Zachary Campbell

Development of a logistic regression model to predict the outcome of NBA games

Introduction

The goal of this logistic regression project was to analyze a dataset of every NBA game from 2014 to 2018 in order to build a model that best predicts if a team will win or lose a game based on a variety of common statistics recorded at all basketball games. For the sake of limiting the number of variables in the model as well as creating a predictive model that is solely reliant on the team of interest's statistic, the opponent's statistics were not considered during the process of building the best possible model.

As for the statistics that were considered, they included location (home or away), how many points were scored, field goals made, field goals attempted, field goal percentage, three pointers made, three pointers attempted, three point percentage, free throws made, free throws attempted, free throw percentage, offensive rebounds, total rebounds, assists, steals, blocks, turnovers, and total fouls. These 18 statistics give a holistic view of a game and cover as many aspects of basketball as possible as far as easily accessible stats go. The per game summary statistics for these predictors are as follows.

- Location had 4920 home games (2861 wins and 2059 losses) and 4920 road games (2059 wins and 2861 losses)
- Points scored had a mean of 103.65 with a standard deviation of 12.19
- Field goals made had a mean of 38.60 with a standard deviation of 5.03
- Field goals attempted had a mean of 84.90 with a standard deviation of 7.13
- Field goal percentage had a mean of 45.57% with a standard deviation of 5.47%

- Three pointers made had a mean of 9.13 with a standard deviation of 3.60
- Three pointers attempted had a mean of 25.62 with a standard deviation of 7.10
- Three point percentage had a mean of 35.43% with a standard deviation of 9.78%
- Free throws made had a mean of 17.32 with a standard deviation of 6.00
- Free throws attempted had a mean of 22.75 with a standard deviation of 7.39
- Free through percentage had a mean of 76.24% with a standard deviation of 10.43%
- Total rebounds had a mean of 43.52 with a standard deviation 6.41
- Offensive rebounds had a mean of 10.29 with a standard deviation of 3.81
- Assists had a mean of 22.55 with a standard deviation of 5.12
- Steals had a mean of 7.75 with a standard deviation of 2.96
- Blocks had a mean of 4.83 with a standard deviation of 2.54
- Turnovers had a mean of 13.64 with a standard deviation of 3.87
- Fouls had a mean of 20.06 with a standard deviation of 4.32

Table 1 provides a reference to all of these summary statistics.

With this many variables, there are bound to be several significant relationships between predictors and the response variable (outcome). Most likely there will be a strong positive relationship between field goals made, three pointers made, and free throws made and the outcome of winning a game. At the end of the day a team needs to score more points than their opponent in order to win, so while there will be other positive relationships with the response, I would be surprised if any are as strong as the “shot made” predictors. Also, there should be a negative relationship between fouls and turnovers with the response of outcome, simply because the more of each that a team has, the less likely they are to win. It will be interesting to find out if rebounds, assists, steals, and blocks will have a significant positive relationship with the response variable.

Methods

The dataset was published on a dataset website called data.world and contained 9840 observations of 40 variables (Kelepoouris). In order to verify that the dataset was accurate, a random sample of 25 observations from the potential 9840 were chosen. Each game was then looked up on the website Basketball Reference to verify that every single statistic in the dataset was accurate for that specific game. After each of the 25 observations were proven to be accurate, the dataset was further tested by checking to make sure the inverses of several games were accurate. For example, if the dataset had a game where the Denver Nuggets played the Miami Heat at home on March 31 and won 121-104, there should also be a game entry in the dataset where the Miami Heat play the Denver Nuggets on the road on March 31 and lost 104-121. These ten games random games, 20 total observations, were all also proven to be accurate (“Basketball Statistics and History). Therefore, the dataset appeared to be legitimate.

As for the procedures to help discover the best model fit, first the summary statistics were calculated for the league average of all the quantitative variables for the 4-season period of NBA games.

Then it needed to be determined if there was a significant difference between wins and losses for all the variables. Therefore, independent samples t-tests were performed on the 17 quantitative predictors. These tests assessed each predictor independently to make sure that they were significant, which would justify including them in the model.

For the variable location, a chi-squared test was run to see if there was a difference in the number of wins at home and the number of wins on the road. This test assessed if there were statistically significant differences than what would be expected for the four categories of wins at home, losses at home, wins on the road, and losses on the road. If there was not a significant difference between the four categories, the chi-squared test would find the location variable insignificant, but if there is a significant difference between the categories, the test would show that (Agresti).

After this, the issue of multicollinearity needed to be checked to make sure that there was not an issue of high correlations between predictors which would hurt the reliability of the logistic regression model (Agresti). To check this, two methods were used, correlation matrices and variance inflation factors. The correlation matrices were used to shorten the dataset before model building and the variance inflation factors were used to assess the legitimacy of models and compare them after the model building process occured.

First, a correlation matrix between all the quantitative predictors was created to see if there were any correlations that were above the 0.80 mark would indicate severe multicollinearity issues (Agresti). If there was a predictor that had extremely high correlations with multiple other predictors, it was removed from the model. Also, if several predictors had multicollinearity issues, two different models were fit to see which terms were more significant to the overall model. More on the logic of these removals is detailed in results section, but for example, due to points scored having a high correlation with four of the other predictors it was removed from the model. After removals, the correlation matrix was run again. This procedure was repeated until there were no apparent issues of multicollinearity with the quantitative predictors. These correlation matrices were computed in SAS by setting the dataset of the project, using the *proc corr* function, and listing the predictors that needed to be assessed.

After there were no significant issues of multicollinearity, it came time for the model building portion of the project. Potential models for predicting the outcome of an NBA game were created from three different methods. The first method was to fit a model in R using the *glm()* function. An example of this function for the model with location, field goals made, and steals would look like this:

```
logitex <- glm(outcome ~ location + fg + steals, data = data2, family = "binomial")
```

when the summary of the logistic regression models were obtained through the *summary()* function, each predictor's p-value was evaluated at a 0.05 level. If it was greater than 0.05, it was dropped from the model. If the p-value was less than 0.05, the predictor was kept in the model. When all the insignificant predictors were identified and removed, a new model was fit once again with the *glm()* function and the summary was obtained to see the new estimates. This process continued until there were only significant terms in the model. Also, for this method, the start model was one with only the predictors from the correlation matrix and the location predictor because it alleviated the concerns of multicollinearity in the logistic regression model.

The second method of model selection was through both forwards and backwards stepwise selection. Forward selection starts with a model that only has the intercept fit and then proceeds to add the most significant terms, one at a time, until there are no more significant terms to add. Backwards selection starts with the full model that includes all the predictors and removes the least significant term, one at a time, until there are only significant terms remaining. The code that was run in R for these processes was from the Toronto website (Brunner).

```
fit.interceptonly<-glm(outcome ~ 1,family = "binomial")
fit.fullmodel<-glm(outcome ~ location + fg + fg_attempted + fg_percent +
        three + three_attempted + three_percent + freethrow +
        freethrow_attempted + freethrow_percent + totalreb + offreb + assists +
        steals + blocks + turnovers + fouls, family = "binomial")
backwards <- step(fit.fullmodel)
summary(backwards)
```

The final method was to determine a list of the 15 most relevant predictors and then the R program calculated the five best models based on the lowest AIC. This statistic stands for Akaike's Information Criterion and the lower the AIC score the better. However, there is no baseline for what a good AIC is. It is dependent on the number of predictors and the specifics of the given dataset (Brown). It is worth noting that the reason it was only the 15 most relevant predictors was because R had a restriction that allowed for a maximum of 15 predictors to be used in the function. The 15 that were chosen were as follows: location, assists, steals, blocks, turnovers, total rebounds, fouls, field goals made, field goals attempted, field goal percentage, three pointers made, three pointers attempted, three point percentage, free throws made, and free throws attempted. Free throw percentage was left out because the two percent difference between losses and wins is smaller than many of the other predictors' differences between wins and losses.

```
data2.pr<-data.frame(location, assists , steals , blocks , turnovers ,
        totalreb , fouls , fg, fg_attempted , fg_percent , three, three_attempted,
        three_percent , freethrow , freethrow_attempted, outcome)
AICbest<-bestglm(Xy=data2.pr, family = binomial, IC = "AIC",TopModels = 5,
        method = "exhaustive")
AICbest$BestModels
```

When the best model from each method was determined, each one was compared to each other through concordance, AIC, and VIF. Concordance is a statistic that measures the agreement between an observed response and a predicted response. The value ranges between 0 and 1 and the higher the value, the better the model fit (Agresti). As for both AIC and VIF, a lower value is desired. For AIC, the low value is subjective, and for VIF, a value below ten is considered good. This is because it means the model does not have multicollinearity issues. The code for concordance in R came from the website R Documentation (Therneau). The model with the combination of the highest concordance, lowest AIC, and lowest VIF was determined to be the best logistic regression model.

Results

None of the t-tests proved to be insignificant at a 5% significance level. In fact, every single p-value was ≤ 0.0012. The chi-squared test statistic was $\chi^2 = 261.405$ with an associated p-value < 0.0001. This means that there was a significant difference between the number of wins at home and the number of wins on the road. Each t-test and associated p-value can be found on Table 2. The chi-squared table and test can be found on Table 3.

Through both of these initial sets of tests, every single one of the 18 potential variables were found to be statistically significant at the 0.05 level, meaning there was a significant difference between a win and a loss for every single predictor. Therefore, after the first step, there was no justification for removing any of the predictors from the model.

From the correlation matrix found on Table 4, a correlation was found to be above |0.80|. This was the correlation between team points scored and field goals made. Furthermore, the correlation between team points and field goal percentage (r=0.71), assists (r=0.575), and three pointers made (r=0.518) were also relatively high. Since logically speaking, the amount of points that a team scores are virtually reliant on every predictor, it was decided to remove the team score predictor from the model.

