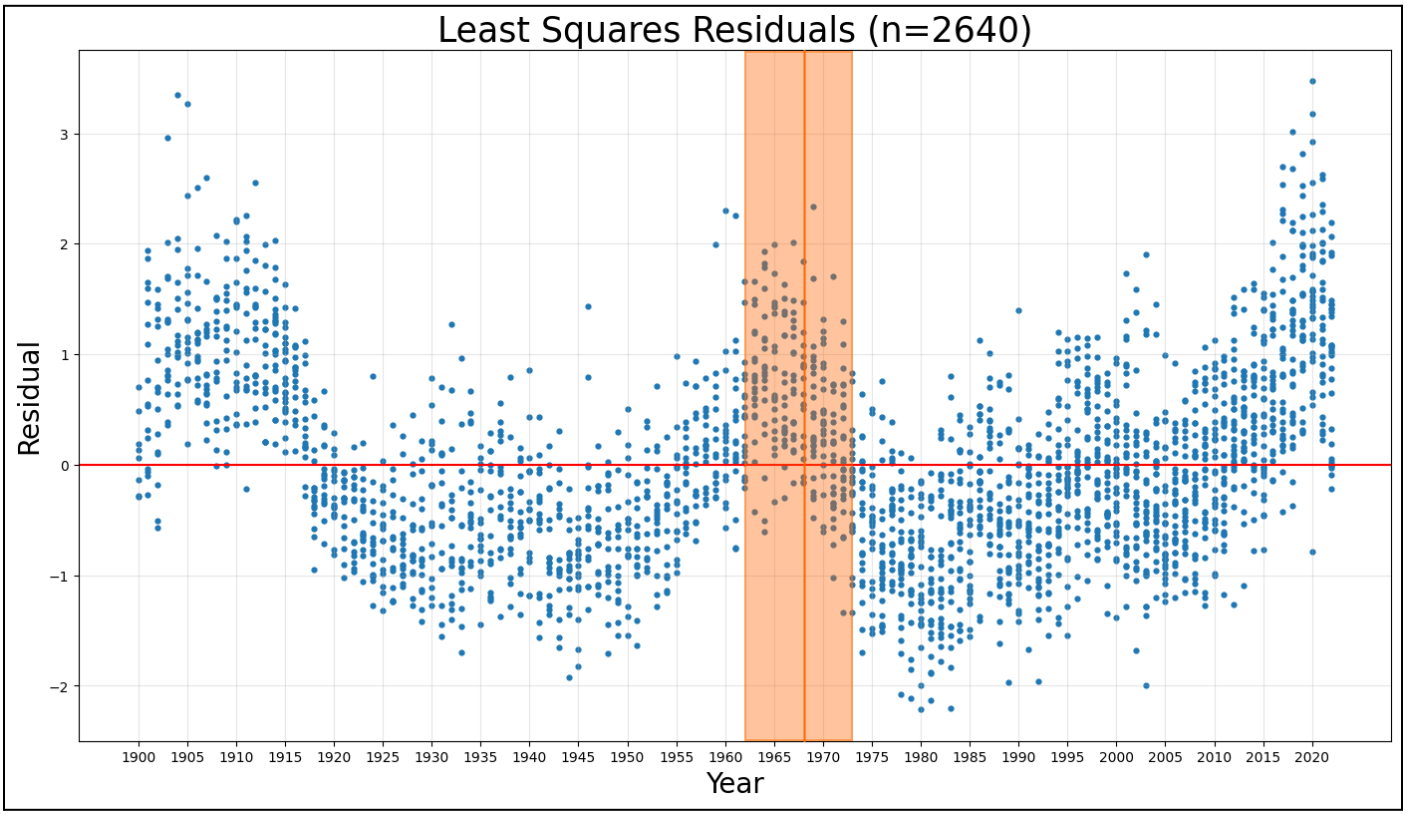
[2] Beneventano, Philip, Paul D. Berger, and Bruce D. Weinberg. "Predicting run production and run prevention in baseball: the impact of Sabermetrics." and*Int J Bus Humanit Technol* 2.4 (2012): 67-75.

