# Predicting NHL Standings Using Multiple... Multiple Linear Regression Models

**Barinder Thind**
**301193363**
**STAT 350**
**SFU**

# Predicting NHL Standings Using Multiple Multiple Linear Regression Models

Barinder Thind

*301193363*

## Abstract

The introduction of the salary cap into the NHL in 2004 has had a substantial impact on the way the league has operated since. The difference in quality between any two given teams has decreased significantly and so, as a result, attempting to predict wins and losses [correctly] has become a very difficult task. The purposes of this paper is to structure a model such that it accurately depicts the standings of any given NHL season after $n$ number of games ($n \leq 82$) have been played. This will be done by taking advantage of a very strong relationship: winning percentage and goal differential. By using advanced statistical methods, the models presented in this paper will serve to be resources of how to predict the total number of goals scored by each team in any given year, the total goals allowed by each team, and then ultimately, by combining the previous results, predicting ordinal AND cardinal performances of each team.

*Keywords:* Ridge, PCA, Hockey, Rank, Differentials, Regression

## 1. Introduction

The NHL has a long standing tradition of being notoriously difficult to predict: particularly when it comes to single games. However, like with most things in statistics, there seems to be a relationship that fits the madness when you look back at the entirety of a season. How is it then that some teams end up 20 games above .500[1] while others end up below? One relationship to explain this discrepancy is that of goal differentials and winning

[1] $\frac{x}{n}$ where x = number of wins and n = games played

percentage. If I were to say to you that a team has scored 50 more goals than it has given up throughout a season, then on average, you would expect that team to score more goals per game than it allows. I'll assume the opposite as well. So, the bigger this difference is (positive or negative), the more you can gauge how a team has performed through a season. In fact, analysis has been performed by statisticians that would showcase this relationship. A popular hockey related statistical website, hockeyanalytics, has run regressions that show a 93 percent predictive relationship of goal differentials ($\beta_1$[2]) to winning percentages[3].

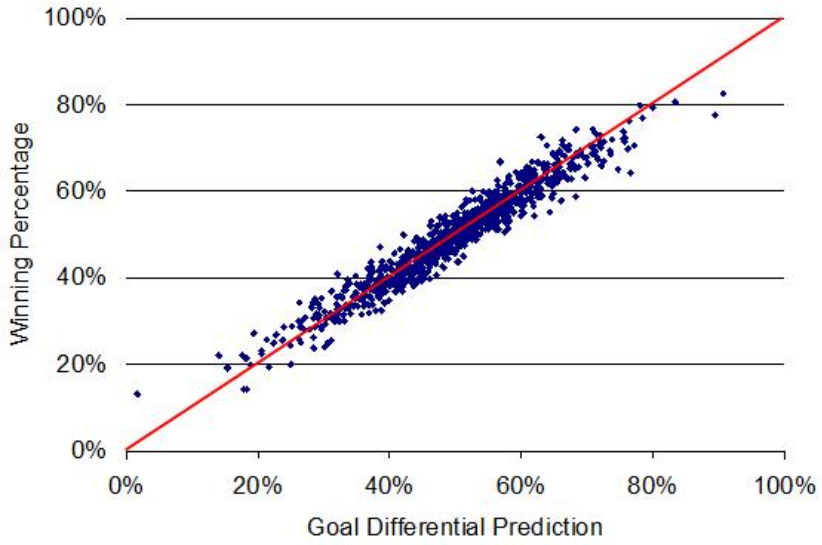


Figure 1: Hockey Analytics Graph: Winning Percentage Vs Goal Differential

Now, since the metric we are looking to satisfy is clear, the question becomes: how will we go about estimating the parameters accurately in a way such that we get concise and useful results? We are left with the task of estimating goal differential for each team and then running a modified simple linear regression. The reasoning being that because goal differential can be a positive or a negative integer, a transformation is required so that an uninterrupted analysis could be done. However, doing such

[2] $\beta_1$ = Coefficient of predictor variable

[3] http://hockeyanalytics.com/2008/01/the-ten-laws-of-hockey-analytics/

things causes problems to arise in other parts of our calculations and so adjustments were made accordingly and are outlined in section 2 [Methods].

