Final Project

Owen Braglio, Dean Bae, Jack Billings

2022-12-02

Report

Research Question:

What is the relationship between a baseball team's runs scored, runs allowed, and their win rate for that season?

We wanted to find if there was a significant link between the runs scored by a team in a season, and their win rate. The initial hypothesis is yes of course, it should be a direct relationship, but factors like blowout games and close defeats may skew the numbers and make them less accurate than initially presumed.

Data Source

We got both of our data sources from the MLB's website as they have collected all relevant data into a couple tables for each Baseball League. As all teams were gather from the same source it also ensured consistency between the numbers reported.

Data Focus

We'll be focusing on Win Rate, Runs Scored and Runs Allowed to determine if there is a significant like between these factors and if a team made it to playoffs or not that season. This allows us to both narrow down our sample sizes as well as having multiple sets with which to compare between.

```
##    Playoff Teams    PCT(Win Rate)  RS(Runs Scored) RA(Runs Allowed)
##      "character"        "numeric"        "numeric"        "numeric"
```

Not Playoff Teams Summary table

count	type	mean	min	firstQuin	secondQuin	median	thirdQuin	fourthQuin	max	standdev
18	Win Rate	0.440	0	0.39	0.42	0.438	0.457	0.492	0.531	0.054
18	Runs Scored	656.222	557	595.80	646.40	665.500	688.000	705.000	735.000	56.774
18	Runs Allowed	748.444	668	688.00	716.20	735.500	748.400	813.000	873.000	64.267

The first statistic, Win Rate, increases by close to .3 points every quintile, but increases moderately as it nears to top which shows a slight skew by the best teams on the fringe of the cluster. Runs scored has an

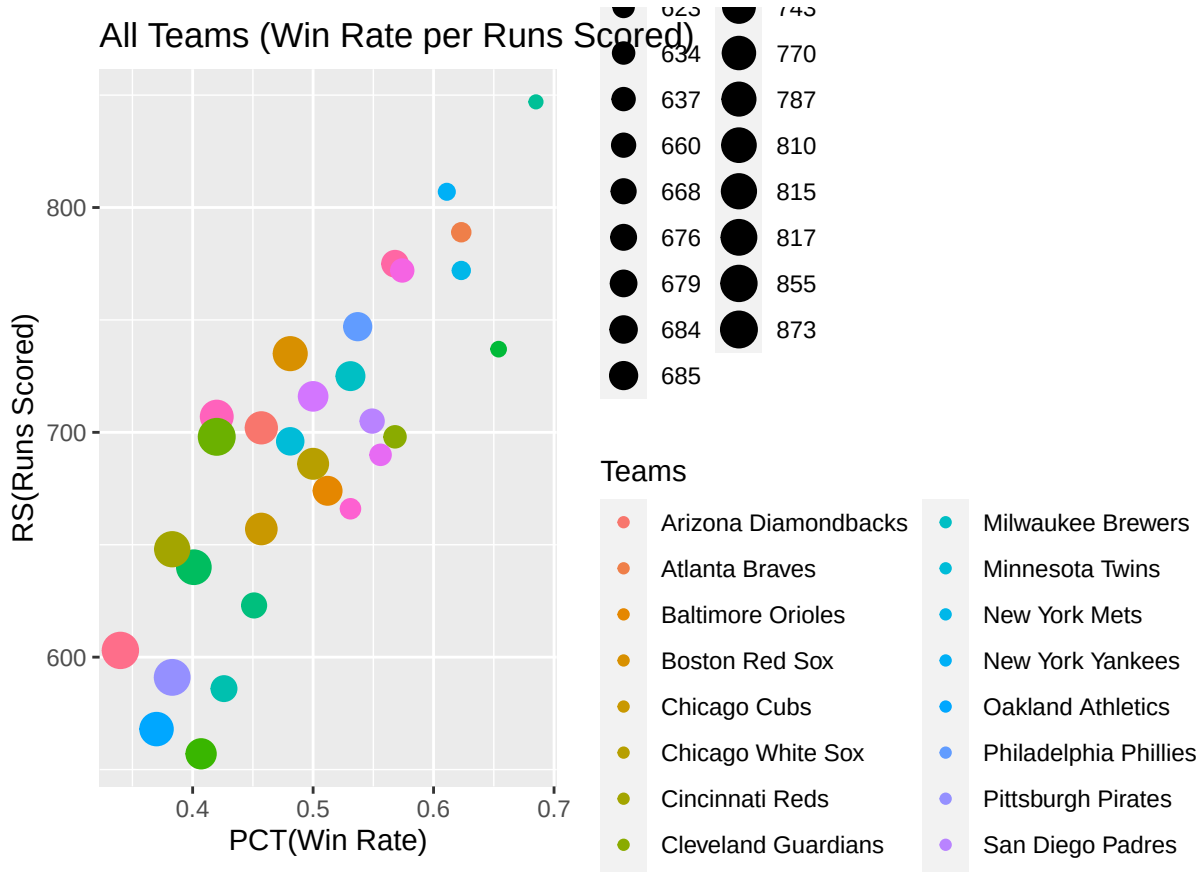
inverse of this phenomenon, the increase to each quintile starts off at about 40 and decreases. This can also be seen on the graphs where the non playoff teams start to drop in runs scored where they meet those who made it to the playoffs. Runs allowed is where the greatest variance can be seen, starting off by an 20 point increase and ending with an increase of 60 from the fourth quintile to maximum. Where runs scored had an obvious link to win rate, runs allowed is less bound.

