

Predicting National Football League (NFL) Stadium Attendance

Jennifer L. Bowley
Bentley University
Waltham, MA. 02452
U.S.A.

Paul D. Berger
Bentley University
Waltham, MA. 02452
U.S.A.

Abstract

In this paper we explore the factors that determine attendance at NFL games. After an introduction and literature review, we perform and discuss the results of various statistical analyses. We first investigate the impact on attendance of a variety of factors on a univariate level. Then we consider two multivariate analyses, multiple regression and stepwise regression, the former to consider all of the variables simultaneously, and the latter to ferret out levels of multi-collinearity to get a clearer picture of which variables are significant. We find that the two most important variables are the number of wins a team had previously (which, of course, indicates potential for future wins and the quality of the team,) and the number of individual "star players" on a team. We end the paper with a discussion of limitations of the study and directions for future research.

Introduction

The aim of this paper is to attempt to determine which factors predict stadium attendance for regular season games in the NFL. The NFL is a huge market in the United States, with millions of fans attending games every season. This is a particularly relevant topic today, since in recent years, the NFL has considered moving several teams to new locations. They have already followed through with one move, sending the Rams to Los Angeles, CA from St. Louis, MO, in an effort to attract a larger fan base. More recently, the San Diego Chargers have also been moved to Los Angeles for the start of the upcoming 2017-2018 season. Additionally, the Oakland Raiders have arranged to move to Las Vegas, NV, although they will spend the next two seasons in Oakland before moving to Las Vegas for the 2019-2020 season. Throughout this paper, we will aim to define a list of factors that we suspect may contribute to NFL stadium attendance, perform analyses to determine which of these factors relate more strongly to stadium attendance, and document any relevant conclusions or suggestions to improve future research in this field.

Literature Review

Many NFL fans typically focus on player performance and statistics and are not aware of literature regarding analyses of patterns and fan behavior within the NFL, as well as other sports. Several types of studies have been conducted regarding what impacts stadium attendance in a number of different sports, both professional and collegiate. DeSchraver (2002) analyzed football attendance specifically at Division II NCAA stadiums. Some of the factors he analyzed were similar to the ones examined in this study (given in further detail in the Methodology section below), such as team wins and ticket prices, and his study informed our study.

However, NFL attendance analysis deals with different figures and fan bases, and we include additional variables exclusive to the NFL. DeSchriver (2011) also conducted similar research to analyze the attendance factors of Major League Soccer (MLS) in an article on the effect of expansion teams and soccer-specific stadiums. This analysis showed how this mode of data analysis is relevant and useful for any type of sports at any level of playing.

Another analysis of college football attendance was conducted by McKnight et al. (2016), who conducted research to analyze which demographic variables can predict stadium attendance in order to improve marketing in college football in all divisions. They looked at factors such as team performance, division, and college type in order to analyze what contributes to attendance at collegiate football games. Their specific intent was to be able to develop marketing strategies that may attract more fans to college football games; our intent is to investigate what predicts NFL attendance and how that could potentially explain or contribute to the NFL team moves happening and being considered in recent years.

There have also been several articles and news stories in recent years regarding the relocation of several NFL teams. Prior to the move of the St. Louis Rams back to Los Angeles, where they were previously located from 1946-1995, the most recent relocation of an NFL team was the Houston Oilers to Tennessee in 1997, to eventually become the Tennessee Titans. There has been almost a 20 year gap since any teams were moved or even considered for moves, and now, in addition to the Rams moving, and as we noted earlier, the Oakland Raiders and the San Diego Chargers have made plans to move cities. We are interested in what factors may have caused this recent increase in concern for relocating certain franchises. A set of articles on the mathematical analysis behind stadium attendance was written by Coates & Humphreys (e.g., Coates & Humphreys, 2007), which proposes several different strategies and models regarding moving teams or creating teams in new cities. Their research suggests that one of the main factors in determining where to place a team is the city's population, but also brings up the perspective of looking at how a city can benefit from a sports team. While the NFL is trying to find a city where a team can be successful at appealing to fans, that city also needs to be willing to house an NFL stadium, which usually isn't a problem due to "the creation of net new income and jobs by professional sports teams," drawing attention to the fact that moving a team to a new city is a two way relationship. This is important to keep in mind, although our analysis may not reveal much insight into the process of how an NFL team selects a city and how that city agrees with the NFL's terms.

Methodology

In order to ensure consistency and better determine trends, we decided to gather data for more than one season. Since we are investigating attendance in recent years and how it corresponds to the recent discussions about relocating teams, we focused on the 2013, 2014, and 2015 NFL seasons. Another important point is that looking solely at the attendance numbers for each team would be misleading, since some NFL stadiums have much higher capacities than others. Instead, it is necessary to compute the stadium attendance *percentage* for each team. Every season, each team has 8 home games and 8 away games¹. The yearly average attendance percentage for all home games is calculated by dividing the sum of total attendees at all home games throughout the season by the total potential stadium capacity throughout the season (multiplied by 100 to convert to a percentage). Average stadium attendance percentage will be the dependent variable being investigated, so that the relationship between stadium attendance and a number of other factors can be explored; data were derived from Pro Football Attendance (2016), Weekly League Attendance (2016), and ESPN (2016).

We discovered that the recorded stadium attendance for several teams in many instances throughout the past three seasons is greater than that team's official stadium capacity. This is because some stadiums are able to incorporate additional seating and expand their capacity for games that are in high demand. This is why some average attendance percentages exceed 100% for some teams in the analyses we conducted. Another note of caution about stadium attendance numbers in general comes from a *Forbes Magazine* article written by Brown (2011), which describes how attendance is often tracked by tickets sold and may not necessarily reflect the number of spectators present. It additionally explains that some teams can oversell tickets and provide additional seating in stadiums or as standing room only, in order to fill their stadium as much as possible.

¹ Each season, lately, features a game in London, and possibly another game in Mexico City. These are "official" home games for one of the teams playing. We have not separated out these games, but believe that the results of our analyses are not materially affected by this issue.

While the attendance numbers might not document the exact number of attendees, these figures can still be used for the purpose of this analysis since higher ticket sales should generally reflect the same patterns as higher stadium attendance. As a precaution, we are including each team's stadium capacity as a separate factor in the analysis in order to examine whether in some cases larger stadiums may be harder to fill and thereby have lower attendance rates.

When trying to create a list of factors that could potentially impact stadium attendance, we wanted to be comprehensive and include variables from a variety of sources. Our primary consideration was that NFL teams who perform better will likely attract larger crowds. Therefore, it is vital to include measures of each team's level of success for each season. Thus, one variable included in our analysis is each team's number of wins that season. However, this variable may not be very useful in predicting stadium attendance near the beginning of the season, since early in the season they will have played very few (or zero) games and may therefore have zero wins - or, at minimum, a small sample of games. Consequently, we are also including each team's number of total wins from the previous regular season, which will provide some insight into how each team was expected to do heading into the current season. It is also important to focus on each team's number of *home* wins, both from the season being analyzed and the previous season, because stadium attendance deals only with each team's 8 home games per season. Some teams have much better records and perform significantly better at their home field, as opposed to away games or vice versa. Teams that perform better at home than on the road could potentially have higher home-field attendance than their overall record may suggest.