Taking another step back however, we see that before we are even graced by the problems above, we have to figure out the particulars of what makes up goal differentials. That is, we need to precisely predict to the best of our ability the goals-for and goals-against for each team.[4]

Model 1 will from now on refer to the analysis that allowed us to predict goals-for. My approach to this was rather trivial: I say that if you can predict the number of goals each player will score in a year, then you can figure out the total number of goals scored by each team by taking a partition $G_\alpha$ ($\alpha : 1, 2, ..., 30$; where 1, 2, 3,... corresponds to a particular hockey team in alphabetical order) of E (where E is the set containing every player belonging to an NHL team) $\ni$ each member of $G_\alpha$ only belongs to that 1 corresponding NHL team. Then $\forall\ G_\alpha$:

(1) $\sum_{i=1}^{n} G_{\alpha i}$ = Total Goals Scored By Each Team

The variables I chose to consider for this model include:

- Age
- Shots Taken
- Shot Percentage
- Games Played

I will go into more details as to why in section [2].

Model 2 will from now on refer to the analysis that allowed us to predict goals-against. This model required the usage of logistic regression to factor in the odds that a team will have an injury to a goalie as that is a significant

[4]Total number of goals scored by and against each team throughout the season

factor in a team's performance.[5] Predictive factors that were considered include:

- FaceOff Percentage
- Blocked Shots
- CorsiFor Percentage
- Save Percentage
- FenwickFor Percentage

The methodologies, as mentioned earlier, will be explored more in depth in the next section. As a final note, Model 3 will refer to the one talked about initially (Winning percentage, Goal differentials).

## 2. Methods

### *2.1. Model 1 - GOALS FOR*

The initial expectation of this model to begin with was to find an aggregate of goals scored by each team. While the goal has remained the same, the method to getting there went through a couple of changes. First, I needed to take into account that, because my data set had 30,000+ observations, there was room for a lot of outliers and irrelevant pieces of information that would significantly impact my results. The following is the steps taken to remedy situations like that with the justifications to go along with it.

1. **Age** - The variable of age was adjusted so that every player who played while he was over 40 was removed. This is because most NHL players careers typically end before they reach 40 and so, when a player does in fact get to an age of 40+, it usually means they will only serve to be outliers.

2. **Goals** - This variable was adjusted so that all goal totals of 60 or more are removed from the data set. This is because it is extremely rare to find

[5]That is not to say that goal tending is the most important aspect of a team's ability to be successful, but rather that while there are no injuries, the effect is mostly noise, but when an injury does occur, the result has the potential to be devastating

players who score at such a pace. The data set I used takes in account every player who has played since the early 1960s and the game was a lot different back then[6]. Because of this disparity, any particularly outlandish goal results from those earlier eras would give us inaccurate results and so, were removed.

3. **Games Played** - This variable was a bit difficult to deal with because a player that plays a small amount of games will obviously score fewer goals (than they would had they played the entire season), so to adjust, small game totals were removed. The number I settled on was 25 and there is no particularly reasoning for this number. However, this is a moot point as ultimately, this variable was left out of the analysis.

4. **Shots** - For this variable, I removed all results with shots less than 30. This is mostly a judgment call as it is tough to find statistics on this, but it isn't completely without base. Calculating the mean number of shots (with all observations of 10 shots or less removed), we get approximately 108 with a standard deviation of about 67. I felt that I encompassed most reasonable shot totals by excluding any below 30.

5. **Shot Percentage** - This last variable had analysis performed such that any ridiculous totals[7] were removed. I did this by only including values belonging to the 95 percent confidence interval. Shot percentages greater than 16 and less than 6 were excluded.

While initially the plan was to predict the total number of goals scored by each player on a hockey team, a different approach had to be carved to take into account injuries. Instead then, the decision to predict goals-per-game was made. This allowed me to adjust for any players being taken in and out of the line up throughout a season and thus resulted in a much more dynamic and flexible model.

[6]As an example for perspective, the goals per game by a team in 1980-1981 was 3.84 and the goals per game in the 2014-2015 season was 2.73

[7]For example, player A plays one hockey game, takes 1 shot, and scores 1 goal. Then player A has a shot percentage of 100.

The next step was to check for the strength of the relationship between the predictor variables and goals-per-game[8].