Playoff Teams summary table

count	type	mean	min	firstQuin	secondQuin	median	thirdQuin	fourthQuin	max	standdev
12	Win Rate	0.590	0.531	0.55	0.568	0.571	0.596	0.623	0.685	0.049
12	Runs Scored	750.417	666.000	699.00	741.000	759.500	772.000	786.000	847.000	53.371
12	Runs Allowed	612.083	513.000	575.00	611.000	618.000	630.000	655.000	685.000	55.647

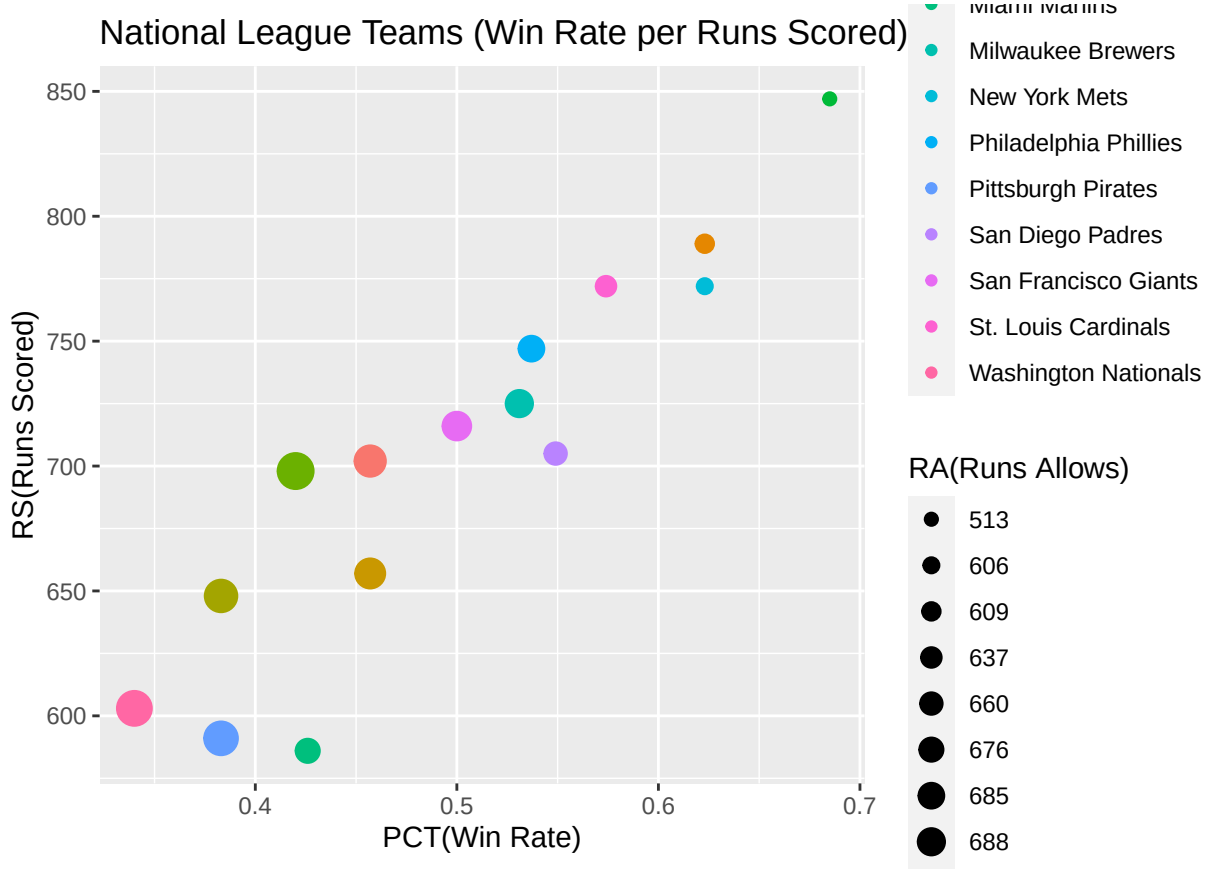
The teams which made it to the playoffs are the 6 top teams from each league, and usually the teams skewing the data on the edges. Initially we can see a much less deviance from the win rates. these teams are more consistently clustered near .57 with the outliers being on the high end. Runs scored are also on average much higher, as expected, but cluster on the lower side due to teams like the Dodgers raising the maximum so high. Runs allowed is also much lower than its non playoff counterpart, but is more consistent with its value which have a range of 172 as opposed to non playoff's 205. As expected from taking the best teams into their own set, the data is much more of a higher consistent standard than when measured with a larger population.