Another relevant factor that is more difficult to measure is a team's popularity or number of highly popular *individual* players on their team. Thus, we sought a way to determine whether a team has any "star" players who become very popular among fans and draw in larger crowds. The way we decided to measure this is by looking at the top 50 jersey sales from each year and documenting the number of players each team had in the top 50 jersey sales that year. This may help identify which teams have influential players whose fans may be more likely to pay money to watch them play *live*.

Geographical data is also of importance. It is possible that stadiums in warmer areas of the country will draw in more attendees than stadiums in the Northeast or other colder regions. Since the football season occurs during the fall and winter, stadiums in colder regions could be at a disadvantage, since they encounter very cold weather and even snowstorms, which could lower attendance numbers. It is also possible that football is historically a larger cultural phenomenon in some regions of the country as opposed to others, so this needs to be investigated. For this research, the country was divided into 5 regions: Northeast, Southeast, Midwest, Southwest, and West. These regions, at least to a moderate extent, encompass somewhat similar climatic and cultural trends.

Finally, a last factor that needed to be added to this analysis is the financial aspect of attending an NFL game. Average ticket price at each stadium each year will be included and analyzed to determine whether ticket price impacts attendance levels (NFL Tickets: Team by Team Average Prices, 2016). We might see a pattern that higher ticket price is associated with lower attendance, or *just the opposite*².

² This issue provides a good example of why we have avoided the idea of a cause-and-effect relationship. Perhaps we have the traditional relationship of a higher price associated with a lower demand. However, it is also possible that higher demand for tickets leads a team to charge higher prices.

Data, Analysis, and Discussion of Results

These are the average attendance percentage results for every NFL team over the past three seasons (in order of decreasing average over the 3-year period) based on recorded statistics from ESPN (2016):

Team	2013	2014	2015	3-year average
Green Bay Packers	106.9	107.1	107.2	107.1
Dallas Cowboys	110.1	112.6	91.5	104.7
Indianapolis Colts	104.7	103.8	104.8	104.4
Philadelphia Eagles	102.3	103	102.8	102.7
Seattle Seahawks	101.8	102.1	103	102.3
San Francisco 49ers	99.3	103.3	103.4	102.0
Minnesota Vikings	99.8	104.5	99.8	101.4
Denver Broncos	101	101.1	101	101.0
Houston Texans	100.9	101	101	101.0
Chicago Bears	101.4	100.3	100.9	100.9
Baltimore Ravens	100.2	100.1	100	100.1
Carolina Panthers	99.5	99.8	100.4	99.9
New England Patriots	100	100	97.2	99.1
Atlanta Falcons	98.6	99	98.8	98.8
New Orleans Saints	99.9	100.2	95.5	98.5
Arizona Cardinals	96.3	97.8	98.7	97.6
Kansas City Chiefs	98.2	97.7	96.5	97.5
Detroit Lions	98.9	97.7	94.4	97.0
Tennessee Titans	100	100	90.1	96.7
New York Giants	97.1	95.7	95.8	96.2
Cincinnati Bengals	96.6	92.7	93.7	94.3
New York Jets	93.3	94.7	94.7	94.2
Pittsburgh Steelers	88.2	95.7	98.3	94.1
Miami Dolphins	85.5	92.7	102.9	93.7
Buffalo Bills	90.7	93.8	95.6	93.4
Cleveland Browns	97.3	92.1	90.4	93.3
San Diego Chargers	90.1	91.8	94.6	92.2
Jacksonville Jaguars	89.2	93.7	91.5	91.5
Tampa Bay Buccaneers	89.6	90.9	93.5	91.3
Washington Redskins	84.2	85	89.7	86.3
St. Louis Rams	87.2	87.3	80.2	84.9
Oakland Raiders	80	85.2	86.5	83.9

Before any analyses are performed using this data, it is immediately noticeable that the St. Louis Rams, who were relocated for the 2016 season, had the second lowest attendance over the past three years. The team with the lowest attendance in this time span has been the Oakland Raiders, who, as noted earlier, are headed for Las Vegas. The other team in serious consideration for being moved to a different city has been the San Diego Chargers - possibly moving to Los Angeles - who are 27th on this list of the 32 NFL teams sorted by attendance.

The attendance data alone tells us that the teams the NFL has been considering for relocation (primarily, St. Louis, Oakland, San Diego) have been (relatively speaking!) struggling to sell tickets and fill their stadiums for home games.

Table 1 presents some key descriptive statistics for the attendance percentages for each year:

<u>2013</u>		<u>2014</u>	
Mean	96.5	Mean	97.6
Median	98.8	Median	98.4
Standard Deviation	6.9	Standard Deviation	6.1
Minimum	80.0	Minimum	85.0
Maximum	110.1	Maximum	112.6
<u>2015</u>			
	Mean		96.7
	Mode		91.5
	Standard Deviation		5.8
	Minimum		80.2
	Maximum		107.2

Table1: Some descriptive statistics for team attendance

The mean attendance percentages don't vary by much, but out of these three years it was the lowest in 2013 (96.5%), closely followed by 2015 (96.7%), and the highest in 2014 (97.6%). Since all these means are very similar, it suggests a lack of any obvious drastic changes or trends in average attendance percentage in the NFL in recent years. The maximum and minimum were also highest in 2014, while 2013 had the lowest minimum and 2015 had the lowest maximum attendance percentage for an individual team. In all three years, the median attendance was slightly higher than the mean, suggesting that some of the teams with the lower attendance values dragged down the mean, suggesting a minor degree of left skewness to the distribution of attendance figures.

Univariate Analyses

One of the relationships we wished to investigate is whether location appears to have an effect on a team's attendance. Table 2 contains the average attendance percentages by region of the country.

<u>Region</u>	<u>Average Attendance (2013-2015)</u>	<u># of Teams</u>
Midwest	97.9%	9
Northeast	95.8%	8
Southeast	95.8%	7
Southwest	101.1%	3
West	96.3%	5

Table 2: Attendance by region of the country

From these data alone, it's not entirely clear why Los Angeles was chosen as the new destination for the Rams. They were moved from the Midwestern region, which had the second highest attendance percentage in the past three years, to the West, which had the third highest attendance percentage. One possibility is that Los Angeles is predicted to be a successful market based on high city-population alone.

Another is that they were taken out of the Midwest, the region with the most NFL teams, and brought to the West, the region with the second lowest number of NFL teams, in order to try to even the distribution of teams across the country. One interesting conclusion is that the 3 NFL teams in the Southwestern region have very highly successful attendance rates. This could potentially be because there are fewer NFL teams in this region, or could be related to a number of other factors, such as the consistently warm weather or the huge cultural impact of football in the Southwest. Regardless, it is interesting that the NFL hasn't seriously considered relocating a team to a city within this region based on its current success.

