

Statistical Analysis

Curtis Higa

January 8, 2018

Approach

I approached this project in two ways to see if there was a correlation between various 3-point statistics and winning basketball games. The first was to investigate any patterns between 3-pointers made, 3-point percentage, and the percentage of points contributed by 3's for each team in a game. The second was to look at those same basketball statistics but by each team averaged over the entire season.

Game-By-Game

By plotting each of the three statistics, no relationships could be definitively recognized but small differences between losing and winning teams numbers. Drawing histograms also showed there was a slight difference in the numbers for winning teams and losing teams but looking at each of the three different basketball statistics (see below for table) for all 1230 unique games and analyzing the numbers, no patterns could be concluded about who would win a game. First taking the difference of three pointers made by the losing team from the winning team (`diff_tpm_mean`) for each game yielded an average of 1.9 with a standard deviation of 4.5. For the 95% confidence interval, that's roughly 1.65σ above and below the mean which equates to $-2.59 \leq \mu \leq 6.4$. That means making more three-point shots doesn't determine if you win or lose in a particular game.

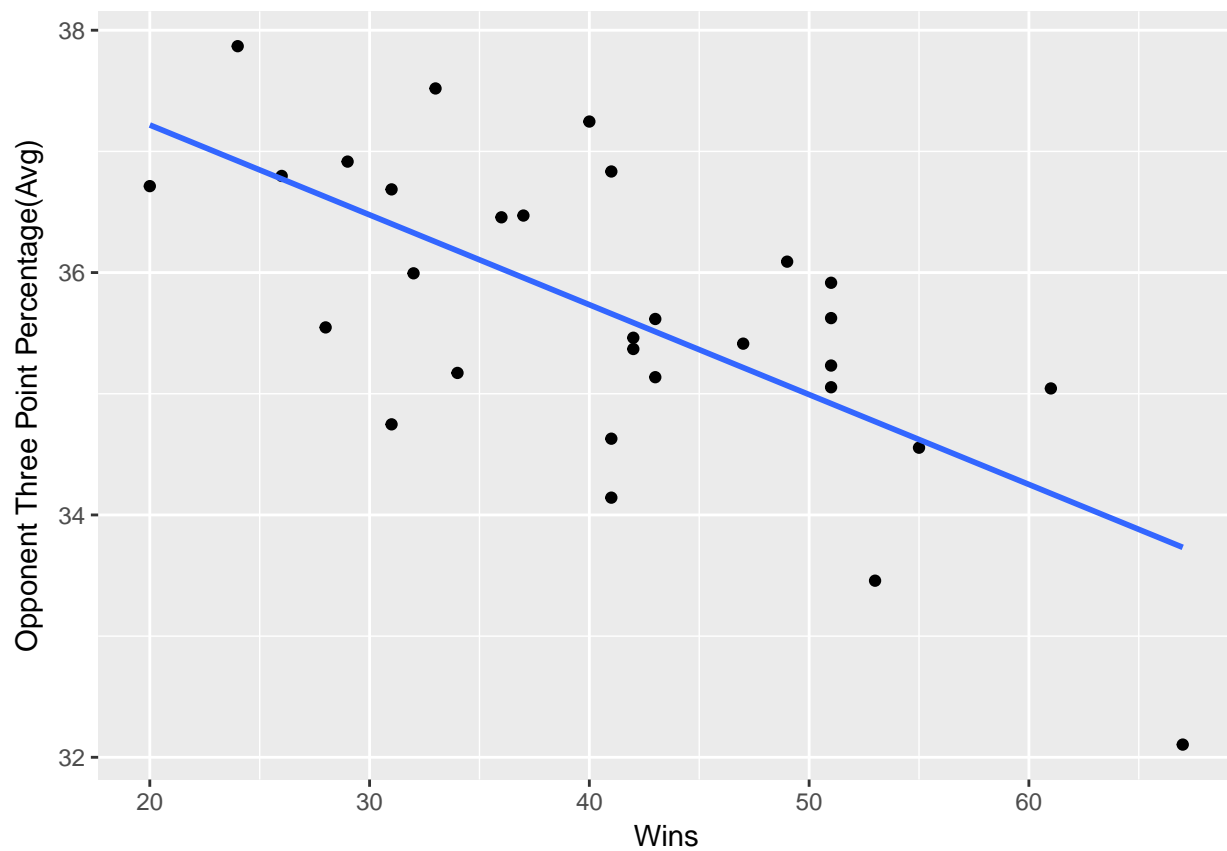
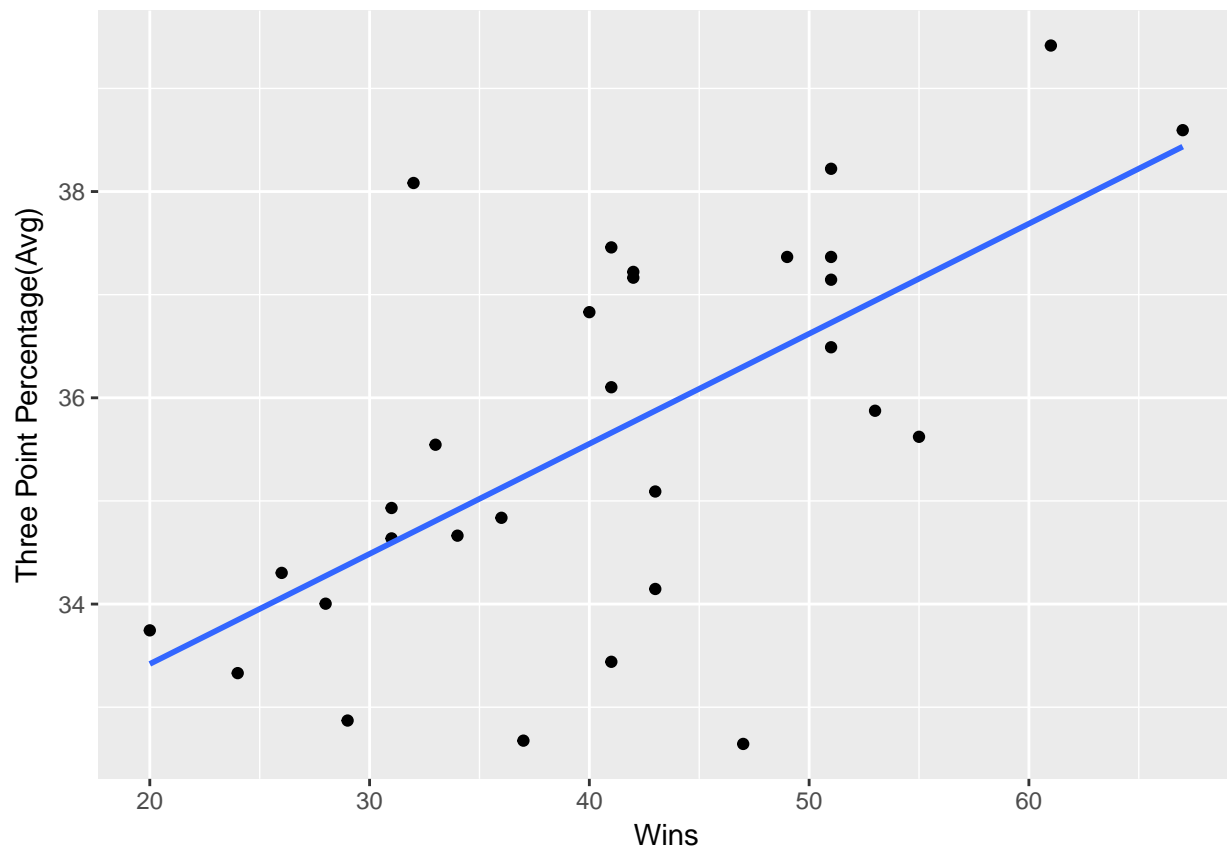
The second statistic we looked at was the difference in three-point percentage per game of the losing team from the winning team (`diff_tpp_mean`). This statistic yielded a mean of 6.81 with a standard deviation of 4.5. Again we would need $\pm 1.65\sigma$ around the mean for a 95% confidence interval, giving $-5.15\% \leq \mu \leq 18.77\%$. As with three-pointers made, simply shooting better or worse from three won't determine the outcome of the game.