After running another correlation matrix on the model without team points, which is found on Table 5, the correlations improved a decent amount but there were still a few notably high ones. The correlation between field goal percent and field goals made was 0.777, the correlation between three pointers attempted and three pointers made was 0.747, and the correlation between free throws attempted and free throws made was 0.924. Since each correlation involved a shooting metric (field goals, threes, and free throws), the "makes" predictor for each metric was removed from the model. The reasoning behind this was because by knowing how many shots were attempted as well as the percent of makes allowed for a quick calculation of makes. Also, percentage statistics were involved in the highest correlations of the shooting statistics. Finally, for interpretability, it was decided to have the same two metrics for each of the shooting categories rather than having one category have makes and percentage, another have percentage and attempts, for example.

The resulting correlation matrix, found on Table 6, looked much better. There were two correlation higher than 0.50, one between assists and field goal percentage with a correlation of 0.555 and the other between field goals attempted and offensive rebounds with a correlation of 0.51. However, almost all the remaining correlations were quite low (around 0.20 or lower). Therefore, this proved that the 13 predictors of assists, steals, blocks, turnovers, total rebounds, offensive rebounds, total fouls, field goals attempted, field goal percentage, three pointers attempted, three point percentage, free throws attempted, and free throw percentage were found to have no significant issues of multicollinearity.

The R output from the full model can be found on Figure 1. The estimated logistic regression function was:

$$\frac{e^{-9.68+0.39l+0.26fm-0.28fa+11.56fp+0.19tm-0.03ta+1.46tp+0.12ftm-0.06fta+1.99ftp+0.37r-0.02or+0.02a+0.40s+0.12b-0.33t-0.06f}}{1+e^{-9.68+0.39l+0.26fm-0.28fa+11.56fp+0.19tm-0.03ta+1.46tp+0.12ftm-0.06fta+1.99ftp+0.37r-0.02or+0.02a+0.40s+0.12b-0.33t-0.06f}}$$

* l = location, fm = field goals made, fa = field goals attempted, fp = field goal percentage, tm = three pointers made, ta = three pointers attempted, tp = three point percentage, ftm = free throws made, fta = free throws attempted, ftp = free throw percentage, r = total rebounds, or = offensive rebounds, a = assists, s = steals, b = blocks, t = turnovers, f = fouls

In order to assure that there actually was a logistic relation for this model, a test of logistic regression relation was run. The output is found on Figure 1 and the test is below

H_o: There is not a logistic regression relationship ($\beta_1 = \beta_2 = \ldots = \beta_{17} = 0$)
H_a: There is a logistic regression relationship (At least one $\beta_1, \beta_2, \ldots, \beta_{17} \neq 0$)
Test Statistic: $\chi^2 = 6947.80$
P-value ≈ 0
Conclusion: Since the p-value < 0.05, there was significant evidence to conclude that there was a logistic regression relationship for the full model at a 5% significance level.

This model served as the baseline for comparison for the rest of the models.

When the first method of model building was used, a logistic regression model was fit with the 13 predictors remaining after the correlation matrices and the location predictor. The R output from this model can be found on Figure 2. The estimated logistic regression function was:

$$\frac{e^{-23.08+0.38l-0.16fa+33.90fp+0.04ta+5.68tp+0.03fta+4.51ftp+0.37r-0.02or+0.03a+0.39s+0.12b-0.33t-0.06f}}{1+e^{-23.08+0.38l-0.16fa+33.90fp+0.04ta+5.68tp+0.03fta+4.51ftp+0.37r-0.02or+0.03a+0.39s+0.12b-0.33t-0.06f}}$$

However, one of the predictors was found to be statistically insignificant at the 0.05 significance level. This predictor was offensive rebounds, with a p-value of 0.121. Therefore, the predictor needed to be taken out of the model and then the model needed to be refitted. The resulting R output can be found on Figure 3 and the estimated logistic regression function was:

$$\frac{e^{-22.87+0.37l-0.17fa+34.09fp+0.04ta+5.66tp+0.03fta+4.53ftp+0.36r+0.03a+0.39s+0.12b-0.33t-0.06f}}{1+e^{-22.87+0.37l-0.17fa+34.09fp+0.04ta+5.66tp+0.03fta+4.53ftp+0.36r+0.03a+0.39s+0.12b-0.33t-0.06f}}$$

This newly fitted logistic regression model had no VIF scores above 3.5, and all the predictors were significant. This suggested that this was an adequate model to predict the outcome of an NBA game based off readily available statistics.

For backwards selection, the full model, which included all of the predictors, was the starting point. The first and only step removed three point percentage from the model. After this step, the program failed to remove any other predictors. The program's process can be seen on Figure 4. As for the logistic regression function, it was:

$$\frac{e^{-9.55+0.39l+0.26fm-0.28fa+12.44fp+0.24tm-0.05ta+0.12ftm-0.06fta+1.98ftp+0.37r-0.02or+0.02a+0.40s+0.12b-0.33t-0.06f}}{1+e^{-9.55+0.39l+0.26fm-0.28fa+12.44fp+0.24tm-0.05ta+0.12ftm-0.06fta+1.98ftp+0.37r-0.02or+0.02a+0.40s+0.12b-0.33t-0.06f}}$$

This model still had two insignificant terms at the 0.05 level: field goal percentage with a p-value of 0.109 and offensive rebounds with a p-value of 0.111. The model was refitted after it dropped both terms. The resulting logistic regression model summary is on Figure 5 and the function was:

$$\frac{e^{-3.68+0.38l+0.40fm-0.35fa+0.24tm-0.04ta+0.12ftm-0.06fta+2.05ftp+0.36r+0.03a+0.40s+0.12b-0.34t-0.06f}}{1+\ e^{-3.68+0.38l+0.40fm-0.35fa+0.24tm-0.04ta+0.12ftm-0.06fta+2.05ftp+0.36r+0.03a+0.40s+0.12b-0.34t-0.06f}}$$

As for forward selection, even though the process started with the intercept only model and added significant terms, resulting model at the conclusion of the code was the exact same as in backwards selection. The model had the same predictors with the same intercepts. This meant in forwards selection that both field goal percentage and offensive rebounds also were found to be insignificant and needed to be removed from the model. When they were both dropped, the resulting model and function was one again identical to the one found with backwards selection. The forward selection process and R output that proves this is found on Figure 6.

The final method of finding the best model for the prediction of outcomes of NBA games based on game statistics was the best AIC approach in R. Although briefly mentioned in the methods section, the reason for the 15 variables that were used to determine the best AIC was because of the correlation matrices and the fact that the two percent difference in free throw percentages between losses and wins was smaller than many of the other predictors’ differences between wins and losses.

With this criterion established, the best model according to AIC was one that included location, assists, steals, blocks, turnovers, total rebounds, fouls, field goals made, field goals attempted, field goal percent, three pointers made, three pointers attempted, free throws made and free throws attempted. When fitting this logistic regression model, based on the output from R found on Figure 7, the function was:

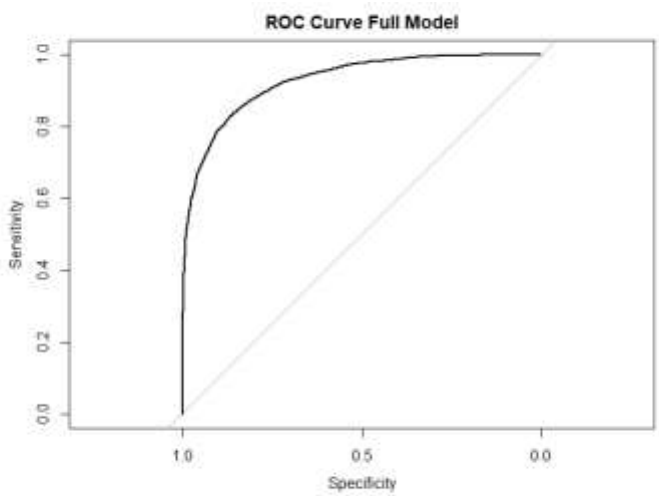
$$\frac{e^{-7.89+0.39l+0.25fm-0.28fa+12.83fp+0.24tm-0.04ta+0.21ftm-0.13fta+0.36r+0.02a+0.40s+0.12b-0.34t-0.06f}}{1+e^{-7.89+0.39l+0.25fm-0.28fa+12.83fp+0.24tm-0.04ta+0.21ftm-0.13fta+0.36r+0.02a+0.40s+0.12b-0.34t-0.06f}}$$

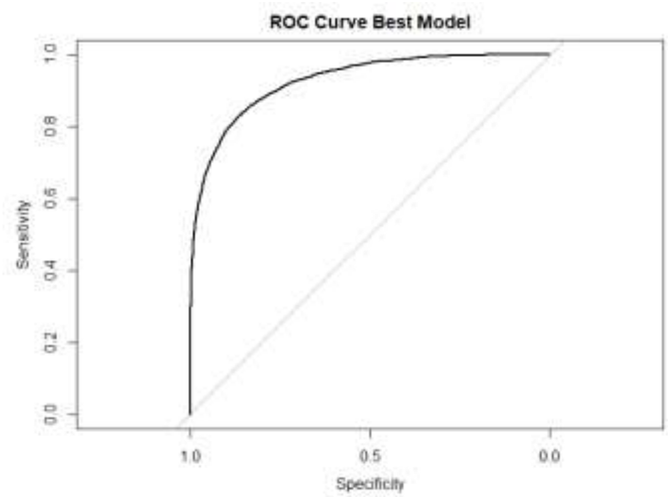
From each method there was a potential “best model” which was then tested with concordance, AIC, and VIF to determine what the actual best model was. Starting off with AIC, the full model had an AIC of 6729.3, the first method model had an AIC of 6733.4, the second method model had an AIC of 6730.0, and the third method model had an AIC of 6730.6. From these statistics, in terms of AIC, the best model is the full model.

As for concordance, the full model had a concordance of 0.9278. This meant 92.78% of the actual outcomes of games are in agreement with the predictions of games. The concordance of the model derived from the first method was 0.9270. The concordance for the second method was 0.9277. Finally, the concordance for the third method was 0.9276. Once again, this statistic suggest that the full model is the best model to predict the outcome of NBA games.