All League Teams chart



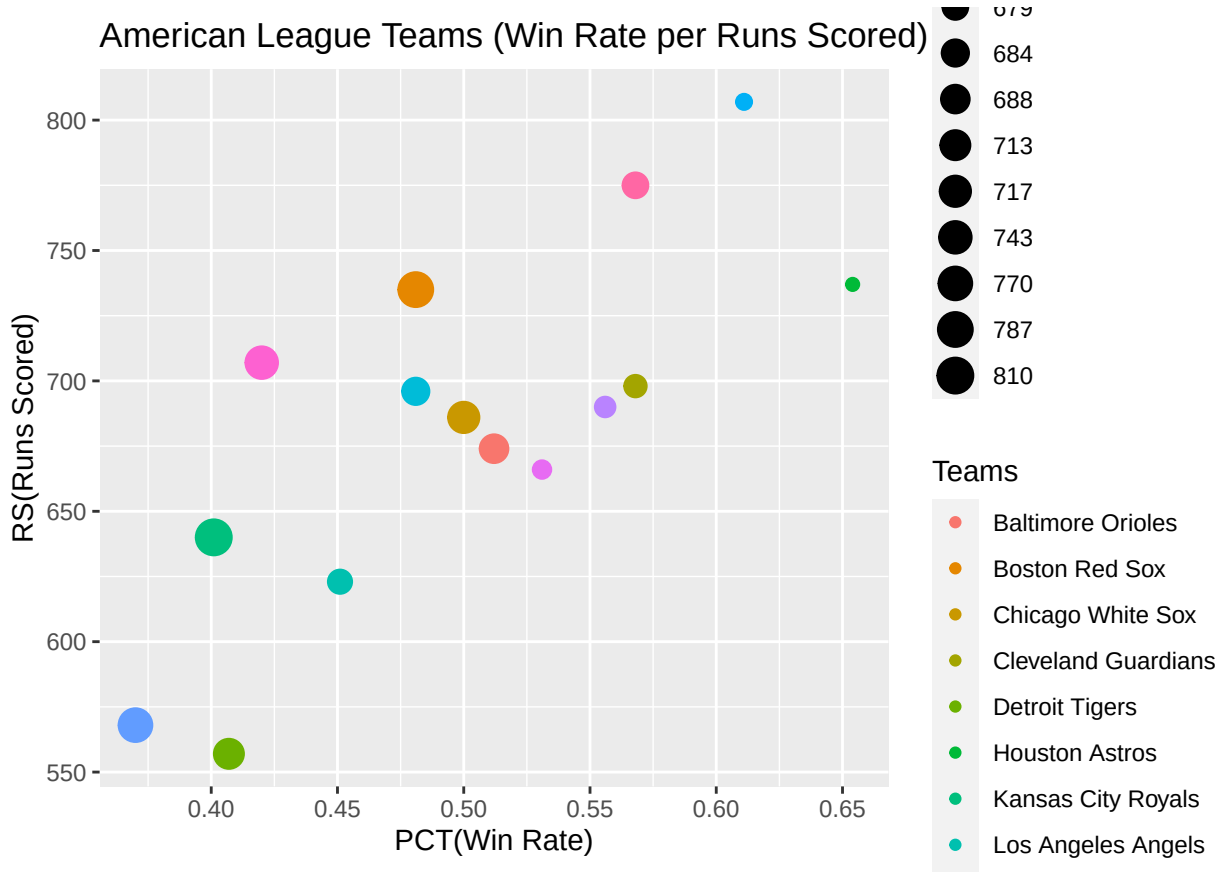
As can be observed by the general trend of the graph, throughout the entire league there is a distinct linear correlation between runs scored and the win rate of a team. The data present can be fitted nicely to a best fit line, which insinuates that the runs scored could be used to find a slope that dictates the graph as a whole within a margin of error. A general trend can also be observed between win rate and runs allowed, or the size of each point. As a team's win rate increases it quite consistently also sees a decrease in runs allowed, but this is not the case between runs scored and runs allowed. There are multiple instances of two teams having the same win rate where the team with more has also more runs allowed. From this we can infer that the relationship between runs scored and allowed is an inverse linear relationship where the difference between them indicates the win rate of a team.