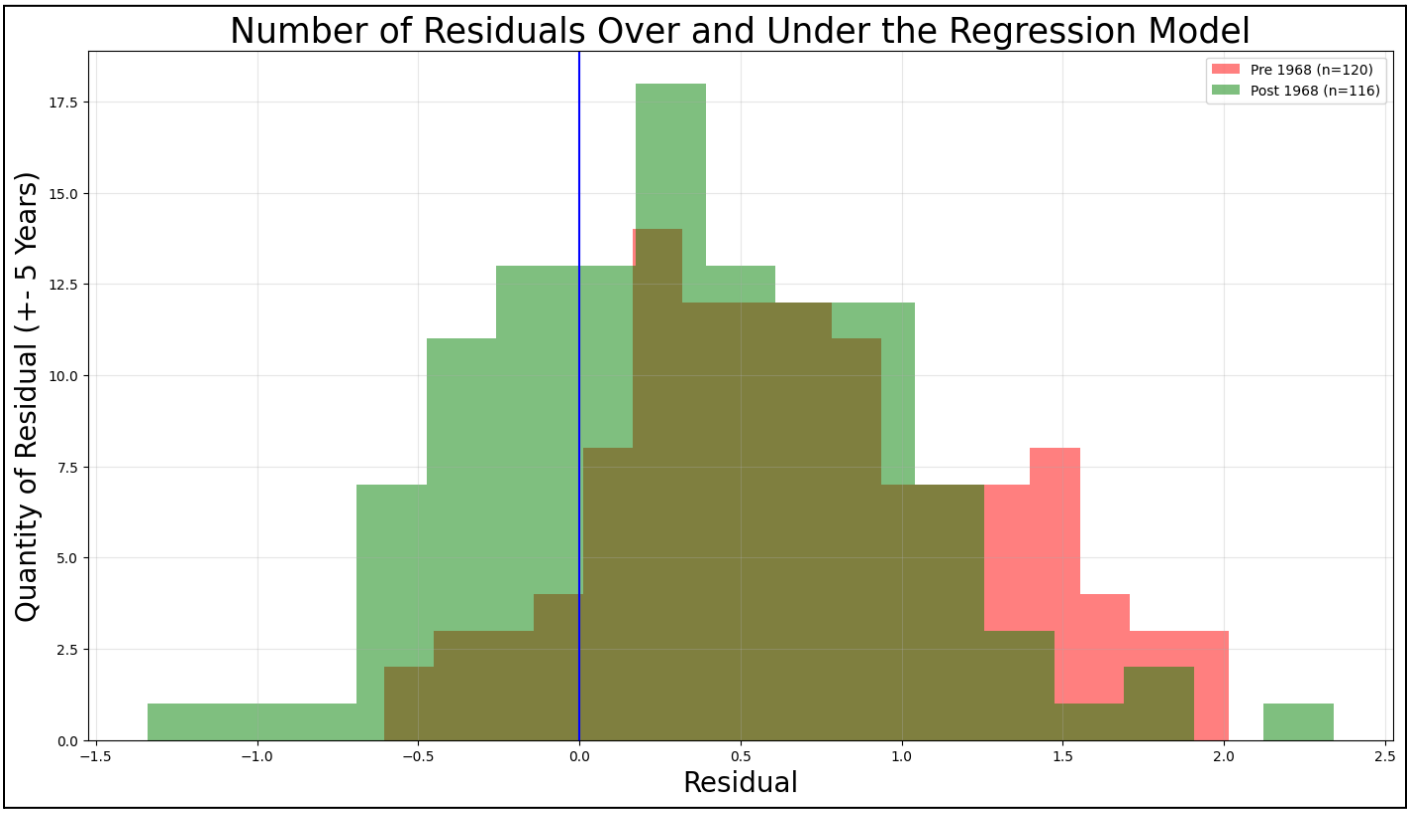
[3] McMurray, John. "Examining Stolen Base Trends by Decade from the Deadball Era through the 1970s." Society for American Baseball Research, Society for American Baseball Research, Oct. 2015, https://sabr.org/journal/article/examining-stolen-base-trends-by-decade-from-the-deadball-era-through-the-1970s/.