The final statistic of VIF was used to assess multicollinearity in the models. For the full model, there were three predictors with VIFs above 100 and two more predictors that had VIFs above 50, this indicated some serious multicollinearity issues. The first method model had every single VIF below 3.5; since all of these were below 10, there was not an issue of multicollinearity. The second method model had low VIFs except for free throws made and free throws attempted, which were 60.89 and 54.83, respectively. The third method model had three VIFs above 100.

Finally, to compare the adequacy of the best model, the ROC curves will be compared to show that the best model is better than the full model.





The ROC curves help explain concordance, as the more area that is under the curve, the better (Agresti). The two curves looked almost identical, and that is because the concordance of the two models were only separated by 0.08% ((.9278-.9270)*100). With the concordances being so similar and the concerns of multicollinearity being very prevalent in the full model and non-existent in the best model, there is little doubt the best model was the best at predicting the outcome of NBA games based on the predictors.

This “best model” included the predictors location, field goals attempted, field goal percentage, three pointers attempted, three point percentage, free throws attempted, free throw percentage, total rebounds, assists, steals, blocks, turnovers, and fouls. While the other models had slightly higher concordances and slightly lower AIC values, this could have been due to the multicollinearity. If the other models addressed their correlation issues, there is a good chance the final model from each method would look very similar to the selected best model. Therefore, the best model and corresponding logistic regression function was:

$$\frac{e^{-22.87+0.37l-0.17fa+34.09fp+0.04ta+5.66tp+0.03fta+4.53ftp+0.36r+0.03a+0.39s+0.12b-0.33t-0.06f}}{1+e^{-22.87+0.37l-0.17fa+34.09fp+0.04ta+5.66tp+0.03fta+4.53ftp+0.36r+0.03a+0.39s+0.12b-0.33t-0.06f}}$$

In order to assure that there actually was a logistic relation for the best model, a test of logistic regression relation was run. The output is found on Figure 3 and the test is below

H_o: There is not a logistic regression relationship ($\beta_1 = \beta_2 = \ldots = \beta_{13} = 0$)
H_a: There is a logistic regression relationship (At least one $\beta_1, \beta_2, \ldots, \beta_{13} \neq 0$)
Test Statistic: $\chi^2 = 6910.50$
P-value ≈ 0
Conclusion: Since the p-value < 0.05, there was significant evidence to conclude that there was a logistic regression relationship for the best model at a 5% significance level.

To help make the significance of this model better understood, the odds ratio of all the predictors was calculated by taking the exponential of each predictor's coefficient.

- $e^{0.374} = 1.454$ The estimated odds of winning a game at home are 1.45 times that of winning a game on the road while controlling for all the other predictors in the model
- $e^{-0.167} = 0.846$ The estimated odds of winning increase by 0.85 times (decrease by 15%) for every one unit increase in field goals attempted while controlling for all the other predictors in the model
- $e^{34.091} = 6.390\text{x}10^{14}$ The estimated odds of winning a game increase by $6.390\text{x}10^{14}$ times for every one percent increase in field goal percentage while controlling for all the other predictors in the model (this is very surprising)

- $e^{0.041} = 1.042$ The estimated odds of winning a game increase by 1.04 times for every one unit increase in three pointers attempted while controlling for all the other predictors in the model
- $e^{5.660} = 287.149$ The estimated odds of winning a game increase by 287.15 times for every one percent increase in three pointer percentage while controlling for all the other predictors in the model
- $e^{0.027} = 1.027$ The estimated odds of winning a game increase by 1.03 times for every one unit increase in free throws attempted while controlling for all the other predictors in the model
- $e^{4.533} = 93.037$ The estimated odds of winning a game increase by 93.04 times for every one percent increase in free throw percentage while controlling for all the other predictors in the model
- $e^{0.362} = 1.436$ The estimated odds of winning a game increase by 1.44 times for every one unit increase in total rebounds while controlling for all the other predictors in the model
- $e^{0.028} = 1.028$ The estimated odds of winning a game increase by 1.03 times for every one unit increase in assists while controlling for all the other predictors in the model
- $e^{0.392} = 1.480$ The estimated odds of winning a game increase by 1.48 times for every one unit increase in steals while controlling for all the other predictors in the model
- $e^{0.122} = 1.130$ The estimated odds of winning a game increase by 1.13 times for every one unit increase in blocks while controlling for all the other predictors in the model

- $e^{-0.335} = 0.715$ The estimated odds of winning a game increase by 0.72 times (decrease by 28%) for every one unit increase in turnovers while controlling for all the other predictors in the model
- $e^{-0.060} = 0.942$ The estimated odds of winning a game increase by 0.94 times (decrease by 6%) for every one unit increase in fouls while controlling for all the other predictors in the model

Also, to see the probability of winning a game with the league averages, the average from Table 1 were put into the function both for location = home and location = road.

Home:

$\hat{\pi}(l = 1, fa = 84.902, fp = 0.456, ta = 25.624, tp = 0.354, fta = 22.749, ftp = 0.762, r = 43.521, a = 22.547, s = 7.751, b = 4{,}820, t = 13.639, f = 20.059)$

= 1.396/(1+1.396) = 0.5826

Road:

$\hat{\pi}(l = 0, fa = 84.902, fp = 0.456, ta = 25.624, tp = 0.354, fta = 22.749, ftp = 0.762, r = 43.521, a = 22.547, s = 7.751, b = 4{,}820, t = 13.639, f = 20.059)$

= 0.9603/(1+0.9603) = 0.4899

Therefore, the estimated probability of winning a game when the team has league averages at home is 58.26% while the estimated probability of winning a game when the team has league averages on the road is 48.99%. By taking the ratio of these two percentages, the relative risk was 1.189. This meant the estimated risk of winning for home games is 1.189 times that of road games. The percent level of risk of winning for home games is 18.9% higher than for road games. This makes sense, because home teams are favored to win games because of the advantage that home courts offer, such as fans, familiarity, no travel, etc.

Discussion

Overall, all the significant finding seemed to make logical sense. The higher the percentages for field goals, threes, and free throws, the higher the chance a team has to win. However, the high estimates and odd ratios for these shooting percentages were very surprising, but since ever 1% increase is only 0.01 in the equation, it makes sense that a high estimate doesn't influence the model as much as it might seem. Furthermore, the more assists, rebounds, and steals a team gets, the higher their chance of winning. Similarly, the more turnovers and fouls a team commits, the less likely they are to win. However, the negative relationship between attempted field goals and the probability of winning was somewhat confusing, considering there was a positive relationship between three point attempts and the probability of winning as well as between free throw attempts and the probability of winning. Perhaps, a reason for this is that threes are worth more than twos, so taking more of them results in a higher amount of points scored per possession on average. Also, since free throws are rewarded by the other team committing fouls and they are essentially free shots with a high percentage of making them, it makes sense having more free throws attempted helps a team's probability to win.

The best model seemed to be doing a good job of predicting the outcomes of games based on the league averages. The 58.26% and 48.99% estimated probability of winning at home and on the road respectively while shooting league averages seem reasonable.

However, a limitation of the dataset was that while the it was large, having 9840 observations, it only spanned four NBA seasons. This might seem like a lot, but the league is currently undergoing a philosophical change of sorts. In the last two seasons the pace in the league is increasing, meaning virtually more of every single statistic for each game. Also, more three pointers are being attempted per game than before ("Basketball Statistics and History").

Therefore, these older 4 seasons that the project was based on might not put enough of a weight into three pointers and the lower averages of all the statistics might limit the reliability of the model going forward.

The dataset also did not include defensive rebounds. While this is seemingly minor, there could have been a better fit model that included defensive rebounds.

Future studies could improve the best model by including defensive rebounds, including more recent datasets, and testing to see if there is a better model that includes points scored. Two other potential variables that could be included to improve the model would be to include predictors such as which team was playing and who their opponent is. Finally, an interesting study would be to see if this model could extend to other levels of basketball by testing its validity at the college or even high school level.

Appendix

Table 1.

Predictor	**Mean**	**Standard Deviation**
Points Scored	103.6523	12.18804
Field Goals Made	38.60244	5.029992
Field Goals Attempted	84.90244	7.130043
Field Goal Percentage	45.56893%	5.471074%
Threes Made	9.126829	3.598084
Threes Attempted	25.62368	7.102501
Three Point Percentage	35.43212%	9.778227%
Free Throws Made	17.32063	6.00262
Free Throws Attempted	22.74939	7.39045
Free Throw Percentage	76.23949	10.4297%
Total Rebounds	43.52063	6.410428
Offensive Rebounds	10.2876	3.80712
Assists	22.54654	5.122989
Steals	7.750508	2.958886
Blocks	4.827642	2.536845
Turnovers	13.63862	3.869543
Fouls	20.05854	4.317611

Table 2.

Predictor	Difference (W-L)	t	P-value
Points Scored	11.16	51.10	< 0.0001
Field Goals Made	3.86	41.18	< 0.0001
Field Goals Att.	-0.59	-4.13	< 0.0001
Field Goal %	4.9%	49.56	< 0.0001
Threes Made	1.73	24.57	< 0.0001
Threes Att.	0.46	3.23	0.0012
Three Point %	6.2%	33.20	< 0.0001
Free Throws Made	1.72	14.34	< 0.0001
Free Throws Att.	1.67	11.31	< 0.0001
Free Throw %	2.2%	10.55	< 0.0001
Total Rebounds	3.58	26.70	< 0.0001
Offensive Rebounds	-0.15	-4.00	< 0.0001
Assists	3.19	32.54	< 0.0001
Steals	0.83	14.08	< 0.0001
Blocks	0.85	16.89	< 0.0001
Turnovers	-0.92	-11.93	< 0.0001
Fouls	-0.97	-11.26	< 0.0001

Table 3.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of location by outcome			
		outcome		
	location	0	1	Total
	0	2861 29.08 58.15 58.15	2059 20.92 41.85 41.85	4920 50.00
	1	2059 20.92 41.85 41.85	2861 29.08 58.15 58.15	4920 50.00
	Total	4920 50.00	4920 50.00	9840 100.00

Statistics for Table of location by outcome

Statistic	DF	Value	Prob
Chi-Square	1	261.4650	<.0001
Likelihood Ratio Chi-Square	1	262.6355	<.0001
Continuity Adj. Chi-Square	1	260.8134	<.0001
Mantel-Haenszel Chi-Square	1	261.4385	<.0001
Phi Coefficient		0.1630	
Contingency Coefficient		0.1609	
Cramer's V		0.1630	

*in this table for location 0 = home and 1 = road
for outcome 0 = win and 1 = loss

Table 4.