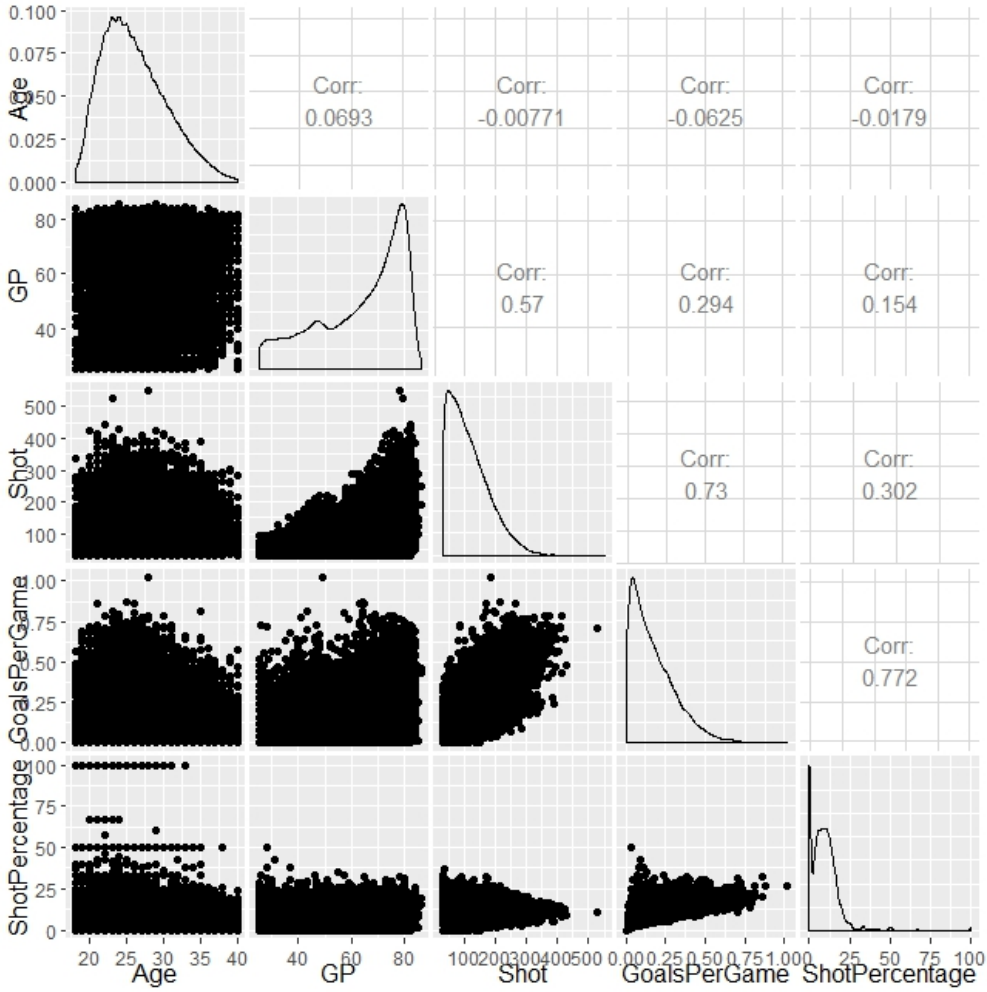


Figure 2: Scatterplot Matrix for Model 1 Variables

As you can discern from figure 2, it seems that games played and age have a very weak relationship with the number of goals a player scores per

[8]This variable was actually not a part of the data set and was created by me. Suppose x = goals scored by a player and n = games played, then goals-per-game = $\frac{x}{n}$.

game. There could be a number of reasons for this such as, for example for games played, it doesn't really matter if a player plays 40 games or 80 games, their scoring pace will not necessarily change. You could make an argument that some players tend to do better during the second half of a season or vice-versa, but due to the sheer volume of players there are, having to sift through that much information was simply not within the realm of possibility due to time constraints. Looking at the scatter plots, we see that shot percentage and number of shots taken have a strong relationship with goals-per-game, but don't concede much in the way of multicollinearity. And so, we had the primary form of our model as follows:

Model 1: $Y = \beta_0 + \beta_1 \cdot Shots + \beta_2 \cdot ShotPercentage$

Where Y is equal to our response variable, goals-per-game. After performing analysis that allowed us to create our linear model (this analysis is found in appendix 1), the following estimates were found:

$$\hat{\beta}_0 = \text{-0.107077}$$
$$\hat{\beta}_1 = 0.001039$$
$$\hat{\beta}_2 = 0.017214$$

And this results in our model now becoming:

Model 1: GPG = -0.107077 + 0.001039$\cdot Shots$ + 0.017214$\cdot ShotPercentage$

Now, something to note before we on. It may be of concern that only two variables are being considered to predict this statistic. According to our model, if for example, player A is a lot better than player B, but player A and player B take the same amount of shots and score at the same rate, then player A and player B will score the same amount of goals. This may lead one to believe that another variable should be added that takes into account the quality of the player. While this suggestion would have merit, the current model can also still be justified. That is because if you consider a statistic like shot percentage, within it is contained the element of skill that would otherwise be seemingly left out. Player A in our scenario, because of being

more skilled, will naturally have a higher shot percentage. So, while it is true that identical stat lines will result in identical goal totals, the likelihood that something like that occurs would actually say something about the difference in skill between those players more so than it would about the quality of this model.