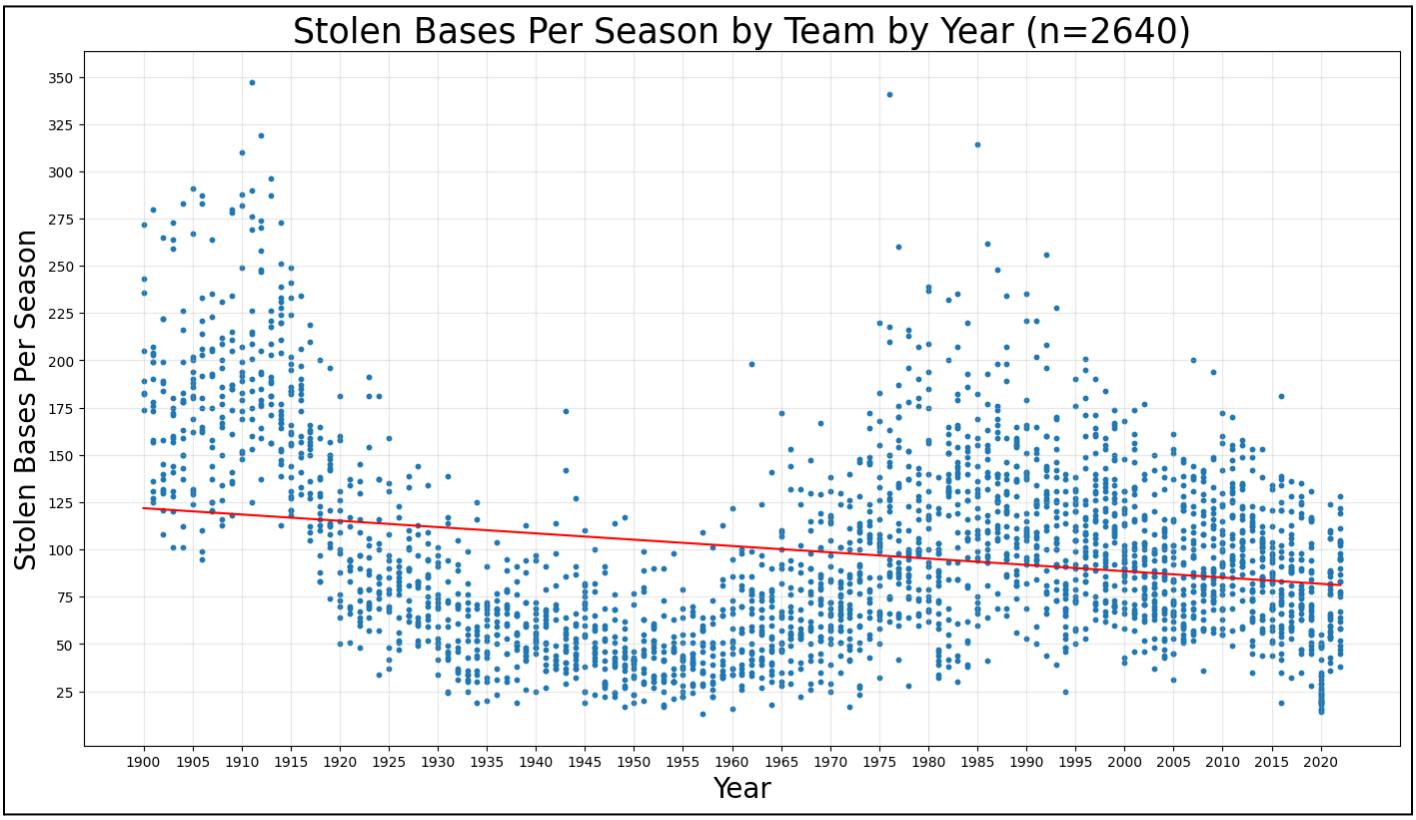
FULL SIZE PLOT IMAGES

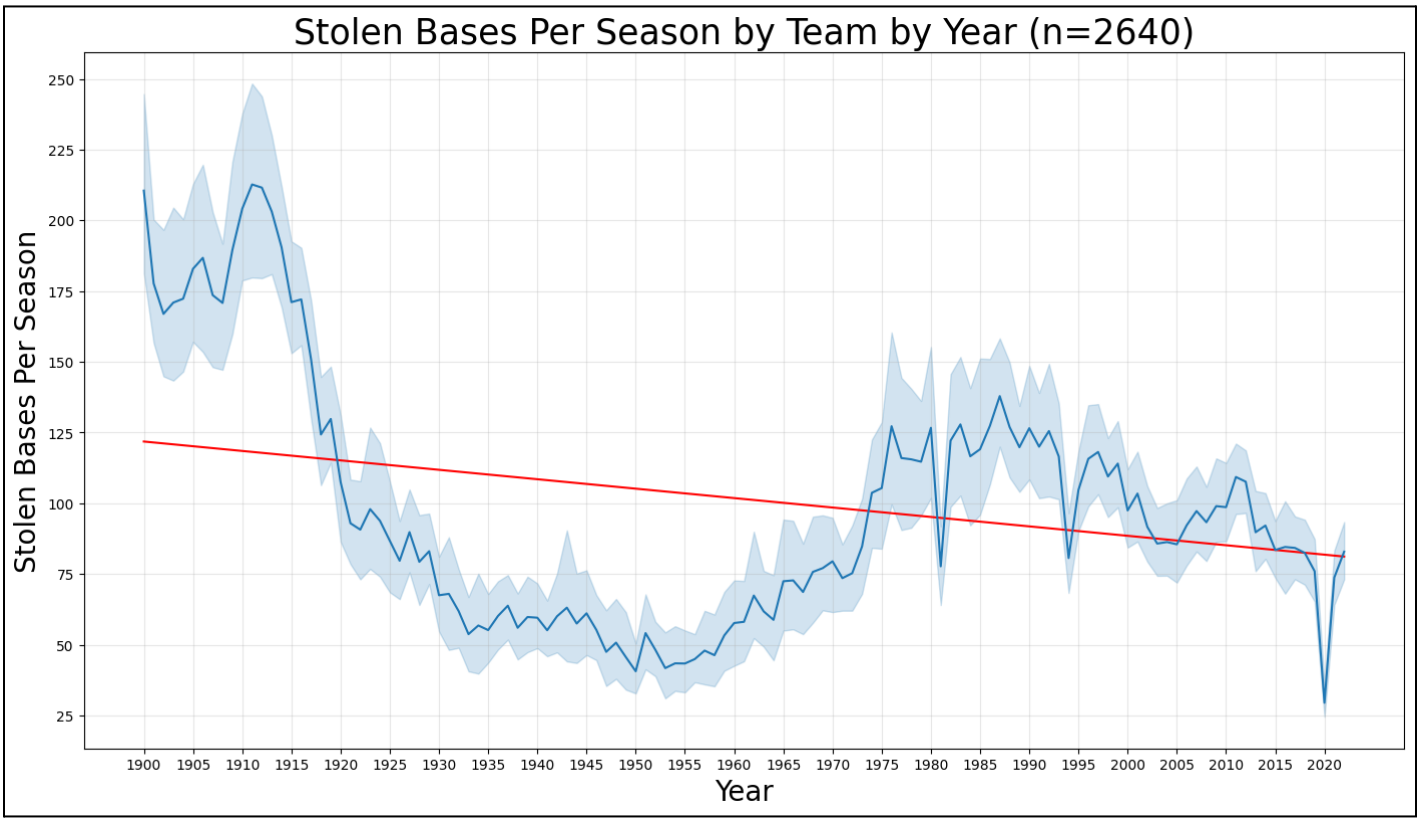