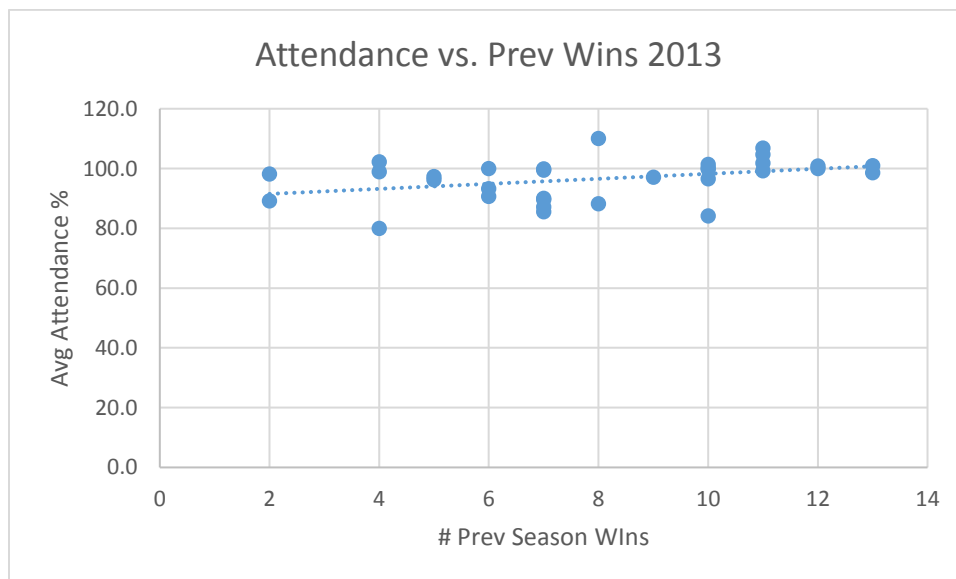
Interestingly, there is no clear relationship between the 3-year attendance average percentage and stadium capacity for NFL teams. The regression results for these two variables are shown in Table 3:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	83.40	10.73	7.77	1.13E-08
Stadium Capacity	0.00019	0.00015	1.27	0.215

Table 3: Linear Regression results with Y = average attendance %, X = stadium capacity

The p-value of 0.215 does not allow us (at traditional significance level of .05, or, perhaps, .10) to conclude that there is a linear relationship between attendance percentage and stadium capacity. The sign of the coefficient is positive, indicating that a larger stadium capacity corresponds with a larger attendance percentage³, but the magnitude of the coefficient (.00019) is too close to zero to provide convincing evidence that the true coefficient is really non-zero; thus, we conclude there is no linear relation between the two variables; a look at the scatter diagram (not provided here) would also clearly indicate the lack of a non-linear relationship.

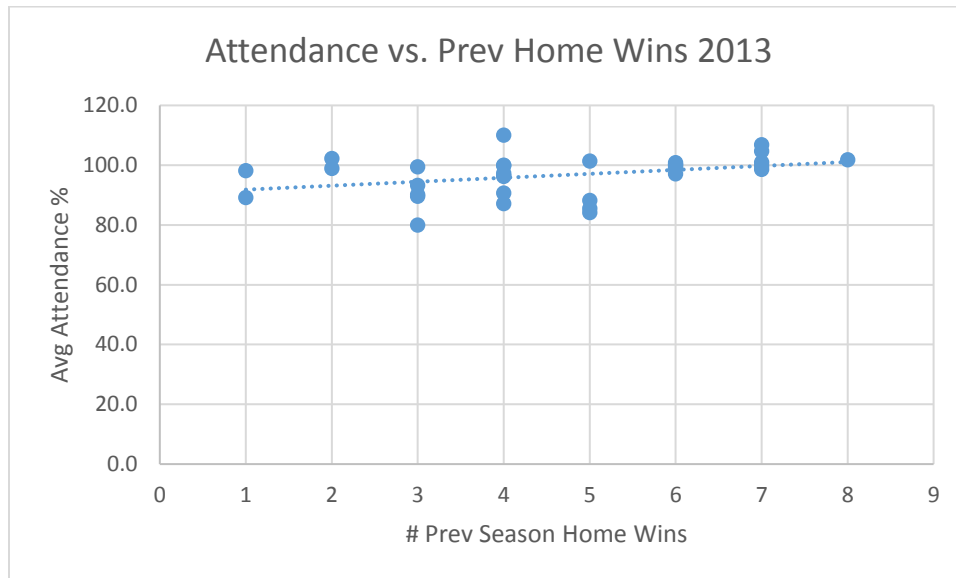
Our next step in the analysis was to search for a relationship between each team's home game attendance and their performance. Relating each team's attendance to their wins for the current season and the previous season (for both *total wins* and *home wins only*), the results are below for each of the three seasons. Each analysis includes a graphical depiction of the relationship with a linear trend line, the regression results including the p-value which determines significance, and both the R-squared and adjusted R-squared values which measure how much of the variation in attendance is accounted for by each model. Tables 4a-d considers a) attendance vs. total wins the previous year, b) attendance vs. home wins the previous year, c) attendance vs. total wins in the current year, and d) attendance vs. home wins in the current year. Tables 5a-d considers the same analyses for 2014, and Tables 6a-d the same analyses for 2015. We then discuss these results.



³ Certainly, this is not an impossible direction for such a relationship, if we acknowledge that the decision on a stadium's capacity could be based on demand for seating.

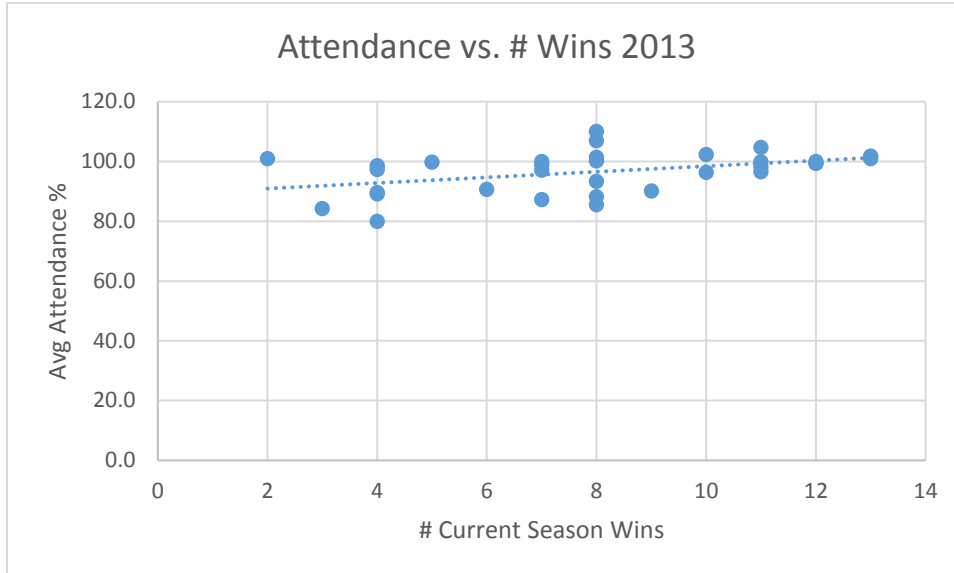
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	89.8	3.2	27.98	4.73E-23
# Wins Prev. Season	0.84	0.38	2.23	0.033
	R Square	0.142		
	Adjusted R Square	0.114		

Table 4a: Attendance vs. Total Wins from Previous Season, 2013



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	90.5	3.1	29.05	1.60E-23
# Prev Season Wins	1.32	0.63	2.08	0.045
	R Square	0.126		
	Adjusted R Square	0.097		