National League Teams chart



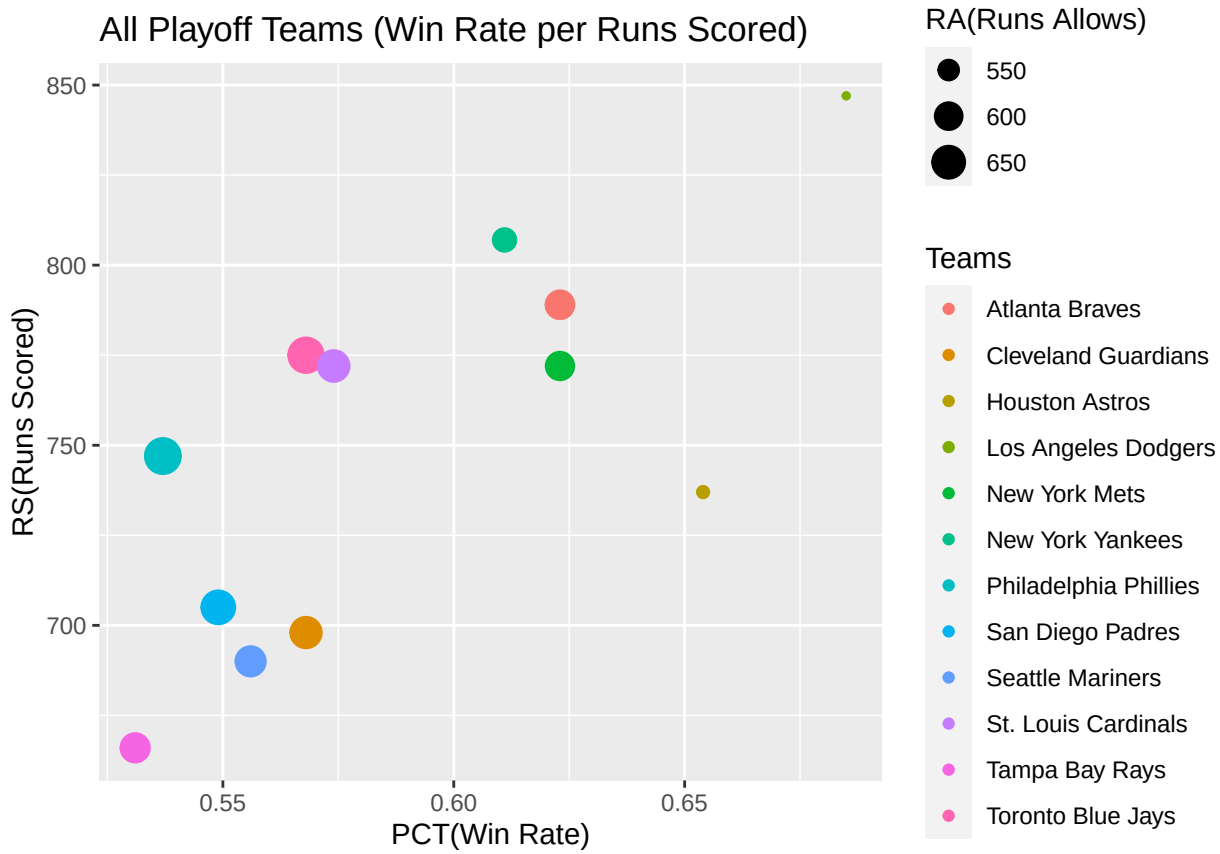
Here can be seen the most obvious example of the linearity that exists between win rate and run scored. While there are some outliers and gaps, the trend reinforces the hypothesis and observations made about the league wide statistics.