Pearson Correlation Coefficients, N = 9840 Prob > |r| under H0: Rho=0																	
	TeamPoints	Assists	Steals	Blocks	Turnovers	TotalRebounds	OffRebounds	TotalFouls	FieldGoals	FieldGoalsAttempted	FieldGoals_	X3PointShots	X3PointShotsAttempted	X3PointShots_	FreeThrows	FreeThrowsAttempted	FreeThrows_
TeamPoints	1.00000	0.57461 <.0001	0.10242 <.0001	0.05699 <.0001	-0.11590 <.0001	0.09331 <.0001	-0.01159 0.2505	0.15118 <.0001	0.83620 <.0001	0.28443 <.0001	0.71019 <.0001	0.51802 <.0001	0.26573 <.0001	0.47754 <.0001	0.31852 <.0001	0.26783 <.0001	0.17570 <.0001
Assists	0.57461 <.0001	1.00000	0.09377 <.0001	0.08764 <.0001	-0.03205 0.0015	0.01426 0.1572	-0.11555 <.0001	-0.00487 0.6288	0.63756 <.0001	0.19686 <.0001	0.55459 <.0001	0.41929 <.0001	0.22585 <.0001	0.37481 <.0001	-0.15313 <.0001	-0.16692 <.0001	0.01541 0.1263
Steals	0.10242 <.0001	0.09377 <.0001	1.00000	0.00544 0.5894	0.12575 <.0001	-0.10977 <.0001	0.03543 0.0004	0.02780 0.0058	0.09768 <.0001	0.12144 <.0001	0.02082 0.0389	0.00809 0.4223	0.04661 <.0001	-0.02905 0.0040	0.03940 <.0001	0.05120 <.0001	-0.01771 0.0789
Blocks	0.05699 <.0001	0.08764 <.0001	0.00544 0.5894	1.00000	0.03301 0.0011	0.16724 <.0001	0.00782 0.4382	-0.01144 0.2566	0.06232 <.0001	0.03716 0.0002	0.04164 <.0001	-0.00210 0.8347	-0.00561 0.5780	0.00076 0.9399	0.01254 0.2136	0.01572 0.1190	0.00326 0.7463
Turnovers	-0.11590 <.0001	-0.03205 0.0015	0.12575 <.0001	0.03301 0.0011	1.00000	0.12567 <.0001	0.04467 <.0001	0.16630 <.0001	-0.15189 <.0001	-0.27690 <.0001	0.02761 0.0062	-0.00575 0.5683	-0.01200 0.2338	0.00508 0.6143	0.02267 0.0245	0.02763 0.0061	-0.01028 0.3078
TotalRebounds	0.09331 <.0001	0.01426 0.1572	-0.10977 <.0001	0.16724 <.0001	0.12567 <.0001	1.00000	0.56438 <.0001	0.00505 0.6165	0.08015 <.0001	0.42475 <.0001	-0.20653 <.0001	-0.01344 0.1825	0.09028 <.0001	-0.11536 <.0001	0.06319 <.0001	0.09115 <.0001	-0.04979 <.0001
OffRebounds	-0.01159 0.2505	-0.11555 <.0001	0.03543 0.0004	0.00782 0.4382	0.04467 <.0001	0.56438 <.0001	1.00000	0.04638 <.0001	-0.00092 0.9271	0.50914 <.0001	-0.36360 <.0001	-0.12397 <.0001	0.01671 0.0975	-0.20134 <.0001	0.05234 <.0001	0.09246 <.0001	-0.08969 <.0001
TotalFouls	0.15118 <.0001	-0.00487 0.6288	0.02780 0.0058	-0.01144 0.2566	0.16630 <.0001	0.00505 0.6165	0.04638 <.0001	1.00000	0.05137 <.0001	0.07434 <.0001	0.00609 0.5457	0.03306 0.0010	0.04743 <.0001	0.00535 0.5953	0.20105 <.0001	0.21383 <.0001	0.00027 0.9788
FieldGoals	0.83620 <.0001	0.63756 <.0001	0.09768 <.0001	0.06232 <.0001	-0.15189 <.0001	0.08015 <.0001	-0.00092 0.9271	0.05137 <.0001	1.00000	0.44173 <.0001	0.77702 <.0001	0.32909 <.0001	0.10958 <.0001	0.36324 <.0001	-0.17532 <.0001	-0.18550 <.0001	0.00488 0.6287
FieldGoalsAttempted	0.28443 <.0001	0.19686 <.0001	0.12144 <.0001	0.03716 0.0002	-0.27690 <.0001	0.42475 <.0001	0.50914 <.0001	0.07434 <.0001	0.44173 <.0001	1.00000	-0.21643 <.0001	0.09168 <.0001	0.26396 <.0001	-0.13680 <.0001	-0.21774 <.0001	-0.21722 <.0001	-0.03697 0.0002
FieldGoals_	0.71019 <.0001	0.55459 <.0001	0.02082 0.0389	0.04164 <.0001	0.02761 0.0062	-0.20653 <.0001	-0.36360 <.0001	0.00609 0.5457	0.77702 <.0001	-0.21643 <.0001	1.00000	0.29308 <.0001	-0.06620 <.0001	0.49359 <.0001	-0.03591 0.0004	-0.04704 <.0001	0.03073 0.0023
X3PointShots	0.51802 <.0001	0.41929 <.0001	0.00809 0.4223	-0.00210 0.8347	-0.00575 0.5683	-0.01344 0.1825	-0.12397 <.0001	0.03306 0.0010	0.32909 <.0001	0.09168 <.0001	0.29308 <.0001	1.00000	0.74701 <.0001	0.68588 <.0001	-0.09913 <.0001	-0.11535 <.0001	0.02779 0.0058
X3PointShotsAttempted	0.26573 <.0001	0.22585 <.0001	0.04661 <.0001	-0.00561 0.5780	-0.01200 0.2338	0.09028 <.0001	0.01671 0.0975	0.04743 <.0001	0.10958 <.0001	0.26396 <.0001	-0.06620 <.0001	0.74701 <.0001	1.00000	0.06839 <.0001	-0.09188 <.0001	-0.09937 <.0001	0.00724 0.4726
X3PointShots_	0.47754 <.0001	0.37481 <.0001	-0.02905 0.0040	0.00076 0.9399	0.00508 0.6143	-0.11536 <.0001	-0.20134 <.0001	0.00535 0.5953	0.36324 <.0001	-0.13680 <.0001	0.49359 <.0001	0.68588 <.0001	0.06839 <.0001	1.00000	-0.05027 <.0001	-0.06625 <.0001	0.03267 0.0012
FreeThrows	0.31852 <.0001	-0.15313 <.0001	0.03940 <.0001	0.01254 0.2136	0.02267 0.0245	0.06319 <.0001	0.05234 <.0001	0.20105 <.0001	-0.17532 <.0001	-0.21774 <.0001	-0.03591 0.0004	-0.09913 <.0001	-0.09188 <.0001	-0.05027 <.0001	1.00000	0.92385 <.0001	0.33192 <.0001
FreeThrowsAttempted	0.26783 <.0001	-0.16692 <.0001	0.05120 <.0001	0.01572 0.1190	0.02763 0.0061	0.09115 <.0001	0.09246 <.0001	0.21383 <.0001	-0.18550 <.0001	-0.21722 <.0001	-0.04704 <.0001	-0.11535 <.0001	-0.09937 <.0001	-0.06625 <.0001	0.92385 <.0001	1.00000	-0.02928 0.0037
FreeThrows_	0.17570 <.0001	0.01541 0.1263	-0.01771 0.0789	0.00326 0.7463	-0.01028 0.3078	-0.04979 <.0001	-0.08969 <.0001	0.00027 0.9788	0.00488 0.6287	-0.03697 0.0002	0.03073 0.0023	0.02779 0.0058	0.00724 0.4726	0.03267 0.0012	0.33192 <.0001	-0.02928 0.0037	1.00000

Table 5.