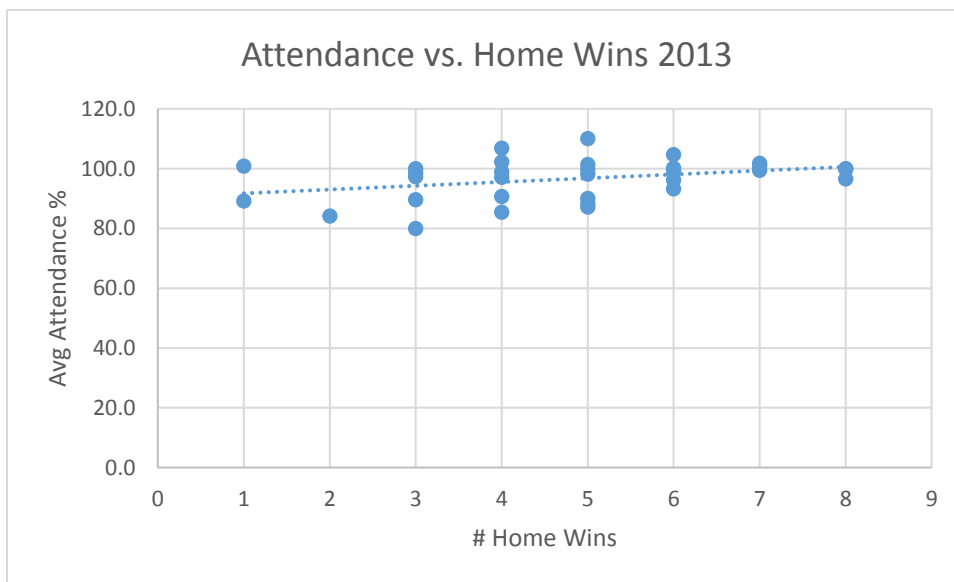
Table 4b: Attendance vs. Home Wins from Previous Season, 2013



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	89.1	3.1	28.62	2.43E-23
# Wins	0.94	0.36	2.57	0.015

R Square 0.181
Adjusted R Square 0.153

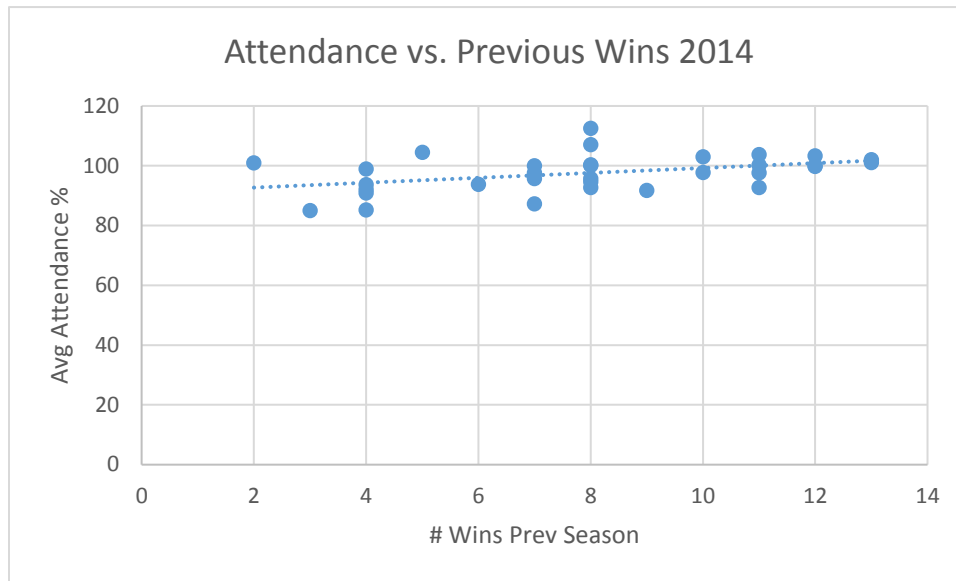
Table 4c: Attendance vs. Total Wins Current Season, 2013



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	90.48995147	3.211180698	28.17965103	3.84131E-23
# Home Wins	1.262232372	0.626385558	2.015104525	0.052931073

R Square 0.119
Adjusted R Square 0.090

Table 4d: Attendance vs. Home Wins Current Season, 2013

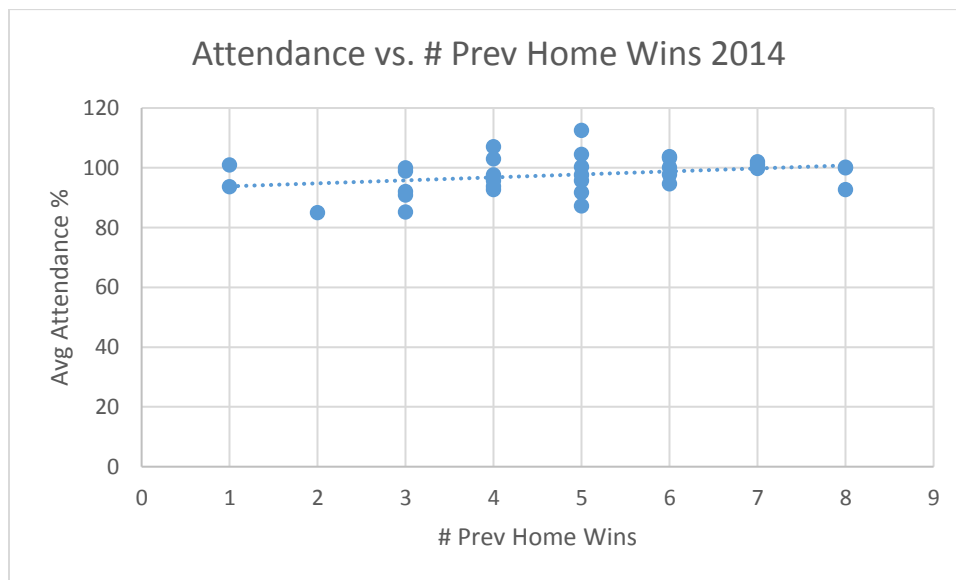


	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	91.1	2.8	32.62	5.45E-25
# Wins Prev. Season	0.82	0.33	2.50	0.018