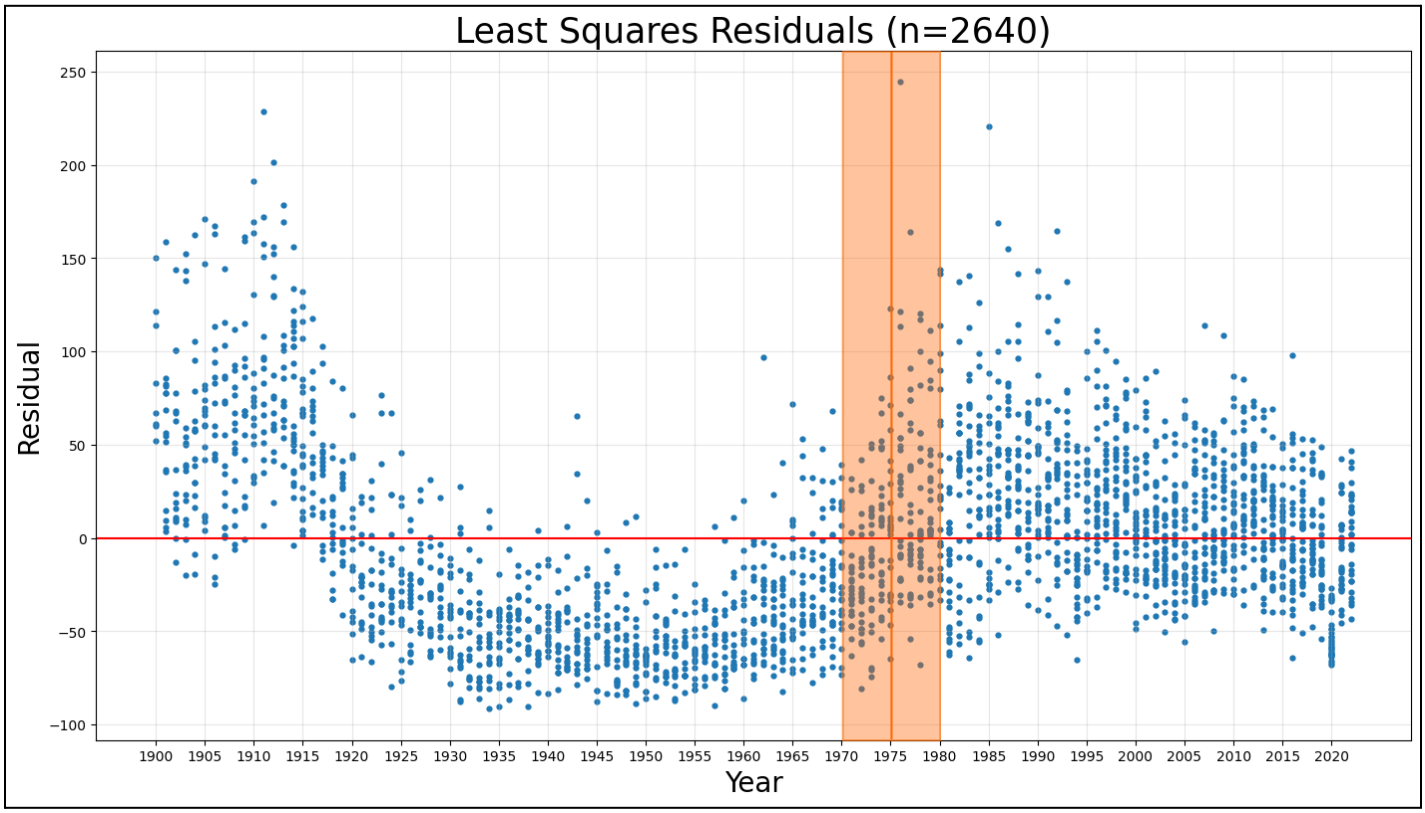


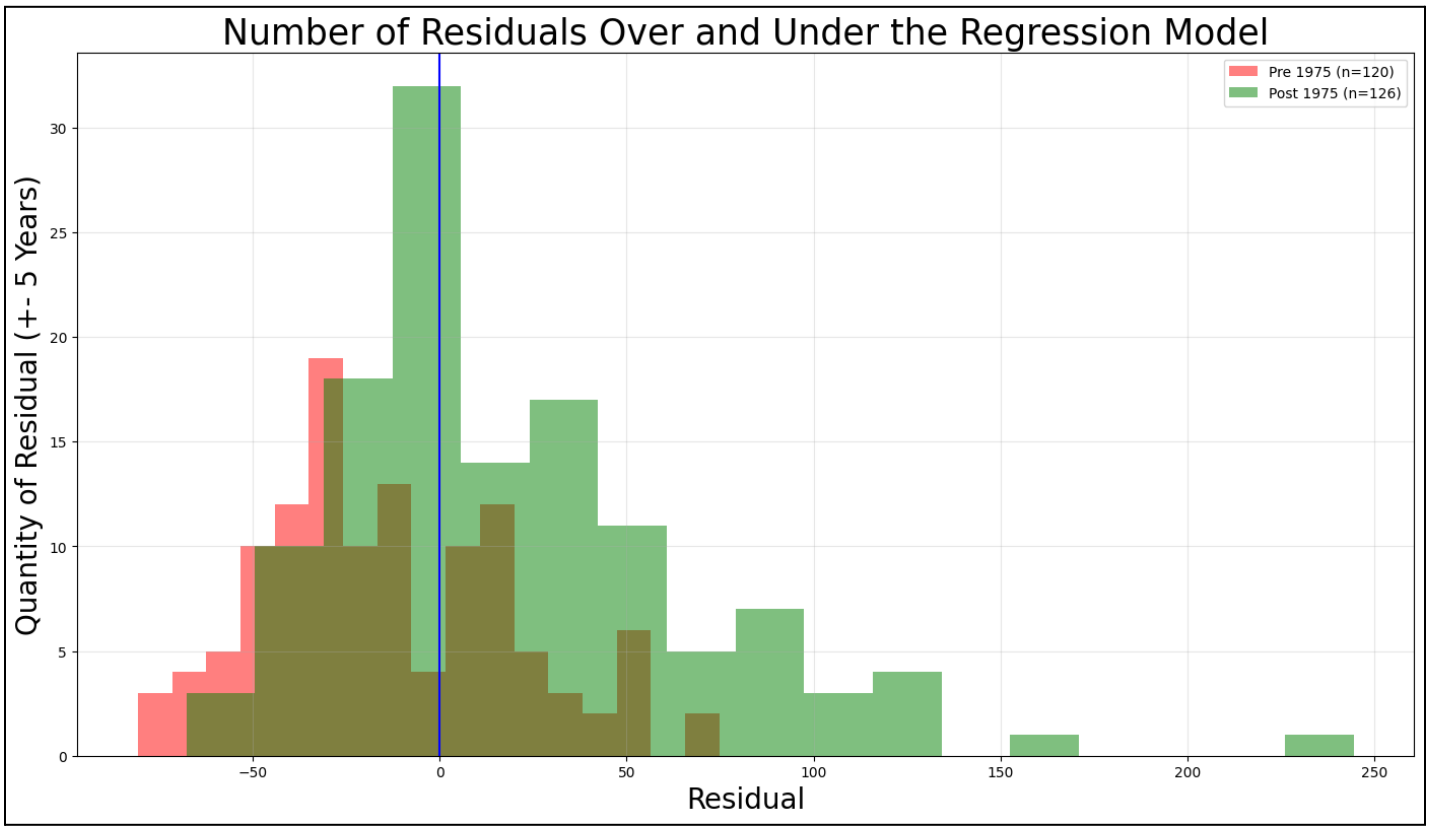

Least Squares Residuals (n=2640)
Residual
Year
3
2
1
0
−1
−2
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