American League Teams chart



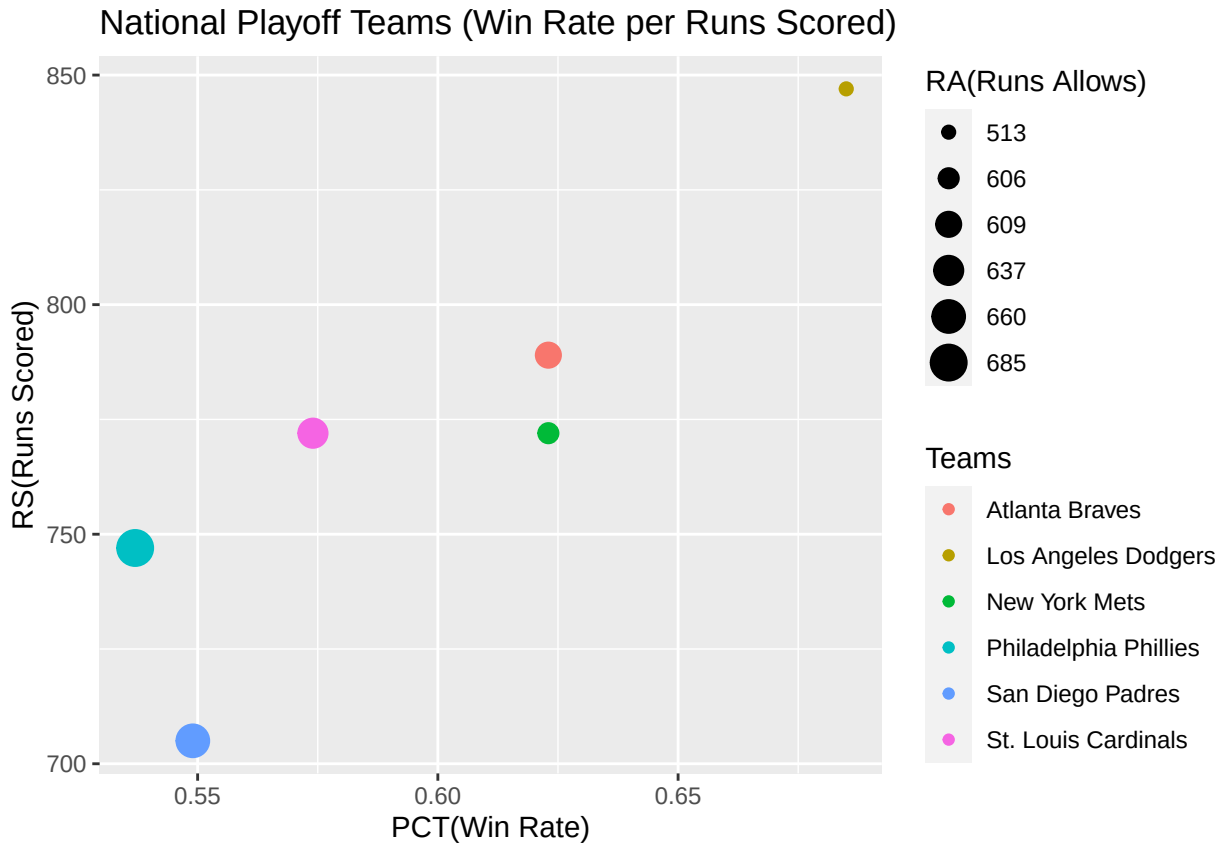
While this graph doesn't appear to be as linear as the others, it serves to corroborate the observations made regarding the runs allowed. For the points excluding the top fringe case, runs allowed is dictated be the win rate of a team and its runs scored. This shows the negation of these two statistics upon each other, where a team which tends to have higher scoring, close, games and a team with lower scoring, close, games will have similar win rates and blowout games have much less of an effect than presumed.

