

# DSC520 10.3 Final Project Part 3

Phil Han  
8/14/2021

## For the Love of Games

- Introduction

In the summer, there's no other sport like baseball in American sports. Fans go to their team's ball parks or stadiums to watch their favorite teams and players play while eating hot dogs and cold drinks. No wonder it is called National Pastime. The game filled with so many memorable moments and legends, like Babe Ruth and Ted Williams, have charmed the fans young and old. Once October arrives, the romance of the games in the summer gives way to an unscripted drama of Fall Classic games in the post season.

- Issues Addressed

However, in the recent years, baseball fans have been losing interest in the games and found baseball games to be either too long or boring. The attendance around the leagues has been steadily declining while fast-paced games like football and basketball have thrived in terms of popularity.

Theo Epstein, a former president of baseball operations for the Boston Red Sox and Chicago Cubs, recently said that he regretted that analytics used in the game has produced unintended consequences—less actions on the field (Svrluga, 2021).

- How Issues Are Addressed

This research paper seeks to find some of the main issues that might have contributed to the declining attendance for the games. It turned out that fans do not get to watch much action and athletic plays on the field when watching baseball games. Namely, there has been a decreasing number of doubles, triples, and stolen base produced on bases which all create more actions around the diamond.

I investigated whether or not there's been a historical declining pattern with the above three categories. Specifically, I looked to see if the correlation among the attendance figures and the aforementioned three categories existed. If we can see the correlation, it will provide a clue as to how to fix the current problems that baseball is facing, if not at least help improve the game by introducing more action-inducing elements.

- Analysis

I looked at the over 100 years worth of the data on batting (doubles & triples) & stolen bases from *Sean Lahman's Baseball Archive*. I also looked at the attendance data from *The Baseball Cube*. Because there's an expansion of teams and the total number of teams vary over the years, I used the mean of the categories. And also, because the average attendance per game data only existed starting in 1950, I limited the size of the data to about 70 years (1950 - 2019). I did not use the data from 2020 since no fans, except for the postseason, were allowed during the pandemic shortened season.