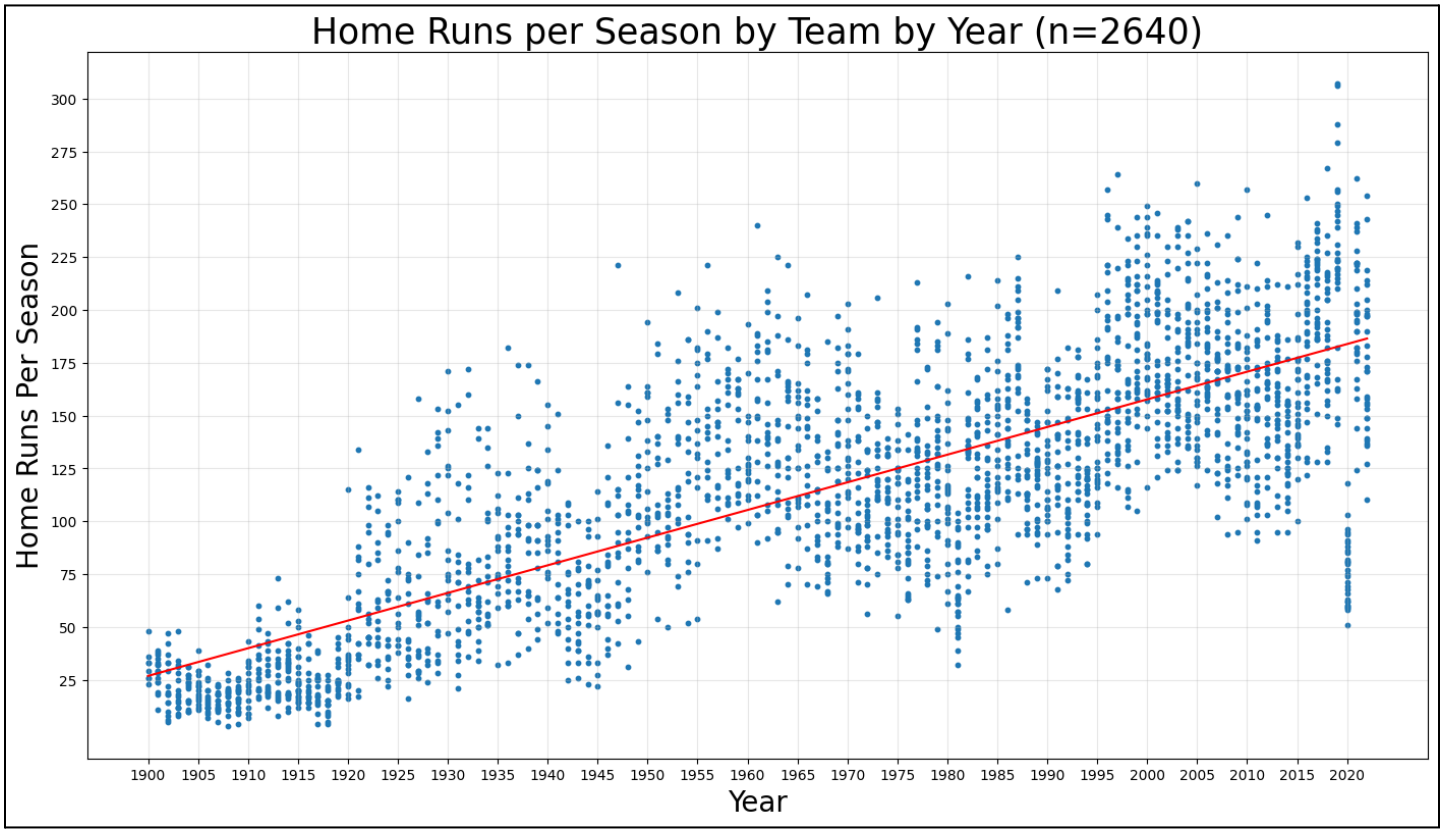

Number of Residuals Over and Under the Regression Model
Pre 1968 (n=120)
Post 1968 (n=116)
Quantity of Residual (+- 5 Years)
Residual
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