Pearson Correlation Coefficients, N = 9840
Prob > |r| under H0: Rho=0

	Assists	Steals	Blocks	Turnovers	TotalRebounds	OffRebounds	TotalFouls	FieldGoals	FieldGoalsAttempted	FieldGoals_	X3PointShots	X3PointShotsAttempted	X3PointShots_	FreeThrows	FreeThrowsAttempted	FreeThrows_
Assists	1.00000	0.09377 <.0001	0.08764 <.0001	-0.03205 0.0015	0.01426 0.1572	-0.11555 <.0001	-0.00487 0.6288	0.63756 <.0001	0.19686 <.0001	0.55459 <.0001	0.41929 <.0001	0.22685 <.0001	0.37481 <.0001	-0.15313 <.0001	-0.16692 <.0001	0.01541 0.1263
Steals	0.09377 <.0001	1.00000	0.00544 0.5894	0.12575 <.0001	-0.10977 <.0001	0.03543 0.0004	0.02780 0.0058	0.09768 <.0001	0.12144 <.0001	0.02082 0.0389	0.00809 0.4223	0.04661 <.0001	-0.02905 0.0040	0.03940 <.0001	0.05120 <.0001	-0.01771 0.0789
Blocks	0.08764 <.0001	0.00544 0.5894	1.00000	0.03301 0.0011	0.16724 <.0001	0.00782 0.4382	-0.01144 0.2566	0.06232 <.0001	0.03716 0.0002	0.04164 <.0001	-0.00210 0.8347	-0.00561 0.5780	0.00076 0.9399	0.01254 0.2136	0.01572 0.1190	0.00326 0.7463
Turnovers	-0.03205 0.0015	0.12575 <.0001	0.03301 0.0011	1.00000	0.12567 <.0001	0.04467 <.0001	0.16630 <.0001	-0.15189 <.0001	-0.27690 <.0001	0.02761 0.0062	-0.00575 0.5683	-0.01200 0.2338	0.00508 0.6143	0.02267 0.0245	0.02763 0.0061	-0.01028 0.3078
TotalRebounds	0.01426 0.1572	-0.10977 <.0001	0.16724 <.0001	0.12567 <.0001	1.00000	0.56438 <.0001	0.00505 0.6165	0.08015 <.0001	0.42475 <.0001	-0.20653 <.0001	-0.01344 0.1825	0.09028 <.0001	-0.11536 <.0001	0.06319 <.0001	0.09115 <.0001	-0.04979 <.0001
OffRebounds	-0.11555 <.0001	0.03543 0.0004	0.00782 0.4382	0.04467 <.0001	0.56438 <.0001	1.00000	0.04638 <.0001	-0.00092 0.9271	0.50914 <.0001	-0.35350 <.0001	-0.12397 <.0001	0.01671 0.0975	-0.20134 <.0001	0.05234 <.0001	0.09246 <.0001	-0.08969 <.0001
TotalFouls	-0.00487 0.6288	0.02780 0.0058	-0.01144 0.2566	0.16630 <.0001	0.00505 0.6165	0.04638 <.0001	1.00000	0.05137 <.0001	0.07434 <.0001	0.00609 0.5457	0.03306 0.0010	0.04743 <.0001	0.00535 0.5963	0.20105 <.0001	0.21383 <.0001	0.00027 0.9788
FieldGoals	0.63756 <.0001	0.09768 <.0001	0.06232 <.0001	-0.15189 <.0001	0.08015 <.0001	-0.00092 0.9271	0.05137 <.0001	1.00000	0.44173 <.0001	0.77702 <.0001	0.32909 <.0001	0.10958 <.0001	0.36324 <.0001	-0.17532 <.0001	-0.18550 <.0001	0.00488 0.6287
FieldGoalsAttempted	0.19686 <.0001	0.12144 <.0001	0.03716 0.0002	-0.27690 <.0001	0.42475 <.0001	0.50914 <.0001	0.07434 <.0001	0.44173 <.0001	1.00000	-0.21643 <.0001	0.09168 <.0001	0.26396 <.0001	-0.13680 <.0001	-0.21774 <.0001	-0.21722 <.0001	-0.03697 0.0002
FieldGoals_	0.55459 <.0001	0.02082 0.0389	0.04164 <.0001	0.02761 0.0062	-0.20653 <.0001	-0.35350 <.0001	0.00609 0.5457	0.77702 <.0001	-0.21643 <.0001	1.00000	0.29308 <.0001	-0.06620 <.0001	0.49359 <.0001	-0.03591 0.0004	-0.04704 <.0001	0.03073 0.0023
X3PointShots	0.41929 <.0001	0.00809 0.4223	-0.00210 0.8347	-0.00575 0.5683	-0.01344 0.1825	-0.12397 <.0001	0.03306 0.0010	0.32909 <.0001	0.09168 <.0001	0.29308 <.0001	1.00000	0.74701 <.0001	0.68588 <.0001	-0.09913 <.0001	-0.11535 <.0001	0.02779 0.0058
X3PointShotsAttempted	0.22685 <.0001	0.04661 <.0001	-0.00561 0.5780	-0.01200 0.2338	0.09028 <.0001	0.01671 0.0975	0.04743 <.0001	0.10958 <.0001	0.26396 <.0001	-0.06620 <.0001	0.74701 <.0001	1.00000	0.06839 <.0001	-0.09188 <.0001	-0.09937 <.0001	0.00724 0.4726
X3PointShots_	0.37481 <.0001	-0.02905 0.0040	0.00076 0.9399	0.00508 0.6143	-0.11536 <.0001	-0.20134 <.0001	0.00535 0.5963	0.36324 <.0001	-0.13680 <.0001	0.49359 <.0001	0.68588 <.0001	0.06839 <.0001	1.00000	-0.05027 <.0001	-0.06625 <.0001	0.03267 0.0012
FreeThrows	-0.15313 <.0001	0.03940 <.0001	0.01254 0.2136	0.02267 0.0245	0.06319 <.0001	0.05234 <.0001	0.20105 <.0001	-0.17532 <.0001	-0.21774 <.0001	-0.03591 0.0004	-0.09913 <.0001	-0.09188 <.0001	-0.05027 <.0001	1.00000	0.92385 <.0001	0.33192 <.0001
FreeThrowsAttempted	-0.16692 <.0001	0.05120 <.0001	0.01572 0.1190	0.02763 0.0061	0.09115 <.0001	0.09246 <.0001	0.21383 <.0001	-0.18550 <.0001	-0.21722 <.0001	-0.04704 <.0001	-0.11535 <.0001	-0.09937 <.0001	-0.06625 <.0001	0.92385 <.0001	1.00000	-0.02928 0.0037
FreeThrows_	0.01541 0.1263	-0.01771 0.0789	0.00326 0.7463	-0.01028 0.3078	-0.04979 <.0001	-0.08969 <.0001	0.00027 0.9788	0.00488 0.6287	-0.03697 0.0002	0.03073 0.0023	0.02779 0.0058	0.00724 0.4726	0.03267 0.0012	0.33192 <.0001	-0.02928 0.0037	1.00000

Table 6.

Pearson Correlation Coefficients, N = 9840
Prob > |r| under H0: Rho=0

	Assists	Steals	Blocks	Turnovers	TotalRebounds	OffRebounds	TotalFouls	FieldGoalsAttempted	FieldGoals_	X3PointShotsAttempted	X3PointShots_	FreeThrowsAttempted	FreeThrows_
Assists	1.00000	0.09377 <.0001	0.08764 <.0001	-0.03205 0.0015	0.01426 0.1572	-0.11555 <.0001	-0.00487 0.6288	0.19686 <.0001	0.55459 <.0001	0.22685 <.0001	0.37481 <.0001	-0.16692 <.0001	0.01541 0.1263
Steals	0.09377 <.0001	1.00000	0.00544 0.5894	0.12575 <.0001	-0.10977 <.0001	0.03543 0.0004	0.02780 0.0058	0.12144 <.0001	0.02082 0.0389	0.04661 <.0001	-0.02905 0.0040	0.05120 <.0001	-0.01771 0.0789
Blocks	0.08764 <.0001	0.00544 0.5894	1.00000	0.03301 0.0011	0.16724 <.0001	0.00782 0.4382	-0.01144 0.2566	0.03716 0.0002	0.04164 <.0001	-0.00561 0.5780	0.00076 0.9399	0.01572 0.1190	0.00326 0.7463
Turnovers	-0.03205 0.0015	0.12575 <.0001	0.03301 0.0011	1.00000	0.12567 <.0001	0.04467 <.0001	0.16630 <.0001	-0.27690 <.0001	0.02761 0.0062	-0.01200 0.2338	0.00508 0.6143	0.02763 0.0061	-0.01028 0.3078
TotalRebounds	0.01426 0.1572	-0.10977 <.0001	0.16724 <.0001	0.12567 <.0001	1.00000	0.56438 <.0001	0.00505 0.6165	0.42475 <.0001	-0.20653 <.0001	0.09028 <.0001	-0.11536 <.0001	0.09115 <.0001	-0.04979 <.0001
OffRebounds	-0.11555 <.0001	0.03543 0.0004	0.00782 0.4382	0.04467 <.0001	0.56438 <.0001	1.00000	0.04638 <.0001	0.50914 <.0001	-0.35350 <.0001	0.01671 0.0975	-0.20134 <.0001	0.09246 <.0001	-0.08969 <.0001
TotalFouls	-0.00487 0.6288	0.02780 0.0058	-0.01144 0.2566	0.16630 <.0001	0.00505 0.6165	0.04638 <.0001	1.00000	0.07434 <.0001	0.00609 0.5457	0.04743 <.0001	0.00535 0.5953	0.21383 <.0001	0.00027 0.9788
FieldGoalsAttempted	0.19686 <.0001	0.12144 <.0001	0.03716 0.0002	-0.27690 <.0001	0.42475 <.0001	0.50914 <.0001	0.07434 <.0001	1.00000	-0.21643 <.0001	0.26396 <.0001	-0.13680 <.0001	-0.21722 <.0001	-0.03697 0.0002
FieldGoals_	0.55459 <.0001	0.02082 0.0389	0.04164 <.0001	0.02761 0.0062	-0.20653 <.0001	-0.35350 <.0001	0.00609 0.5457	-0.21643 <.0001	1.00000	-0.06620 <.0001	0.49359 <.0001	-0.04704 <.0001	0.03073 0.0023
X3PointShotsAttempted	0.22685 <.0001	0.04661 <.0001	-0.00561 0.5780	-0.01200 0.2338	0.09028 <.0001	0.01671 0.0975	0.04743 <.0001	0.26396 <.0001	-0.06620 <.0001	1.00000	0.06839 <.0001	-0.09937 <.0001	0.00724 0.4726
X3PointShots_	0.37481 <.0001	-0.02905 0.0040	0.00076 0.9399	0.00508 0.6143	-0.11536 <.0001	-0.20134 <.0001	0.00535 0.5953	-0.13680 <.0001	0.49359 <.0001	0.06839 <.0001	1.00000	-0.06625 <.0001	0.03267 0.0012
FreeThrowsAttempted	-0.16692 <.0001	0.05120 <.0001	0.01572 0.1190	0.02763 0.0061	0.09115 <.0001	0.09246 <.0001	0.21383 <.0001	-0.21722 <.0001	-0.04704 <.0001	-0.09937 <.0001	-0.06625 <.0001	1.00000	-0.02928 0.0037
FreeThrows_	0.01541 0.1263	-0.01771 0.0789	0.00326 0.7463	-0.01028 0.3078	-0.04979 <.0001	-0.08969 <.0001	0.00027 0.9788	-0.03697 0.0002	0.03073 0.0023	0.00724 0.4726	0.03267 0.0012	-0.02928 0.0037	1.00000

Figure 1.