Now that we have our Model 1, we can begin the process of predicting goal scoring total for each team. Recalling and applying (1), we can sum the goal totals for each player. However, before we do this, we must first get a goal total for each player because Model 1 only predicted the goals per game. For simplicity sake, we will say that every player plays 82 games. While this is highly unrealistic, this is the best assumption for us to go by. There is no way to predict whether a particular player would be injured or not with any degree of certainty and so, attempting to do so would be foolish. And, this is not a problem (of injuries) that is particularly hard to deal with because, as mentioned earlier, the versatility of the model allows us to just substitute a player out and in with ease (because again, our model predicts goals per game, NOT total goals scored by a player in a year[9]).

The way I figured out the goals-per-game was through a data set I found in which I used the averages of the players shot percentages and shots throughout their careers (if possible). This was the best way I found to do this because otherwise, it's really not possible to figure out exactly how a player would do. Even though extrapolation is sometimes discouraged, it was the only course of action in this particular case that I could see. After this, a function was written in R that took the observations of my data set as its arguments and outputted the goals-per-game of each player. Players with shot percentages of 0 or total shots less than 10 were removed (because they were throwing off the results for obvious reasons). Finally, a summation was done (again, as in (1)) and the results for each team was computed.

For some final discussions about this model, it has to be said that there will be some points of concern about the results. For example, it was difficult

[9]So, if player A plays 35 games and then gets injured, then you can remove player A's goals for the rest of the year ((85 − 32)*(goals-per-game)) and add in player B's (who substitutes player A) goal total.

to predict how a rookie player will do (like Connor McDavid) because there is no tangible statistics available on such players at the NHL level. Because of restrictions like these, there is an increased variance of our results when compared with what ultimately are the parameters. While there are methods of dealing with this, I left those untouched as it is not in the scope of what we were asked to do here. These factors however, will be explored more deeply in future analysis.

### *2.2. Model 2 - GOALS AGAINST*

The second model (appropriately dubbed "Model 2") is centered around figuring out the amount of goals that a team will allow throughout a year. This required a different approach than the previous model because there is no statistic that is obviously related to the number of teams that a goal allows (unlike in the previous model where total goals scored by each team is a DIRECT result of the goal scoring performances of the players themselves). And so, the decision to look at team statistics (rather than individual players) was made and then when calculating the result, more focused steps could be taken to make it such that the model takes into account the players that a team would have in any given year[10].

The next few lines, as in the previous section, will be dedicated to discussing the various variables considered.

1. **FaceOff Percentage** - This reasoning for this variable was that because winning the faceoff gives control of the puck (to the winning team), the number of chances that could occur would be increased (or decreased) and this could lead to more (or less) goals being scored against a team.

2. **CorsiFor Percentage**[11] - This variable measures whether a team controls play more often than not. So, a percentage about 50 would imply a

[10] A team after all, is just a sum of its parts so it is in no way indicative of the future to know past stats of the team if you do not consider where those stats come from i.e the players

[11] $\frac{CF}{CF+CA}$ where CF and CA (corsiFor and corsiAgainst) are variables that measures the number of shots attempted, blocked, and missed for and against for a player or team (if you take the average of the players)

team that has generally better possession numbers.

3. **FenwickFor Percentage**[12] - This variable measures whether a team controls play more often than not. The difference between this variable and the previous one is that this one does NOT take into account blocked shots. As such, because of how some teams block a lot of shots while others don't, an important distinction has to be made between these two statistics. It is important to include both of these in our model then to take into proper consideration of teams that do not block shots as much and so, incur less of a cost when it comes to what they can do one the ice resulting in other plays to be made and to include corsiFor percentage for those teams that are more dedicated to blocking shots.

4. **Blocked Shots** - This variable was initially included because of the discrepancy between the corsiFor and FenwickFor variables, but as you will soon see, was dropped because it did not do much to help improve the model.

5. **Goalie Save Percentage** - This variable is an obvious one to include because a higher save percentage implies that a team is letting in less goals (provided the shots are the same across the board, which of course is not the case and that is why other variables are needed).

So, initially a model was fitted with a multiple linear regression and the table in figure 3 was produced with contains information about the correlations between the variables. It is clear (from the figure) that faceOff percentage and blocked shots have minimal effect on our model and in fact, probably do nothing more than add unneeded noise to our results. Therefore, the decision to drop those two variables was made.