Play off League Teams chart



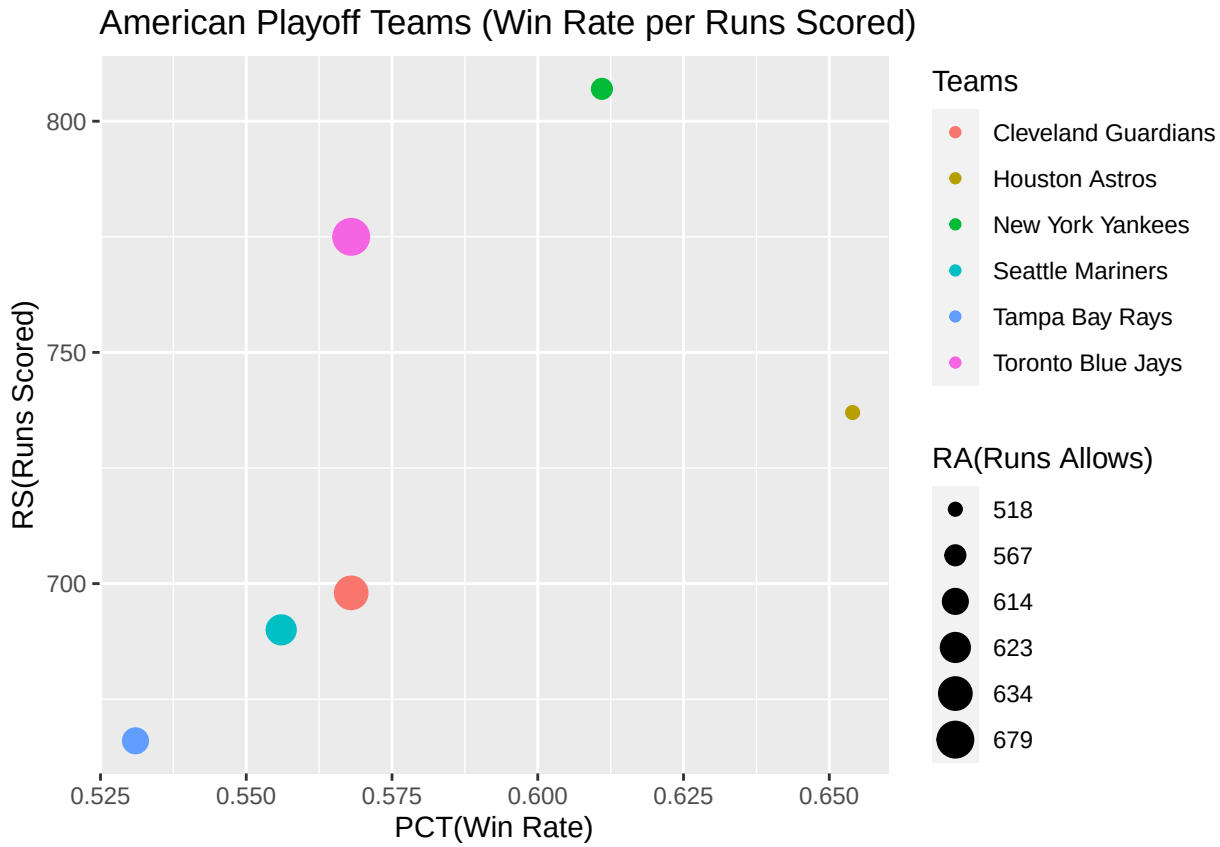
The playoff teams here are the top six teams taken from each league. This may lead to a slight skewing of the data near the top end, due to a greater number of games played, which is why we decided to evaluate them on their own as well. As can be seen from the world series results, the win rate and other statistics here do not have a great effect on their chances of winning. The two teams who made it to the finals were the Houston Astros, which were second in win rate as expected, and the Philadelphia Phillies, which were second last in win rate and a major upset. As the winner is decided by an elimination style bracket, the runs scored vs allowed is not a deciding factor but can be considered a loose indicator.

National Play off Teams chart



The major outlier of this graph is the Los Angeles Dodgers who had a series of blowout games padding their statistics. Even compared to the American League, the Dodgers were an extreme fringe case which were favorites to win the world series, but in the end didn't make it.

American Play off Teams chart



While the highest win rate team here, the Houston Astros, did eventually win the world series, the New York Yankees had almost 75 more runs scored while only 50 more allowed. This disparity appeared to not have affected the overall score, where the lower scoring but more defensive Astros eventually won it all.