R Square 0.173

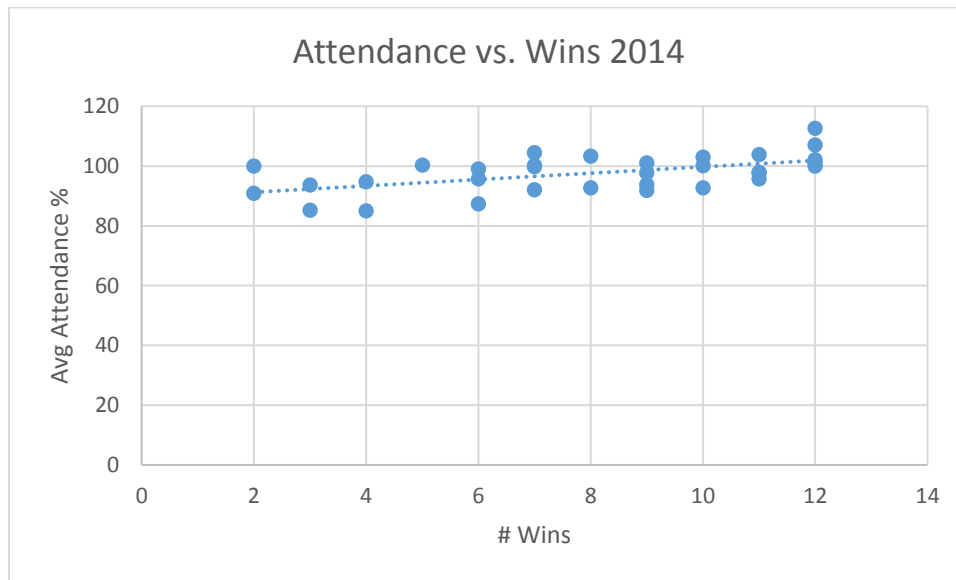
Adjusted R Square 0.145

Table 5a: Attendance vs. Total Wins from Previous Season, 2014



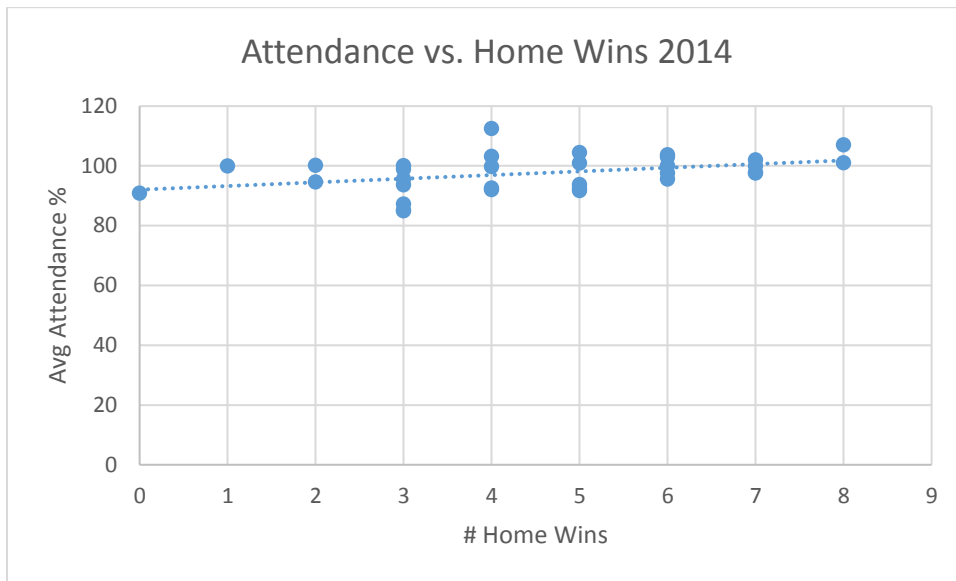
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	92.8	2.9	31.90	1.04E-24
# Prev Home Wins	1.00	0.57	1.76	0.089
R Square			0.093	
Adjusted R Square			0.063	

Table 5b: Attendance vs. Home Wins from Previous Season, 2014



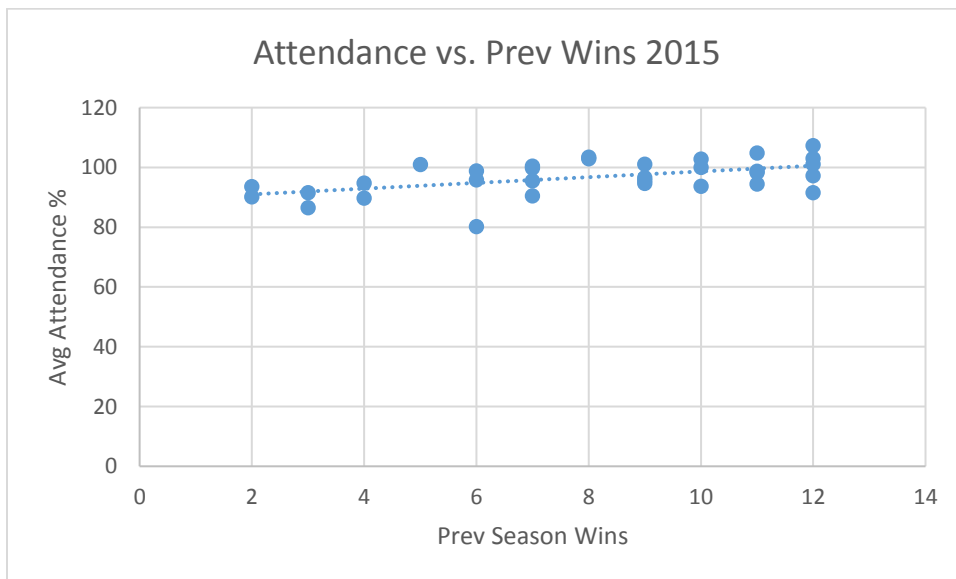
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	89.1	2.5	35.21	5.84E-26
# Wins	1.06	0.30	3.58	0.001
R Square			0.300	
Adjusted R Square			0.277	

Table 5c: Attendance vs. Total Wins Current Season, 2014



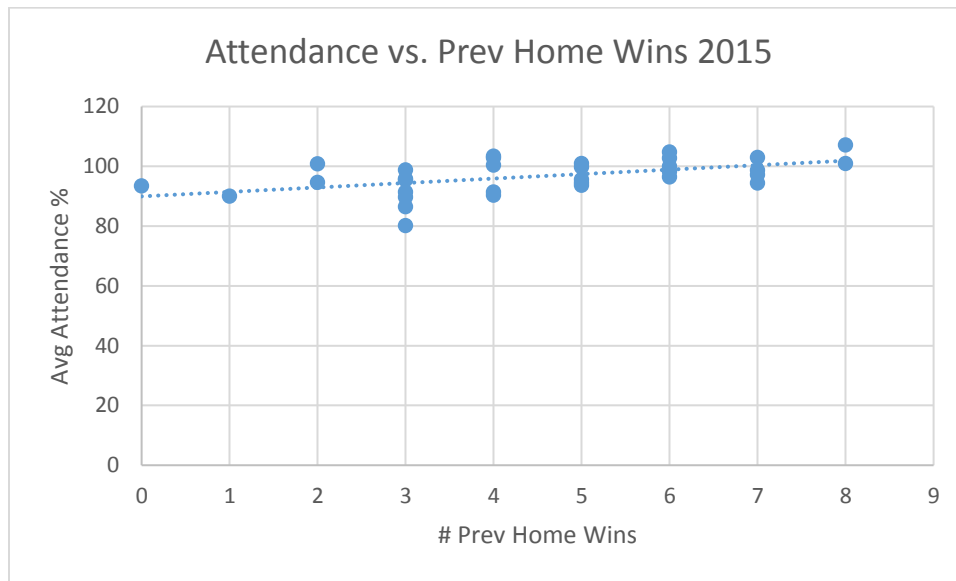
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	92.0	2.5	36.27	2.44E-26
# Home Wins	1.22	0.51	2.38	0.024
R Square		0.159		
Adjusted R Square		0.131		