Full Model R output

Model Summary

```
Call:

glm(formula = outcome ~ location + fg + fg_attempted + fg_percent +
    three + three_attempted + three_percent + freethrow + freethrow_attempted +
    freethrow_percent + totalreb + offreb + assists + steals +
    blocks + turnovers + fouls, family = "binomial", data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9984  -0.4916  -0.0016   0.4537   3.4536

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         -9.679624   3.582469  -2.702 0.006894 **
locationHome         0.386138   0.062155   6.212 5.22e-10 ***
fg                   0.264283   0.092113   2.869 0.004116 **
fg_attempted        -0.281631   0.042385  -6.645 3.04e-11 ***
fg_percent          11.558133   7.782961   1.485 0.137529
three                0.185010   0.047743   3.875 0.000107 ***
three_attempted     -0.025075   0.017673  -1.419 0.155956
three_percent        1.463648   1.148099   1.275 0.202364
freethrow            0.121167   0.039917   3.036 0.002401 **
freethrow_attempted -0.064337   0.030826  -2.087 0.036878 *
freethrow_percent    1.998068   0.887744   2.251 0.024403 *
totalreb             0.367937   0.009213  39.935  < 2e-16 ***
offreb              -0.017723   0.011285  -1.570 0.116304
assists              0.023860   0.008308   2.872 0.004082 **
steals               0.396321   0.013269  29.868  < 2e-16 ***
blocks               0.121609   0.012654   9.610  < 2e-16 ***
turnovers           -0.333897   0.011084 -30.124  < 2e-16 ***
fouls               -0.060801   0.007822  -7.773 7.64e-15 ***
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641.1  on 9839  degrees of freedom
Residual deviance:  6693.3  on 9822  degrees of freedom
AIC: 6729.3

Number of Fisher Scoring iterations: 6
```

VIF

```
           location                  fg        fg_attempted          fg_percent
           1.020442          174.051631          102.789664          129.754941
              three     three_attempted       three_percent           freethrow
          28.438946           16.751974           11.299079           60.790551
freethrow_attempted   freethrow_percent            totalreb              offreb
          54.795805            8.782473            3.352066            1.979752
            assists              steals              blocks           turnovers
           1.597636            1.558736            1.038369            1.896310
              fouls
           1.172617
```

Concordance

```
n= 9840
Concordance= 0.9278 se= 0.002455
concordant discordant     tied.x     tied.y    tied.xy
  22459804    1746596          0   24201480          0
```

Logistic Relation Regression

```
Analysis of Deviance Table

Model 1: outcome ~ 1
Model 2: outcome ~ location + fg + fg_attempted + fg_percent + three +
    three_attempted + three_percent + freethrow + freethrow_attempted +
    freethrow_percent + totalreb + offreb + assists + steals +
    blocks + turnovers + fouls
  Resid. Df Resid. Dev Df Deviance
1      9839    13641.1
2      9822     6693.3 17   6947.8

p-value
[1] 0
```

Figure 2.

Method 1 First Model

Model Summary

```
Call:
glm(formula = outcome ~ location + fg_attempted + fg_percent + 
    three_attempted + three_percent + freethrow_attempted + freethrow_percent + 
    totalreb + offreb + assists + steals + blocks + turnovers + 
    fouls, family = "binomial", data = data2)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.0066  -0.4943  -0.0013   0.4557   3.3926  

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)         -23.075027   0.788522 -29.264  < 2e-16 ***
locationHome          0.376908   0.061922   6.087 1.15e-09 ***
fg_attempted         -0.162540   0.007255 -22.403  < 2e-16 ***
fg_percent           33.898254   1.002821  33.803  < 2e-16 ***
three_attempted       0.040263   0.004757   8.464  < 2e-16 ***
three_percent         5.680238   0.383747  14.802  < 2e-16 ***
freethrow_attempted   0.027665   0.004685   5.905 3.52e-09 ***
freethrow_percent     4.507852   0.307826  14.644  < 2e-16 ***
totalreb              0.365113   0.009146  39.920  < 2e-16 ***
offreb               -0.017390   0.011227  -1.549  0.12138    
assists               0.026491   0.008273   3.202  0.00136 ** 
steals                0.392162   0.013185  29.742  < 2e-16 ***
blocks                0.120345   0.012608   9.545  < 2e-16 ***
turnovers            -0.332953   0.011045 -30.145  < 2e-16 ***
fouls                -0.060430   0.007797  -7.750 9.15e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641.1  on 9839  degrees of freedom
Residual deviance:  6728.3  on 9825  degrees of freedom
AIC: 6758.3

Number of Fisher Scoring iterations: 6
```

VIF

```
           location        fg_attempted          fg_percent     three_attempted 
           1.018741            3.030766            2.161655            1.222110 
      three_percent freethrow_attempted   freethrow_percent            totalreb 
           1.255745            1.267221            1.058059            3.327572 
             offreb             assists              steals              blocks 
           1.973750            1.598414            1.549136            1.037541 
          turnovers               fouls 
           1.895167            1.171908 
```

Concordance

```
n= 9840 
Concordance= 0.9271 se= 0.002474
concordant discordant     tied.x     tied.y    tied.xy 
  22441695    1764705          0   24201480          0 
```

Figure 3.

Method 1 Second Model (Best Model)

Model Summary

```
Call:
glm(formula = outcome ~ location + fg_attempted + fg_percent + 
    three_attempted + three_percent + freethrow_attempted + freethrow_percent + 
    totalreb + assists + steals + blocks + turnovers + fouls, 
    family = "binomial", data = data2)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.9933  -0.4917  -0.0012   0.4568   3.3686  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)        -22.871979   0.776441 -29.457  < 2e-16 ***
locationHome         0.374232   0.061881   6.048 1.47e-09 ***
fg_attempted        -0.166552   0.006793 -24.519  < 2e-16 ***
fg_percent          34.090550   0.996061  34.225  < 2e-16 ***
three_attempted      0.041391   0.004701   8.804  < 2e-16 ***
three_percent        5.659961   0.383431  14.761  < 2e-16 ***
freethrow_attempted  0.026563   0.004629   5.738 9.57e-09 ***
freethrow_percent    4.532662   0.307436  14.743  < 2e-16 ***
totalreb             0.361636   0.008842  40.900  < 2e-16 ***
assists              0.027596   0.008240   3.349 0.000811 ***
steals               0.391583   0.013174  29.724  < 2e-16 ***
blocks               0.121702   0.012579   9.675  < 2e-16 ***
turnovers           -0.334984   0.010975 -30.522  < 2e-16 ***
fouls               -0.060034   0.007792  -7.704 1.31e-14 ***
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641.1  on 9839  degrees of freedom
Residual deviance:  6730.7  on 9826  degrees of freedom
AIC: 6758.7

Number of Fisher Scoring iterations: 6
```

VIF

```
           location        fg_attempted          fg_percent     three_attempted 
           1.017803            2.657984            2.133157            1.194646 
      three_percent freethrow_attempted   freethrow_percent            totalreb 
           1.253705            1.237190            1.055362            3.111985 
            assists              steals              blocks           turnovers 
           1.586809            1.546667            1.032656            1.871233 
              fouls 
           1.170411 
```

Concordance

```
n= 9840 
Concordance= 0.927 se= 0.002475
concordant discordant     tied.x     tied.y    tied.xy 
  22440239    1766161          0   24201480          0 
```

Logistic Regression Relation

```
Analysis of Deviance Table

Model 1: outcome ~ 1
Model 2: outcome ~ location + fg_attempted + fg_percent + three_attempted + 
    three_percent + freethrow_attempted + freethrow_percent + 
    totalreb + assists + steals + blocks + turnovers + fouls
  Resid. Df Resid. Dev Df Deviance
1      9839    13641.1            
2      9826     6730.7 13   6910.5

p-value
[1] 0
```

Figure 4.