Lastly, the third statistic looked at the the percentage of total points were contributed by each team in a game. Like the other statistics, we took the difference of the losing team from the winning team and looked at the numerical statistics. These came out to be $\mu = 2.48$ and $\sigma = 12.1$. Here the 95% confidence interval will be $-9.61\% \leq \mu \leq 14.56\%$ meaning that comprising more of your points of three-pointers doesn't determine the winner of the game.

<code>diff_tpm_mean</code>	<code>diff_tpp_mean</code>	<code>diff_pop_mean</code>	<code>diff_tpm_sd</code>	<code>diff_tpp_sd</code>	<code>diff_pop_sd</code>
1.902	6.809	2.476	4.496	11.96	12.09

By Team

For looking at each teams statistics, each one was averaged out over their 82 games and compared to the number of wins and win percentage for the entire season. Of those statistics, three-point percentage and opponents three-point percentage showed a strong linear relationship to the total number of wins in the teams season (see below).



A table was created to see how well these linear regression lines fit the data and plotted below. The closer the values are to 1 or -1, the better they fit the line. Here it's easier to visualize that when a team has higher three-point percentage or keep their opponents from making three-point shots, they have a better chance of succeeding throughout the season.

win_team_mean	1.0000000
Percent_of_Points_mean	0.3572426
opp_Percent_of_Points_mean	-0.1694386
three_pt_Percentage_mean	0.6315019
opp_three_pt_Percentage_mean	-0.6776963