Table 5d: Attendance vs. Home Wins Current Season, 2014



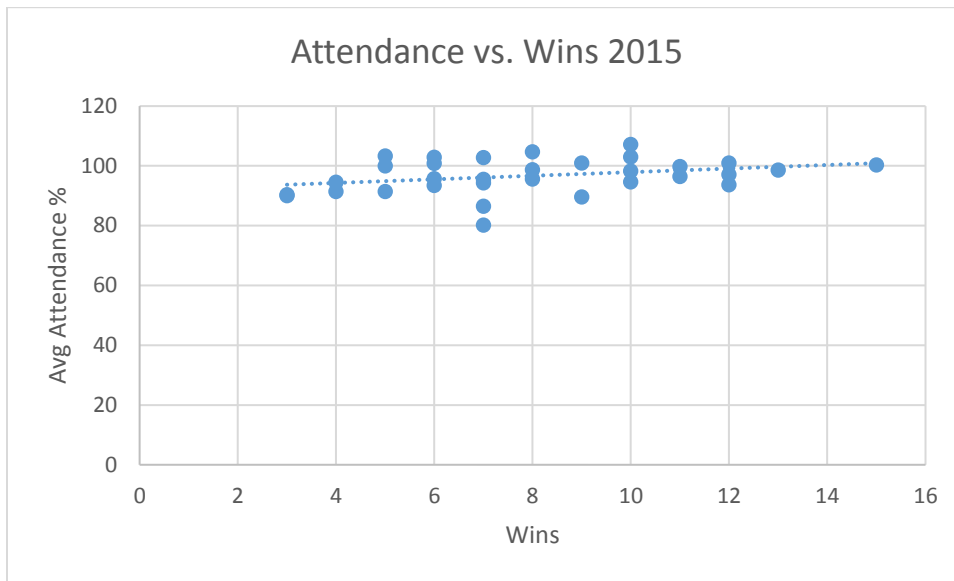
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	89.1	2.41	36.72	1.70E-26
Prev Season Wins	0.96	0.28	3.38	0.002
R Square			0.276	
Adjusted R Square			0.252	

Table 6a: Attendance vs. Total Wins from Previous Season, 2015



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	89.9	2.2	40.28	1.11E-27
Prev Season Home Wins	1.49	0.45	3.30	0.003
R Square			0.266	
Adjusted R Square			0.242	

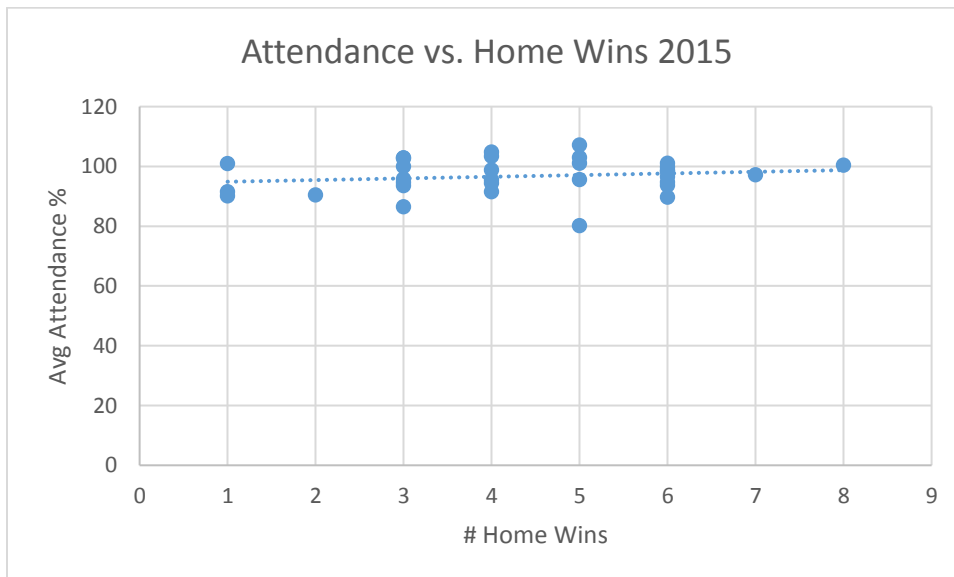
Table 6b: Attendance vs. Home Wins from Previous Season, 2015



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	91.9	2.8	32.76	4.82E-25
# Wins	0.60	0.33	1.83	0.078

R Square 0.100
Adjusted R Square 0.070

Table 6c: Attendance vs. Total Wins Current Season, 2015



	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	94.3	2.7	34.48	1.08E-25
# Home Wins	0.55	0.59	0.94	0.354

R Square 0.029
Adjusted R Square -0.004

Table 6d: Attendance vs. Home Wins Current Season

There are several important conclusions that can be made from these analyses. First of all, the relationship between attendance percentage and each different grouping of wins (whether total wins or home wins, from current season or previous, for each of the three years) is statistically significant (at 0.10 level of significance) in every scenario except for current season home wins in 2015, with 8 of the 11 significant results also significant at .05 significance level. This indicates that number of wins (home or total, for the current or previous year) is usually a useful predictor of that team's attendance percentage. Also, for every scenario, the relationship between attendance percentage and that team's wins was positive, meaning that the more a team wins, the higher attendance percentage they will have at their home stadium on average.

The most significant relationship of any scenario is between attendance percentage and current season wins in 2014, with a p-value of 0.001 and an R squared value of 0.30. The low p-value suggests a highly significant relationship between attendance percentage and wins that year. If we were to conduct a hypothesis test to see if wins impacts attendance percentage, with

H_0 : coefficient for wins = 0

H_1 : coefficient for wins \neq 0

we would look to the p-value for the (slope) coefficient for wins, to determine if the relationship is significant at a given level of significance. Since the p-value in this case is 0.001, lower than all commonly used levels of significance, we are able to reject H_0 and conclude that there is a significant linear relationship between wins and attendance percentage. The R squared value of 0.30 in this scenario indicates that (we estimate that) 30% of the variation in attendance percentage is accounted for by the model between attendance percentage and number of wins in 2014, while the rest is due to other factors (generally labeled "error".) This is the highest R squared value of the 12 regression analyses.

Multivariate Analyses

Multiple-regression analysis

A step that pulls all variables into the analysis simultaneously is to run a multiple regression for each year. Since one of the variables collected, region of the country, is a qualitative/nominal variable, it is necessary to create dummy variables in order to properly perform the analysis. For the five categories, four dummy variables were created:

Northeast: 1 if team is located in the Northeast, 0 otherwise

Southeast: 1 if team is located in the Southeast, 0 otherwise

Midwest: 1 if team is located in the Midwest, 0 otherwise

Southwest: 1 if team is located in the Southwest, 0 otherwise

While the West location does not receive its own dummy variable, (i.e., it is the "dummy" or "base" category), it is accounted for through the intercept.

Tables 7a (2013), 7b (2014), and 7c (2015) present the multiple regression results using all variables for each year to attempt to predict average stadium attendance percentage:

<i>Regression Statistics</i>	
Multiple R	0.816
R Square	0.666
Adjusted R Square	0.483
Standard Error	4.940
Observations	32

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	11	974.9	88.6	3.63	0.006
Residual	20	488.0	24.4		
Total	31	1462.9			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	70.96	10.89	6.51	2.38E-06	48.23	93.69
Stadium Capacity	0.0001	0.00015	0.66	0.516	-0.0002	0.0004
# wins	1.24	0.69	1.79	0.089	-0.21	2.69
# home wins	-0.81	1.10	-0.74	0.468	-3.11	1.48
# wins prev. season	-0.12	0.78	-0.15	0.879	-1.75	1.51
# home wins prev. season	0.75	1.13	0.67	0.5119	-1.60	3.10
Avg. ticket price	-0.004	0.017	-0.25	0.8079	-0.04	0.03
# players in top 50 jersey sales	2.46	1.33	1.85	0.0799	-0.31	5.22
NE	5.73	3.69	1.55	0.136	-1.96	13.42
SE	6.99	3.40	2.06	0.053	-0.10	14.09
MW	10.15	3.07	3.31	0.003	3.75	16.54
SW	10.94	3.94	2.78	0.012	2.72	19.16