Method 2 Backward Selection

```
Start:  AIC=6729.32
outcome ~ location + fg + fg_attempted + fg_percent + three + 
    three_attempted + three_percent + freethrow + freethrow_attempted + 
    freethrow_percent + totalreb + offreb + assists + steals + 
    blocks + turnovers + fouls

                      Df Deviance    AIC
- three_percent        1   6695.0 6729.0
<none>                     6693.3 6729.3
- three_attempted      1   6695.3 6729.3
- fg_percent           1   6695.5 6729.5
- offreb               1   6695.8 6729.8
- freethrow_attempted  1   6697.7 6731.7
- freethrow_percent    1   6698.4 6732.4
- fg                   1   6701.6 6735.6
- assists              1   6701.6 6735.6
- freethrow            1   6702.5 6736.5
- three                1   6708.3 6742.3
- location             1   6732.0 6766.0
- fg_attempted         1   6737.9 6771.9
- fouls                1   6754.8 6788.8
- blocks               1   6788.2 6822.2
- steals               1   7845.9 7879.9
- turnovers            1   7865.7 7899.7
- totalreb             1   9381.2 9415.2

Step:  AIC=6728.96
outcome ~ location + fg + fg_attempted + fg_percent + three + 
    three_attempted + freethrow + freethrow_attempted + freethrow_percent + 
    totalreb + offreb + assists + steals + blocks + turnovers + 
    fouls

                      Df Deviance    AIC
<none>                     6695.0 6729.0
- offreb               1   6697.5 6729.5
- fg_percent           1   6697.5 6729.5
- freethrow_attempted  1   6699.4 6731.4
- freethrow_percent    1   6700.0 6732.0
- fg                   1   6702.6 6734.6
- assists              1   6703.3 6735.3
- freethrow            1   6704.2 6736.2
- three_attempted      1   6730.8 6762.8
- location             1   6734.1 6766.1
- fg_attempted         1   6738.4 6770.4
- fouls                1   6756.2 6788.2
- blocks               1   6789.7 6821.7
- three                1   6942.2 6974.2
- steals               1   7849.9 7881.9
- turnovers            1   7867.0 7899.0
- totalreb             1   9382.5 9414.5


Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.9998  -0.4933  -0.0016   0.4539   3.4623  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)         -9.551453   3.580503  -2.668  0.00764 ** 
locationHome         0.388132   0.062129   6.247 4.18e-10 ***
fg                   0.253812   0.091728   2.767  0.00566 ** 
fg_attempted        -0.276921   0.042212  -6.560 5.37e-11 ***
fg_percent          12.437463   7.751019   1.605  0.10858    
three                0.242526   0.015907  15.246  < 2e-16 ***
three_attempted     -0.045433   0.007623  -5.960 2.52e-09 ***
freethrow            0.121688   0.039905   3.049  0.00229 ** 
freethrow_attempted -0.064814   0.030817  -2.103  0.03545 *  
freethrow_percent    1.983921   0.887799   2.235  0.02544 *  
totalreb             0.367863   0.009211  39.938  < 2e-16 ***
offreb              -0.017953   0.011283  -1.591  0.11159    
assists              0.023938   0.008306   2.882  0.00395 ** 
steals               0.396633   0.013268  29.895  < 2e-16 ***
blocks               0.121538   0.012655   9.604  < 2e-16 ***
turnovers           -0.333759   0.011080 -30.122  < 2e-16 ***
fouls               -0.060659   0.007819  -7.758 8.61e-15 ***
---

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641  on 9839  degrees of freedom
Residual deviance:  6695  on 9823  degrees of freedom
AIC: 6729
```

Figure 5.

Method 2 Backward Selection: Removal of the two predictors

Model Summary

```
Call:
glm(formula = outcome ~ location + fg + fg_attempted + three +
    three_attempted + freethrow + freethrow_attempted + freethrow_percent +
    totalreb + assists + steals + blocks + turnovers + fouls,
    family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0174  -0.4953  -0.0018   0.4552   3.4416

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         -3.676297   0.828518  -4.437 9.11e-06 ***
locationHome         0.384969   0.062071   6.202 5.57e-10 ***
fg                   0.402397   0.011803  34.094  < 2e-16 ***
fg_attempted        -0.348278   0.009037 -38.537  < 2e-16 ***
three                0.242114   0.015883  15.243  < 2e-16 ***
three_attempted     -0.044205   0.007553  -5.853 4.84e-09 ***
freethrow            0.120178   0.039929   3.010  0.00261 **
freethrow_attempted -0.064638   0.030830  -2.097  0.03603 *
freethrow_percent    2.048714   0.888447   2.306  0.02111 *
totalreb             0.364537   0.008909  40.918  < 2e-16 ***
assists              0.025370   0.008271   3.067  0.00216 **
steals               0.396092   0.013255  29.883  < 2e-16 ***
blocks               0.122859   0.012621   9.734  < 2e-16 ***
turnovers           -0.335689   0.011007 -30.498  < 2e-16 ***
fouls               -0.060052   0.007813  -7.687 1.51e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641  on 9839  degrees of freedom
Residual deviance:  6700  on 9825  degrees of freedom
AIC: 6730

Number of Fisher Scoring iterations: 6
```

VIF

```
           location                  fg        fg_attempted               three
           1.018870            2.858351            4.673112            3.143374
    three_attempted           freethrow freethrow_attempted   freethrow_percent
           3.059334           60.886035           54.838229            8.799410
           totalreb             assists              steals              blocks
           3.135521            1.583628            1.556388            1.033628
          turnovers               fouls
           1.872226            1.169769
```

Concordance

```
n= 9840
Concordance= 0.9277 se= 0.002458
concordant discordant     tied.x     tied.y    tied.xy
  22456016    1750384          0   24201480          0
```

Figure 6.

Method 2 Forwards Selection

```
Start:  AIC=13643.14
outcome ~ 1

                     Df Deviance   AIC
+ fg_percent          1    11445 11449
+ fg                  1    12075 12079
+ three_percent       1    12592 12596
+ assists             1    12633 12637
+ totalreb            1    12952 12956
+ three               1    13054 13058
+ blocks              1    13359 13363
+ location            1    13378 13382
+ freethrow           1    13438 13442
+ steals              1    13445 13449
+ turnovers           1    13500 13504
+ freethrow_attempted 1    13514 13518
+ fouls               1    13515 13519
+ freethrow_percent   1    13530 13534
+ fg_attempted        1    13624 13628
+ offreb              1    13625 13629
+ three_attempted     1    13631 13635
<none>                     13641 13643

Step:  AIC=11449.29
outcome ~ fg_percent

                     Df Deviance     AIC
+ totalreb            1   9682.5  9688.5
+ freethrow           1  11130.6 11136.6
+ blocks              1  11161.4 11167.4
+ freethrow_attempted 1  11222.2 11228.2
+ turnovers           1  11225.7 11231.7
+ steals              1  11228.2 11234.2
+ offreb              1  11246.1 11252.1
+ location            1  11261.8 11267.8
+ fouls               1  11279.0 11285.0
+ three               1  11281.3 11287.3
+ three_percent       1  11293.8 11299.8
+ freethrow_percent   1  11341.1 11347.1
+ assists             1  11368.0 11374.0
+ three_attempted     1  11395.7 11401.7
+ fg_attempted        1  11403.8 11409.8
+ fg                  1  11403.8 11409.8
<none>                   11445.3 11449.3

Step:  AIC=9688.52
outcome ~ fg_percent + totalreb

                     Df Deviance    AIC
+ turnovers           1   9162.5 9170.5
+ steals              1   9226.5 9234.5
+ freethrow           1   9399.4 9407.4
+ fouls               1   9467.0 9475.0
+ three_percent       1   9478.8 9486.8
+ freethrow_percent   1   9506.2 9514.2
+ freethrow_attempted 1   9523.5 9531.5
+ three               1   9539.2 9547.2
+ fg_attempted        1   9554.4 9562.4
+ fg                  1   9562.3 9570.3
+ blocks              1   9574.6 9582.6
+ location            1   9584.1 9592.1
+ offreb              1   9594.1 9602.1
+ three_attempted     1   9662.8 9670.8
+ assists             1   9674.0 9682.0
<none>                    9682.5 9688.5

Step:  AIC=9170.51
outcome ~ fg_percent + totalreb + turnovers

                     Df Deviance    AIC
+ steals              1   8510.9 8520.9
+ fg_attempted        1   8698.8 8708.8
+ fg                  1   8718.2 8728.2
+ freethrow           1   8857.7 8867.7
+ three_percent       1   8953.4 8963.4
+ freethrow_percent   1   8982.5 8992.5
+ freethrow_attempted 1   8987.9 8997.9
+ three               1   9023.5 9033.5
+ fouls               1   9034.1 9044.1
+ blocks              1   9047.0 9057.0
+ offreb              1   9064.2 9074.2
+ location            1   9079.2 9089.2
+ three_attempted     1   9145.8 9155.8
<none>                    9162.5 9170.5
+ assists             1   9161.1 9171.1

Step:  AIC=8520.87
outcome ~ fg_percent + totalreb + turnovers + steals

                     Df Deviance    AIC
+ fg_attempted        1   7573.2 7585.2
+ fg                  1   7611.9 7623.9
+ freethrow           1   8229.8 8241.8
+ three_percent       1   8242.1 8254.1
+ freethrow_percent   1   8296.0 8308.0
+ offreb              1   8318.4 8330.4
+ three               1   8364.4 8376.4
+ freethrow_attempted 1   8367.7 8379.7
+ fouls               1   8369.4 8381.4
+ blocks              1   8400.2 8412.2
+ location            1   8428.8 8440.8
+ three_attempted     1   8503.2 8515.2
+ assists             1   8506.3 8518.3
<none>                    8510.9 8520.9

Step:  AIC=7585.2
outcome ~ fg_percent + totalreb + turnovers + steals + fg_attempt
ed

                     Df Deviance    AIC
+ three               1   7220.0 7234.0
+ three_percent       1   7294.7 7308.7
+ freethrow_percent   1   7342.5 7356.5
+ three_attempted     1   7439.5 7453.5
+ blocks              1   7475.2 7489.2
+ assists             1   7489.6 7503.6
+ location            1   7511.1 7525.1
+ freethrow           1   7511.3 7525.3
+ fouls               1   7526.1 7540.1
+ offreb              1   7553.7 7567.7
+ fg                  1   7564.2 7578.2
+ freethrow_attempted 1   7569.4 7583.4
<none>                    7573.2 7585.2

Step:  AIC=7233.95
outcome ~ fg_percent + totalreb + turnovers + steals + fg_attempt
ed +
    three

                     Df Deviance    AIC
+ freethrow_percent   1   6992.4 7008.4
+ blocks              1   7111.0 7127.0
+ freethrow           1   7139.9 7155.9
+ location            1   7157.0 7173.0
+ fouls               1   7170.8 7186.8
+ three_attempted     1   7188.8 7204.8
+ three_percent       1   7189.6 7205.6
+ assists             1   7201.8 7217.8
+ freethrow_attempted 1   7210.0 7226.0
+ fg                  1   7212.1 7228.1
+ offreb              1   7216.0 7232.0
<none>                    7220.0 7234.0

Step:  AIC=7008.36
outcome ~ fg_percent + totalreb + turnovers + steals + fg_attempt
ed +
    three + freethrow_percent

                     Df Deviance    AIC
+ blocks              1   6884.5 6902.5
+ location            1   6929.7 6947.7
+ fouls               1   6941.2 6959.2
+ three_attempted     1   6961.5 6979.5
+ three_percent       1   6962.1 6980.1
+ assists             1   6974.7 6992.7
+ freethrow           1   6977.9 6995.9
+ freethrow_attempted 1   6980.7 6998.7
+ fg                  1   6982.8 7000.8
<none>                    6992.4 7008.4
+ offreb              1   6991.2 7009.2
```

```
Step:  AIC=6902.47
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted  +three + freethrow_percent + blocks

                     Df Deviance    AIC
+ location            1   6831.9 6851.9
+ fouls               1   6834.8 6854.8
+ three_attempted     1   6852.4 6872.4
+ three_percent       1   6853.1 6873.1
+ freethrow           1   6868.5 6888.5
+ freethrow_attempted 1   6871.6 6891.6
+ assists             1   6871.7 6891.7
+ fg                  1   6875.3 6895.3
<none>                    6884.5 6902.5
+ offreb              1   6884.4 6904.4

Step:  AIC=6851.94
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks +
location

                     Df Deviance    AIC
+ fouls               1   6789.2 6811.2
+ three_attempted     1   6799.2 6821.2
+ three_percent       1   6800.8 6822.8
+ freethrow           1   6818.3 6840.3
+ freethrow_attempted 1   6821.2 6843.2
+ assists             1   6822.3 6844.3
+ fg                  1   6823.0 6845.0
<none>                    6831.9 6851.9
+ offreb              1   6831.7 6853.7

Step:  AIC=6811.19
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks
+ location + fouls

                     Df Deviance    AIC
+ freethrow           1   6755.0 6779.0
+ three_attempted     1   6756.6 6780.6
+ three_percent       1   6757.9 6781.9
+ freethrow_attempted 1   6759.9 6783.9
+ fg                  1   6780.9 6804.9
+ assists             1   6781.6 6805.6
<none>                    6789.2 6811.2
+ offreb              1   6789.1 6813.1

Step:  AIC=6779.02
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks
+ location + fouls + freethrow

                     Df Deviance    AIC
+ three_attempted     1   6720.4 6746.4
+ three_percent       1   6721.6 6747.6
+ assists             1   6745.5 6771.5
+ fg                  1   6745.9 6771.9
+ freethrow_attempted 1   6749.0 6775.0
<none>                    6755.0 6779.0
+ offreb              1   6753.5 6779.5

Step:  AIC=6746.42
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted +three + freethrow_percent + blocks
+ location + fouls + freethrow +three_attempted

                     Df Deviance    AIC
+ assists             1   6710.3 6738.3
+ fg                  1   6711.6 6739.6
+ freethrow_attempted 1   6715.2 6743.2
+ offreb              1   6716.4 6744.4
<none>                    6720.4 6746.4
+ three_percent       1   6719.2 6747.2

Step:  AIC=6738.3
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks
+ location + fouls + freethrow +three_attempted
+ assists

                     Df Deviance    AIC
+ fg                  1   6702.1 6732.1
+ freethrow_attempted 1   6705.4 6735.4
+ offreb              1   6707.3 6737.3
<none>                    6710.3 6738.3
+ three_percent       1   6709.2 6739.2

Step:  AIC=6732.06
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted +three + freethrow_percent + blocks
+ location + fouls + freethrow +three_attempted
+ assists + fg

                     Df Deviance    AIC
+ freethrow_attempted  1   6697.5 6729.5
+ offreb               1   6699.4 6731.4
<none>                     6702.1 6732.1
+ three_percent        1   6700.3 6732.3

Step:  AIC=6729.49
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks
+ location + fouls + freethrow + three_attempted
+ assists + fg + freethrow_attempted

                Df Deviance    AIC
+ offreb         1   6695.0 6729.0
<none>               6697.5 6729.5
+ three_percent  1   6695.8 6729.8

Step:  AIC=6728.96
outcome ~ fg_percent + totalreb + turnovers + steals
+ fg_attempted + three + freethrow_percent + blocks
+ location + fouls + freethrow +three_attempted
+ assists + fg + freethrow_attempted + offreb

                Df Deviance    AIC
<none>               6695.0 6729.0
+ three_percent  1   6693.3 6729.3
```