1953	586	6.131399	1.2713311	3.542662	1.139932	17.42833	7.402730	14383797	16	1240	898987	11600
1954	576	5.998264	1.3697917	3.362847	1.206597	17.73438	7.368056	15935983	16	1237	995993	12883
1955	655	4.963359	1.0687023	3.395420	1.059542	16.49313	6.032061	16617383	16	1234	1038586	13466
1956	621	5.376812	1.1674718	3.694042	1.157810	18.52657	6.544283	16543250	16	1239	1033953	13352
1957	615	5.521951	1.0926829	3.580488	1.247154	19.42764	6.614634	17015819	16	1235	1063489	13778
1958	638	5.316614	1.0266458	3.510972	1.161442	19.16144	6.343260	17460630	16	1235	1091289	14138
1959	632	5.503165	0.9351266	3.560127	1.349684	19.94304	6.438291	19143979	16	1238	1196499	15464
1960	637	5.403454	1.0329670	3.340659	1.448980	20.11774	6.436421	19911489	16	1236	1244468	16110
1961	698	5.694842	1.0802292	3.911175	1.498567	21.41404	6.775072	18894518	18	1430	1049695	13213
1962	760	5.675000	1.1223684	3.948684	1.773684	23.11447	6.797368	21375215	20	1621	1068761	13186
1963	752	5.449468	1.0518617	3.595745	1.643617	24.96410	6.501330	20477074	20	1619	1023854	12648
1964	754	5.663130	1.0079576	3.663130	1.559682	25.48011	6.671088	21280341	20	1626	1064017	13088
1965	751	5.591212	1.0479361	3.579228	1.929427	25.67643	6.639148	22441900	20	1623	1122095	13827
1966	774	5.322997	1.0594315	3.543928	1.879845	24.29587	6.382429	25182209	20	1615	1259110	15593
1967	786	5.193384	1.0076336	2.924936	1.746819	24.69847	6.201018	24308353	20	1620	1215418	15005
1968	715	5.411189	0.9748252	2.790210	2.118881	26.77343	6.386014	23102745	20	1625	1155137	14217
1969	932	5.193133	0.9109442	3.346567	1.984979	24.11266	6.104077	27229666	24	1946	1134569	13993
1970	919	5.698409	1.0087051	3.731230	2.076170	24.34603	6.705114	28747333	24	1944	1197806	14788
1971	883	5.584371	0.9150623	3.242356	1.998867	23.73273	6.499434	29193417	24	1938	1216392	15064
1972	889	5.232846	0.8391451	2.850394	2.032621	23.30484	6.071991	26988268	24	1859	1123678	14507
1973	892	5.856502	0.8856502	3.477578	2.280269	22.82287	6.742152	30108926	24	1943	1254539	15496
1974	914	5.695842	0.9266958	2.898249	2.722101	21.32932	6.622538	29576474	24	1945	1232353	15206
1975	907	6.001103	0.9779493	2.974642	2.782800	21.25689	6.979052	29347395	24	1934	1222808	15174
1976	886	5.914221	1.0902935	2.522573	3.446953	21.15688	7.004515	31318331	24	1939	1304930	16152
1977	984	6.545732	1.1890244	3.703252	3.066057	22.07520	7.734756	38709779	26	2103	1488838	18407
1978	960	6.443750	1.0625000	3.079167	3.129167	20.89375	7.506250	40636886	26	2102	1562957	19332
1979	961	6.675338	1.1092612	3.572320	3.104058	20.84807	7.784599	43550398	26	2099	1675015	20748
1980	950	6.678947	1.1326316	3.249474	3.467368	21.27579	7.811579	43014136	26	2105	1654390	20434
1981	944	4.237288	0.6980932	1.886653	2.140890	14.02225	4.935381	26544376	26	1934	1020938	19042
1982	992	6.366935	0.9717742	3.406250	3.201613	21.39214	7.338710	44587874	26	2107	1714918	21162
1983	1006	6.424453	1.0268390	3.281312	3.305169	21.58648	7.451292	45540338	26	2109	1751551	21593
1984	984	6.314024	1.0010163	3.310976	3.081301	22.86585	7.315041	44742863	26	2105	1720879	21256
1985	998	6.435872	0.9669339	3.609218	3.103206	22.49599	7.402806	46824379	26	2103	1800938	22266
1986	1017	6.402163	0.8407080	3.749263	3.256637	24.29302	7.242871	47506203	26	2103	1827162	22590
1987	1048	6.481870	0.8549618	4.253817	3.420802	23.94943	7.336832	52011506	26	2105	2000443	24709
1988	1035	6.170048	0.8115942	3.072464	3.189372	22.56522	6.981643	52998904	26	2100	2038419	25238
1989	1073	5.877912	0.8089469	2.873253	2.904007	22.04101	6.686859	55173096	26	2106	2122042	26198
1990	1115	5.852915	0.7757848	2.974888	2.950673	21.39283	6.628700	54823768	26	2105	2108606	26045
1991	1086	5.984346	0.8232044	3.115101	2.872928	22.45856	6.807551	56813760	26	2104	2185145	27003
1992	1066	5.156660	0.7926829	2.849906	3.061914	22.08068	6.949343	55870466	26	2106	2148864	26529
1993	1180	6.312712	0.7966102	3.415254	2.765254	22.29661	7.109322	70257938	28	2269	2509212	30964
1994	1030	5.556311	0.6815534	3.209709	2.192233	19.19029	6.237864	50010016	28	1600	1786072	31256
1995	1253	5.653073	0.6576217	3.256983	2.340782	20.29130	6.210694	50469236	28	2017	1802473	25022
1996	1253	6.374302	0.6823623	3.960096	2.584996	23.39026	7.056664	60097381	28	2267	2146335	26510
1997	1236	6.475728	0.7144013	3.754045	2.676375	24.22087	7.190129	63168689	28	2266	2256025	27877
1998	1322	6.611952	0.6800303	3.830560	2.484115	24.12481	7.291982	70601147	30	2432	2353372	29030
1999	1299	6.728253	0.7167052	4.255581	2.633564	23.95612	7.444958	70103204	30	2428	2336773	28873
2000	1384	6.431358	0.6878613	4.113439	2.112717	22.65607	7.119220	71358907	30	2429	2378630	29378
2001	1339	6.581777	0.6930545	4.076176	2.317401	24.20015	7.274832	72581101	30	2429	2419370	29881
2002	1319	6.595906	0.6982563	3.835481	2.084913	23.80136	7.294162	67944389	30	2426	2264813	28007
2003	1347	6.553081	0.6933927	3.865627	1.910171	22.86637	7.246474	67630052	30	2430	2254335	27831
2004	1346	6.626300	0.6671620	4.049777	1.923477	23.64636	7.293462	72989219	30	2428	2432974	30061
2005	1330	6.663910	0.6676692	3.772180	1.928571	23.04060	7.331579	74385295	30	2431	2479510	30599
2006	1377	6.633987	0.6913580	3.911402	2.009441	22.98838	7.325345	76043902	30	2429	2534797	31307
2007	1385	6.640433	0.6772563	3.579061	2.106859	23.24116	7.317690	79484718	30	2431	2649491	32696
2008	1385	6.508303	0.6397112	3.522022	2.020939	23.74296	7.148014	78591125	30	2428	2619704	32369
2009	1388	6.294669	0.6837176	3.632565	2.139769	24.20101	6.978386	73418528	30	2430	2447284	30213
2010	1356	6.258112	0.6386431	3.401917	2.182153	25.29941	6.896755	73061123	30	2430	2435371	30066
2011	1389	6.046796	0.6465083	3.277178	2.360691	24.82937	6.693305	73451522	30	2429	2448384	30239
2012	1408	5.867188	0.6583807	3.504261	2.293324	25.87074	6.525568	74859268	30	2430	2495309	30806
2013	1409	5.835344	0.5479063	3.308020	1.911285	26.05394	6.383251	74026895	30	2431	2467563	30451
2014	1435	5.670383	0.5916376	2.917073	1.926132	26.09129	6.262021	73739622	30	2430	2457987	30346
2015	1486	5.546433	0.6318977	3.303499	1.685734	25.19919	6.178331	73780032	30	2429	2458668	30366
2016	1483	5.565745	0.5886716	3.782873	1.701722	26.28591	6.154417	73159068	30	2428	2438636	30131
2017	1494	5.620482	0.5321285	4.086345	1.691432	26.84337	6.152610	72670423	30	2430	2422347	29906
2018	1535	5.383713	0.5517915	3.638436	1.611726	26.84495	5.935505	69649746	30	2431	2321658	28651
2019	1569	5.437221	0.5003187	4.318674	1.453155	27.29318	5.937540	68494845	30	2429	2283162	28199

First, I looked at the relationship between doubles hit (2B) and the average attendance per game (AAPG) based on the data I extracted. As you see 1) Scatter Plot with Regression Model below there is a positive correlation.

Plot of Avg. 2B versus Avg. Attendance Per Game