Regression equation:

$$\text{Avg Attendance\%} = 70.96 - 0.0001*\text{StadiumCapacity} + 1.22*\text{Wins} - 0.81*\text{HomeWins} - 0.12*\text{PrevWins} + 0.75*\text{PrevHomeWins} - 0.004*\text{TicketPrice} + 2.46*\text{PlayersTop50} + 5.73*\text{NE} + 6.99*\text{SE} + 10.15*\text{MW} + 10.94*\text{SW}$$

Table 7a: Multiple regression analysis for 2013

<i>Regression Statistics</i>	
Multiple R	0.75
R Square	0.57
Adjusted R Square	0.33
Standard Error	5.01
Observations	32

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	11	663.4	60.3	2.40	0.043
Residual	20	502.9	25.1		
Total	31	1166.3			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	77.68	11.821	6.57	2.09E-06	53.04	102.33
Stadium Capacity	8.37E-05	0.00016	0.53	0.602	-0.00024	0.00041
# wins	0.53	0.97	0.55	0.589	-1.49	2.59
# home wins	-0.17	1.30	-0.13	0.899	-2.88	2.55
# wins prev season	0.58	0.73	0.80	0.435	-0.95	2.12
# home wins prev season	-0.59	1.07	-0.55	0.589	-2.82	1.641
avg ticket price	0.012	0.02	0.58	0.567	-0.03	0.055
# players in top 50 jersey sales	1.38	1.20	1.15	0.262	-1.12	3.87
NE	0.80	3.43	0.23	0.818	-6.36	7.97
SE	5.76	3.62	1.59	0.127	-1.79	13.31
MW	4.34	3.12	1.39	0.180	-2.17	10.86
SW	7.84	4.73	1.66	0.113	-2.02	17.69

Regression equation:

$$\text{Avg Attendance\%} = 77.68 + 0.00008 * \text{StadiumCapacity} + 0.53 * \text{Wins} - 0.17 * \text{HomeWins} + 0.58 * \text{PrevWins} - 0.59 * \text{PrevHomeWins} - 0.01 * \text{TicketPrice} + 1.38 * \text{PlayersTop50} + 0.80 * \text{NE} + 5.76 * \text{SE} + 4.34 * \text{MW} + 7.84 * \text{SW}$$

Table 7b: Multiple regression analysis for 2014

<i>Regression Statistics</i>					
Multiple R					0.72
R Square					0.512
Adjusted R Square					0.242
Standard Error					5.03
Observations					32

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	11	530.1	48.2	1.91	0.101
Residual	20	505.1	25.3		
Total	31	1035.2			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	76.84	10.66	7.21	5.60E-07	54.61	99.08
Stadium Capacity	9.12E-05	0.00016	0.56	0.580	-0.00024	0.00044
# wins	1.21	0.80	1.37	0.186	-0.64	3.06
# home wins	-1.94	1.52	-1.28	0.215	-5.11	1.22
# wins prev season	0.39	0.839	0.47	0.643	-1.34	2.12
# home wins prev season	1.25	1.32	0.95	0.354	-1.50	4.01
avg ticket price	0.026	0.02	1.21	0.241	-0.02	0.07
# players in top 50 jersey sales	-1.20	1.20	-1.00	0.330	-3.714	1.31
NE	-0.80	3.48	-0.23	0.819	-8.06	6.45
SE	3.02	3.88	0.78	0.446	-5.08	11.13
MW	-1.34	3.57	-0.38	0.7118	-8.79	6.11
SW	-4.08	4.74	-0.86	0.400	-13.97	5.81

Regression equation:

$$\text{Avg Attendance\%} = 76.84 + 0.00009 * \text{StadiumCapacity} + 1.21 * \text{Wins} - 1.94 * \text{HomeWins} + 0.39 * \text{PrevWins} + 1.25 * \text{PrevHomeWins} + 0.03 * \text{TicketPrice} - 1.20 * \text{PlayersTop50} - 0.80 * \text{NE} + 3.02 * \text{SE} - 1.34 * \text{MW} - 4.08 * \text{SW}$$

Table 7c: Multiple regression analysis for 2015

From this, we can first deduce that this group of variables is significantly predictive of attendance percentages for 2013 and 2014 using a standard level of significance 0.05; the two F-statistics have p-values below 0.05. However, for 2015, the p-value for the model is slightly above 0.10 ($p = .101$), so we are unable to conclude at that level of significance that these variables are predictive of attendance, but this p-value is still fairly low. A reassuring sign is that the R squared values for each model are somewhat high, suggesting a moderate amount of variability in attendance is explained by the overall model: 66.64% for 2013, 56.88% for 2014, and 51.21% for 2015.

For each model, the following variables had a p-value below .10 (all with positive coefficients); the variable in bold were significant at $p < .05$:

2013

wins (current season)

players in top 50 jersey sales

Southeast

Midwest

Southwest

2014

None, but $p < .20$ for Southeast, Midwest, and Southwest

2015

None

Stepwise-regression analyses

Another test that examines the significance of variables in a multivariate setting is a *stepwise regression*. Stepwise regression analysis will allow us to identify significant variables (if any) that did not show up in the "regular" multiple-regression analysis due to multi-collinearity among the independent variables. Tables 8a, 8b, and 8c present the results for years 2013, 2014, and 2015, respectively.

Model Summary

Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate
1	.596 ^a	.355	.334		5.60646
2	.665 ^b	.443	.404		5.30229

a Predictors: (Constant),
PlayersInTop50JerseySales

b Predictors: (Constant),
PlayersInTop50JerseySales,
Midwest

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	519.968	1	519.968	16.542	.000 ^b
	Residual	942.972	30	31.432		
	Total	1462.940	31			
2	Regression	647.625	2	323.812	11.518	.000 ^c
	Residual	815.315	29	28.114		
	Total	1462.940	31			

a Dependent Variable: AvgAttendancePct

b Predictors: (Constant), PlayersInTop50JerseySales

c Predictors: (Constant), PlayersInTop50JerseySales, Midwest

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	91.759	1.535		59.785	.000
	PlayersInTop50JerseySales	3.051	.750	.596	4.067	.000
2	(Constant)	90.244	1.616		55.836	.000
	PlayersInTop50JerseySales	3.216	.714	.628	4.506	.000
	Midwest	4.469	2.097	.297	2.131	.042

a Dependent Variable: AvgAttendancePct

Table 8a: Stepwise regression output for 2013

Model Summary

Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate
1	.574 ^a	.329	.307		5.10661

a Predictors: (Constant), PlayersTop50

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	384.036	1	384.036	14.727	.001 ^b
	Residual	782.324	30	26.077		
	Total	1166.360	31			

a Dependent Variable: AvgAttendance

b Predictors: (Constant), PlayersTop50

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	93.682	1.358		68.991	.000
	PlayersTop50	2.491	.649	.574	3.838	.001

a Dependent Variable: AvgAttendance

Table 8b: Stepwise regression output for 2014

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.525a	.276	.252	4.99869

a Predictors: (Constant), PrevWins

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	285.572	1	285.572	11.429	.002b
	Residual	749.608	30	24.987		
	Total	1035.180	31			

a Dependent Variable: AvgAttendance

b Predictors: (Constant), PrevWins

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	89.064	2.426		36.719	.000
	PrevWins	.958	.283	.525	3.381	.002

a Dependent Variable: AvgAttendance

Table 8c: Stepwise regression output for 2015

The results show that the 2013 data contains two variables that are significant at $p < .05$ in predicting average stadium attendance - number of players in the top 50 jersey sales and the Midwest dummy variable. This is somewhat consistent with the original multiple regression model since these variables were among those with the lowest p values and therefore most significant in predicting stadium attendance. The Southwest dummy variable was no longer significant at $p < .05$. We can conclude that the variable, PlayersTop50JerseySales, was overlapping with other variables in the regular multiple regression, but got revealed as significant in the stepwise analysis. Since the coefficients are positive, it indicates that more players in the top 50 in jersey sales for a team increases their stadium attendance on average, and so does being located in the Midwest, in each case, holding the other variable constant. The actual equation is

$$\text{Predicted Average Attendance percent} = 90.244 + 3.216 * \text{Players InTop50JerseySales} + 4.468 * (\text{Midwest } [0,1] \text{ dummy variable})$$