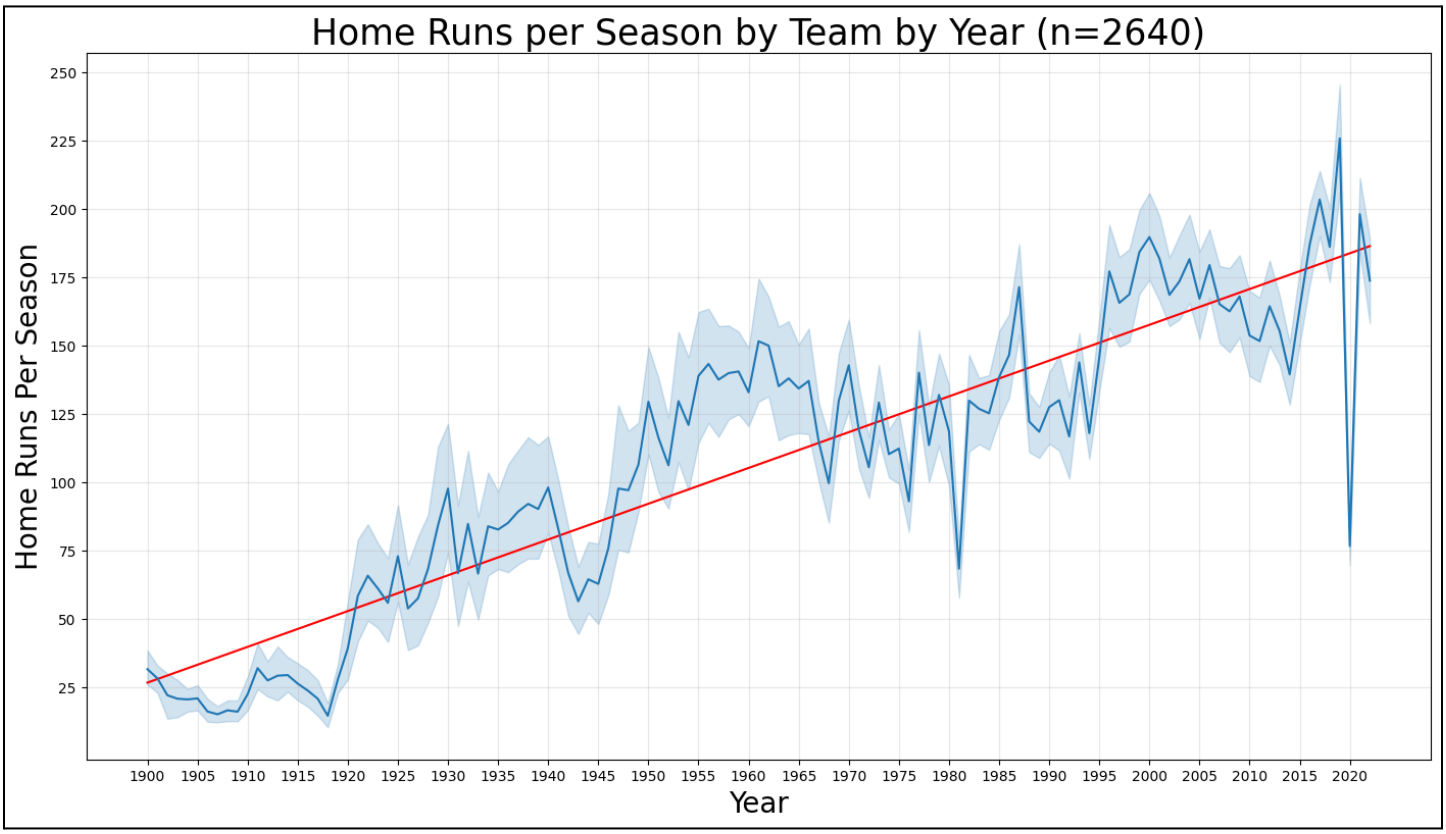

Stolen Bases Per Season by Team by Year (n=2640)
Stolen Bases Per Season
Year

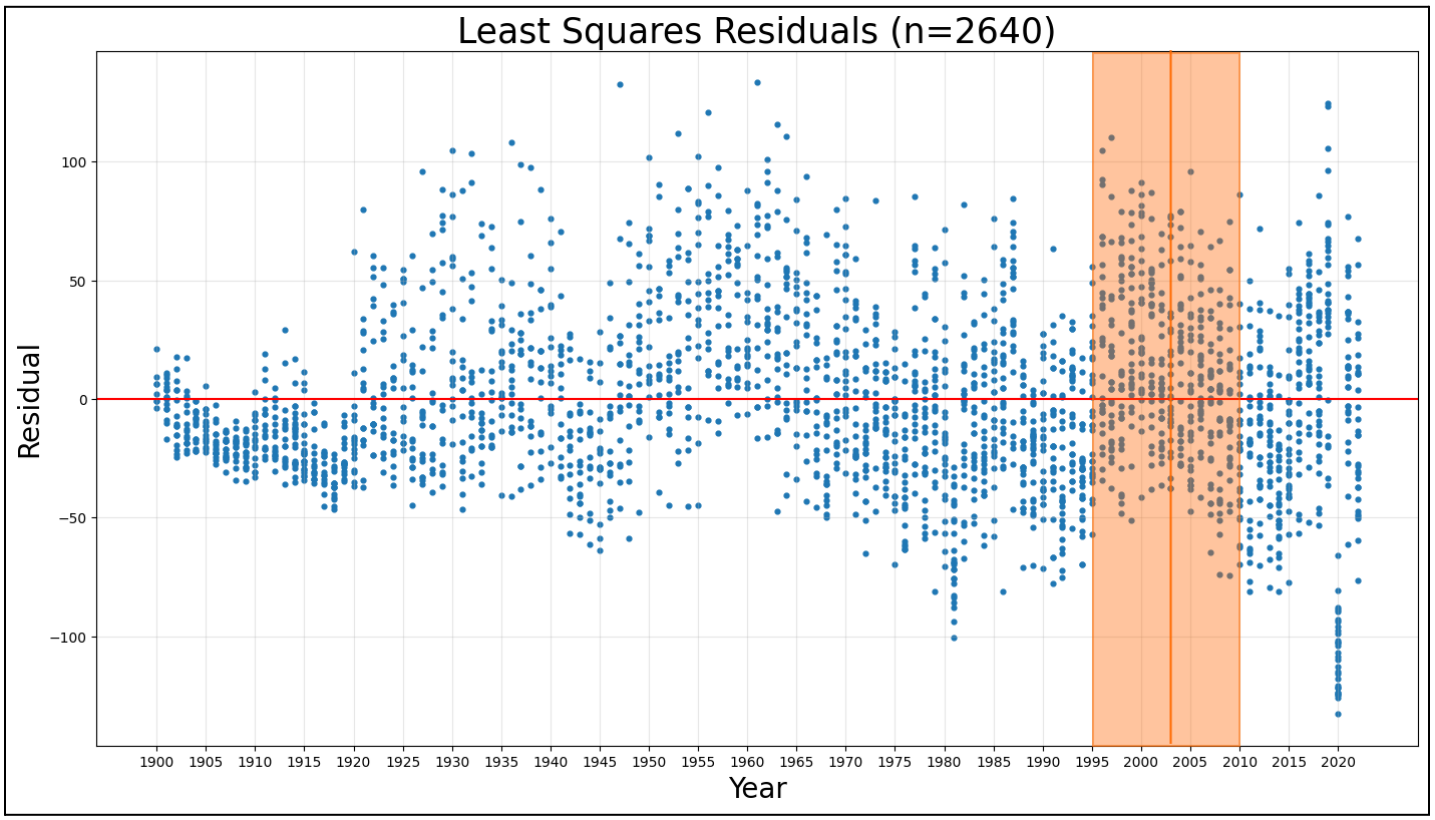

Stolen Bases Per Season by Team by Year (n=2640)
Stolen Bases Per Season
Year

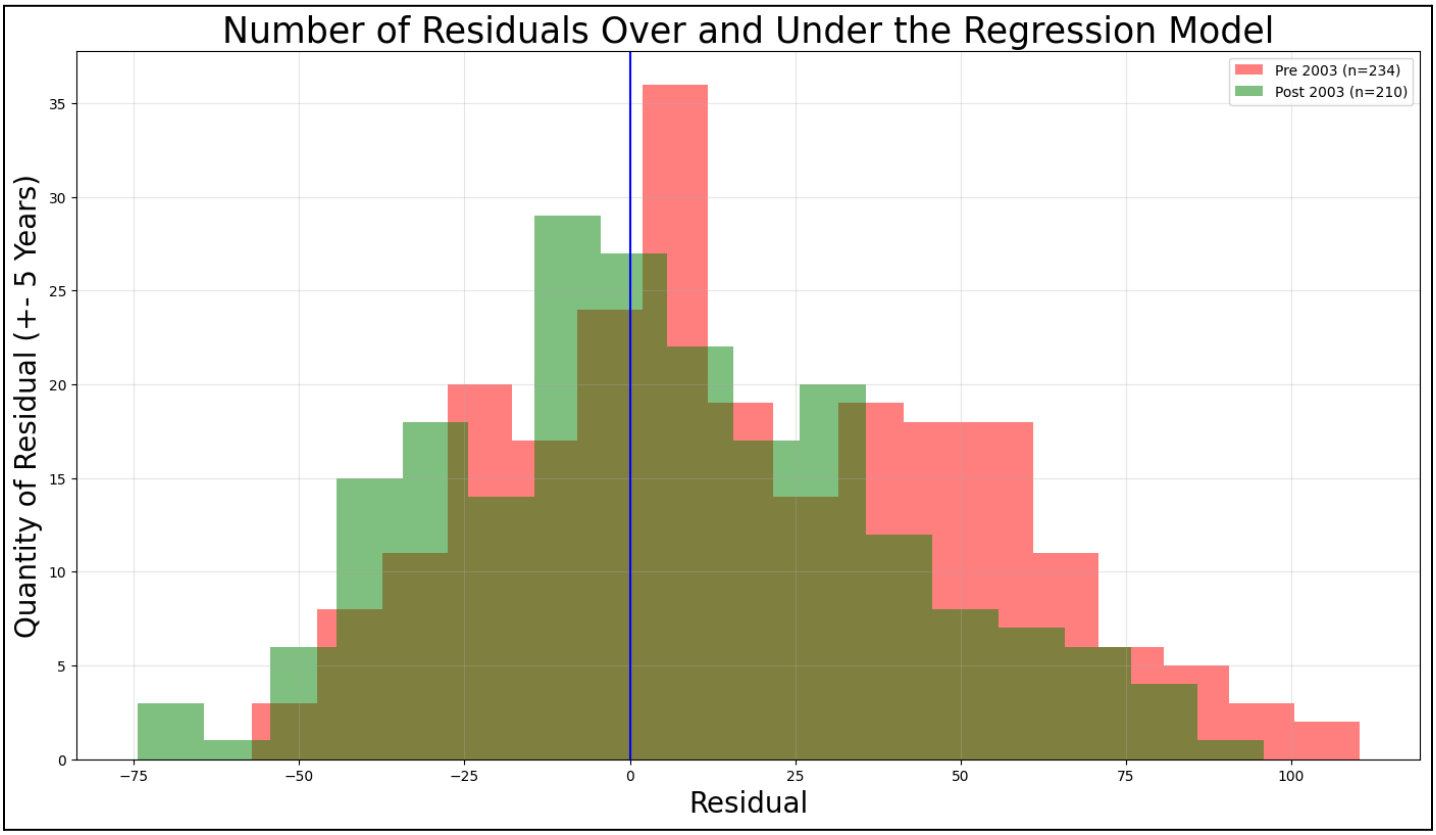

Least Squares Residuals (n=2640)
Residual
Year
250
200
150
100
50
0
−50
−100
1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020


Number of Residuals Over and Under the Regression Model
Pre 1975 (n=120)
Post 1975 (n=126)
Quantity of Residual (+- 5 Years)
Residual
30
25
20
15
10
5
0
−50
0
50
100
150
200
250


Home Runs per Season by Team by Year (n=2640)
Home Runs Per Season
Year


Home Runs per Season by Team by Year (n=2640)
Home Runs Per Season
Year


Least Squares Residuals (n=2640)
Residual
Year
100
50
0
−50
−100
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020


Number of Residuals Over and Under the Regression Model
Pre 2003 (n=234)
Post 2003 (n=210)
Quantity of Residual (+- 5 Years)
Residual
0
5
10
15
20
25
30
35
−75
−50
−25
0
25
50
75
100