Model Summary

```
Call:
glm(formula = outcome ~ fg_percent + totalreb + turnovers
 + steals + fg_attempted + three + freethrow_percent + blocks +
location + fouls + freethrow + three_attempted + assists + fg +
freethrow_attempted +offreb, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9998  -0.4933  -0.0016   0.4539   3.4623

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         -9.551453   3.580503  -2.668  0.00764 **
fg_percent          12.437463   7.751019   1.605  0.10858
totalreb             0.367863   0.009211  39.938  < 2e-16 ***
turnovers           -0.333759   0.011080 -30.122  < 2e-16 ***
steals               0.396633   0.013268  29.895  < 2e-16 ***
fg_attempted        -0.276921   0.042212  -6.560 5.37e-11 ***
three                0.242526   0.015907  15.246  < 2e-16 ***
freethrow_percent    1.983921   0.887799   2.235  0.02544 *
blocks               0.121538   0.012655   9.604  < 2e-16 ***
locationHome         0.388132   0.062129   6.247 4.18e-10 ***
fouls               -0.060659   0.007819  -7.758 8.61e-15 ***
freethrow            0.121688   0.039905   3.049  0.00229 **
three_attempted     -0.045433   0.007623  -5.960 2.52e-09 ***
assists              0.023938   0.008306   2.882  0.00395 **
fg                   0.253812   0.091728   2.767  0.00566 **
freethrow_attempted -0.064814   0.030817  -2.103  0.03545 *
offreb              -0.017953   0.011283  -1.591  0.11159
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641  on 9839  degrees of freedom
Residual deviance:  6695  on 9823  degrees of freedom
AIC: 6729

Number of Fisher Scoring iterations: 6
```

Figure 7.

Method 3 Best AIC

```
  location assists steals blocks turnovers totalreb fouls   fg fg_attempted fg_percent three
1     TRUE    TRUE   TRUE   TRUE      TRUE     TRUE  TRUE TRUE         TRUE       TRUE  TRUE
2     TRUE    TRUE   TRUE   TRUE      TRUE     TRUE  TRUE TRUE         TRUE       TRUE  TRUE
3     TRUE    TRUE   TRUE   TRUE      TRUE     TRUE  TRUE TRUE         TRUE      FALSE  TRUE
4     TRUE    TRUE   TRUE   TRUE      TRUE     TRUE  TRUE TRUE         TRUE       TRUE  TRUE
5     TRUE    TRUE   TRUE   TRUE      TRUE     TRUE  TRUE TRUE         TRUE      FALSE  TRUE

  three_attempted three_percent freethrow freethrow_attempted Criterion
1            TRUE         FALSE      TRUE                TRUE  6730.630
2           FALSE          TRUE      TRUE                TRUE  6730.741
3           FALSE          TRUE      TRUE                TRUE  6730.844
4            TRUE          TRUE      TRUE                TRUE  6731.006
5            TRUE          TRUE      TRUE                TRUE  6731.375
```

```
Call:
glm(formula = outcome ~ location + fg + fg_attempted + fg_percent + 
    three + three_attempted + freethrow + freethrow_attempted + 
    totalreb + assists + steals + blocks + turnovers + fouls, 
    family = "binomial", data = data2)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.0219  -0.4936  -0.0016   0.4568   3.4378  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)         -7.893154   3.528280  -2.237  0.02528 *  
locationHome         0.386820   0.062067   6.232 4.60e-10 ***
fg                   0.251861   0.091708   2.746  0.00603 ** 
fg_attempted        -0.280123   0.042046  -6.662 2.69e-11 ***
fg_percent          12.831110   7.751775   1.655  0.09787 .  
three                0.241032   0.015876  15.182  < 2e-16 ***
three_attempted     -0.043492   0.007550  -5.760 8.39e-09 ***
freethrow            0.206600   0.013801  14.969  < 2e-16 ***
freethrow_attempted -0.130861   0.011339 -11.540  < 2e-16 ***
totalreb             0.364128   0.008904  40.897  < 2e-16 ***
assists              0.024917   0.008268   3.014  0.00258 ** 
steals               0.395498   0.013242  29.868  < 2e-16 ***
blocks               0.123413   0.012624   9.776  < 2e-16 ***
turnovers           -0.335996   0.011008 -30.524  < 2e-16 ***
fouls               -0.060444   0.007809  -7.740 9.94e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13641.1  on 9839  degrees of freedom
Residual deviance:  6702.6  on 9825  degrees of freedom
AIC: 6732.6

Number of Fisher Scoring iterations: 6
```

VIF

```
           location                  fg        fg_attempted          fg_percent 
           1.019184          172.331156          101.142780          128.910743 
              three     three_attempted           freethrow freethrow_attempted 
           3.143889            3.060243            7.280309            7.437876 
           totalreb             assists              steals              blocks 
           3.134172            1.581845            1.554133            1.033434 
          turnovers               fouls 
           1.870055            1.170469 
```

Concordance

```
n= 9840 
Concordance= 0.9276 se= 0.002459
concordant discordant     tied.x     tied.y    tied.xy 
  22453519    1752881          0   24201480          0 
```

References

Agresti, Alan. *An Introduction to Categorical Data Analysis*. Wiley-Interscience, 2007.

"Basketball Statistics and History." *Basketball Reference*, www.basketball-reference.com/.

Brown, Chris. "How Do I Interpret the AIC." *R Bloggers*, 12 Apr. 2018, www.r-bloggers.com/how-do-i-interpret-the-aic/.

Brunner, Jerry. "Stepwise Logistic Regression with R." *University of Toronto Faculty of Arts and Science*, 2011, www.utstat.toronto.edu/~brunner/oldclass/appliedf11/handouts/2101f11StepwiseLogisticR.pdf.

Kelepoouris, Ionas. "NBA Teams Stats for Every Game 2014 - 2018 Period." *Data.world*, 22 Apr. 2018, data.world/ionaskel/nba-teams-stats-for-every-game-2014-2018-period.

Therneau, Terry. "Concordance." *R Documentation*, www.rdocumentation.org/packages/survival/versions/3.1-11/topics/concordance.