In 2014, the only independent variable deemed a significant predictor of average stadium attendance percentage at $p < .05$ was players in the top 50 in jersey sales. The regular multiple regression did not reveal any significant variables at $p < .10$. Again, we can conclude that the variable, PlayersTop50 Jersey Sales, was overlapping with other variables in the regular multiple regression, but got revealed as significant in the stepwise analysis. Since the coefficient is positive (as it was for 2013), this means on average that, the more players a team had in the top 50 in NFL jersey sales, the higher their average attendance would be predicted to be.

The fact that players in the top 50 in jersey sales was a significant predictor of stadium attendance two years in a row indicates that teams with popular or “star” players draw more people in, and teams looking to gain larger crowds may benefit by adding very popular individuals to their team. This fact also adds some consistency to the results over time.

Lastly, the only independent variable that was a significant predictor of average NFL stadium attendance in 2015 at $p < .05$ was the number of wins the previous year; neither this, nor any other variable showed up as significant in the regular multiple regression. So, it is number of wins the previous year that was overlapping with other variables in the regular multiple regression, but got revealed as significant in the stepwise analysis. Since the corresponding coefficient is positive, this means that the more wins a team had the previous year, the higher the stadium attendance percentage was on average predicted to be. This makes sense, since fans are more likely to be drawn to stadiums for teams that had recently proven themselves to be successful.

The overall model is much more significant for the 2013 season, then becomes less significant in 2014 and less again in 2015 (i.e., the R-Square value decreased each year, while still remaining highly significant.) Still, overall, the results make intuitive sense. Fans typically will be more drawn to watch teams with more wins the previous year, while teams who win less may not be as fun to watch or draw large crowds. Additionally, the more players on each team in the top 50 in jersey sales that year, the more fans will often believe that these players, as individuals, are worth seeing in person. For the dummy variables for region of the country, it appears as though the northeast region may not be as consistently popular for watching games as the others, which, as we noted earlier, may make sense based on the climate, since it typically becomes the coldest NFL region in the winter, which may cause fewer fans to attend games.

Limitations and Directions for Future Research

It was difficult to document stadium attendance correctly. We found reliable sources online that gave stadium capacity numbers for each NFL stadium, but many of the recorded average attendance numbers over the last three years exceeded the given stadium capacity. Several of the average attendance percentages are documented as over 100% which further supports the idea that several stadiums sell tickets past capacity. Although sources state that capacity was able to be exceeded, it’s not very clear how much capacity can be expanded by at each stadium, and which games are truly “sell-outs.” The Dallas Cowboys, for example, have had attendance numbers far exceeding the capacity for their stadium in every game in 2013, but some games exceeded capacity by a few thousand and some by over 10,000. There doesn’t seem to be a very reliable way to measure if this is accurate. It would be helpful to have resources that more clearly stated the true maximum capacity for each stadium if it can be expanded past their regular capacity. It would also greatly enhance accuracy if records were kept of the actual fan attendance for each game, rather than going solely by the number of tickets sold.

It may also help to have some measure of the cultural significance of football in different areas of the country. Going into this research, we had the preconceived idea that football is much more popular in the southwestern United States than anywhere else, and especially popular in Texas, and our data proved that to be true. However, there may be a better way to divide the NFL teams into regions, since we used only five major regions in the country, and states in each region may not share the same cultural values. However, it would not be appropriate to conduct an analysis by each state, since several states don’t have professional NFL teams, while others have only one and some have multiple, *and*, such an analysis would necessitate having too many independent variables, given the sample sizes we have. Still, a better way to measure football popularity across the country may result in more significant patterns and could tell us more specifically where the NFL is lacking popularity and where the NFL is most popular.

Lastly, one other variable we wanted to document and use in our analyses was social media following. Especially in recent years, social media has been very telling of which teams have the largest and most passionate fan bases, which might assist in predicting stadium attendance. However, we were unable to find data we believed to be suitable. This was especially true since there are so many forms of social media these days that it would be difficult to determine which application is most indicative of larger fan followings. For example, a team may rank #1 in Instagram followers, but may have a smaller Facebook following than another team. More specific documentation in this field could potentially be useful in this type of research going forward.

References

- Brown, Maury (2011), "How Sports Attendance Figures Speak Lies." *Forbes*. N.p., 25 May. Web. 10 Dec. 2016. <http://www.forbes.com/sites/sportsmoney/2011/05/25/how-sports-attendance-figures-speak-lies/#6f143d1e689c>
- Coates, Dennis; Humphreys, Brad (2007), " Ticket Prices, Concessions and Attendance at Professional Sporting Events," *International Journal of Sport Finance*, 2(3), 161-170.
- DeSchrive, Timothy (2002) "Determinants of Spectator Attendance at NCAA Division II Football Contests," *Journal of Sport Management*, 311-30. Web. 10 Dec. 2016. <http://www.fitnessforlife.org/AcuCustom/Sitename/Documents/DocumentItem/149.pdf>.
- DeSchrive, Timothy (2011), "If We Build It, Will They Come? Examining the Effect of Expansion Teams and Soccer-specific Stadiums on Major League Soccer Attendance," *Sport, Business and Management: An International Journal*, 6(2), 205-27. Web. <http://www.emeraldinsight.com/doi/abs/10.1108/SBM-05-2014-0025>
- McKnight, Oscar T., Ronald Paugh, Jamie Waltz, and Elena Kirkbride (2016), "Improving College Football A Endurance with Demographic Analysis: Understanding the Importance of Benchmarking," *Sport Marketing*. Web. 10 Dec. 2016. https://works.bepress.com/oscar_mcknight/60/download/
- "NFL Attendance." ESPN. N.p., n.d. Web. 10 Dec. 2016. http://www.espn.com/nfl/attendance/_/year/2015/sort/homePct
- "NFL Tickets: Team by Team Average Prices." TicketIQ. N.p., n.d. Web. 10 Dec. 2016. <http://blog.tiqiq.com/2013/08/2013-nfl-tickets-team-team-average-prices/#comments>
- Stadiums of Pro Football. N.p., n.d. Web. 10 Dec. 2016. <http://www.stadiumsofprofootball.com/stadiums>.
- "Weekly League Attendance." Pro Football Reference. N.p., n.d. Web. 10 Dec. 2016. <http://www.pro-football-reference.com/years/2015/attendance.htm>.