The next to consider was the two correlated variables: FenwickFor and corsiFor. As you can see, we have a clear problem that will affect our results, but the problem is that, as explained earlier, I did NOT want to get rid of

[12] $\frac{FW}{FW+FA}$ where FW and FA (FenwickFor and FenwickAgainst) are variables that measures the number of shots attempted and missed for and against for a player or team (if you take the average of the players)

these results because removing one of them I felt would cause results that are not accurate enough of a representation of what we should have. Therefore, the decision to carry out a ridge regression was made.

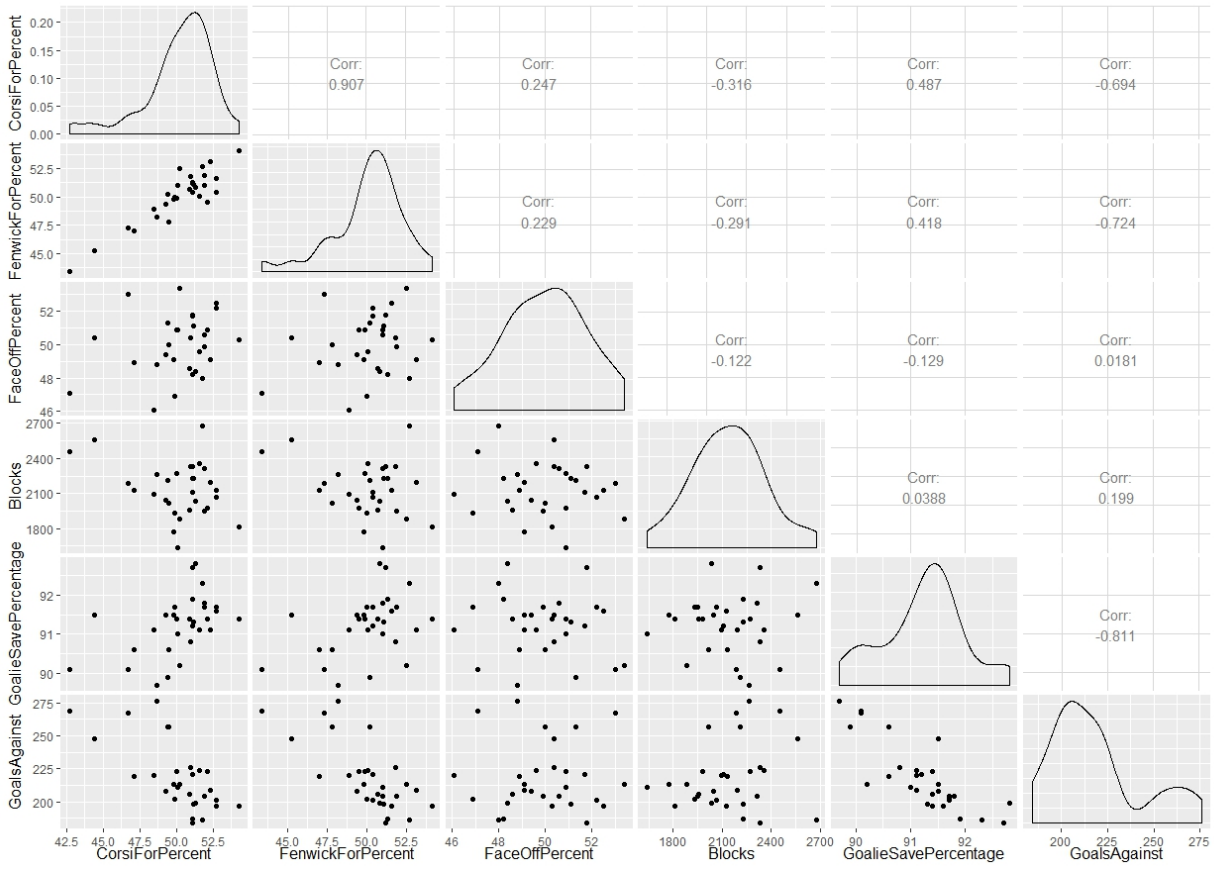


Figure 3: Scatterplot Matrix for Model 2 Variables

When you carry out a ridge regression, you can no longer have the best linear unbiased estimator, but in return, your variance can be much lower than previously. This is very useful because even though there is some slight bias now, the accuracy we make up for it in (less) variance will result in a very good model still and the advantage is that we no longer have to get rid of parameters due to the correlation.

The method of doing a ridge regression is essentially that you run the same model as you do with a normal linear regression, except that you add a constraint. As shown in the graph below, you can see that if the focal point of the level curves is where you would get BLUE (best linear unbiased estimators), then the best estimator (i.e without being unbiased) would be found where the level curves first touch the ridge. We use the L2 NORM here (otherwise, we'd have a lasso), so our ridge will be the intersection of a circle and (the above mentioned) level curves.

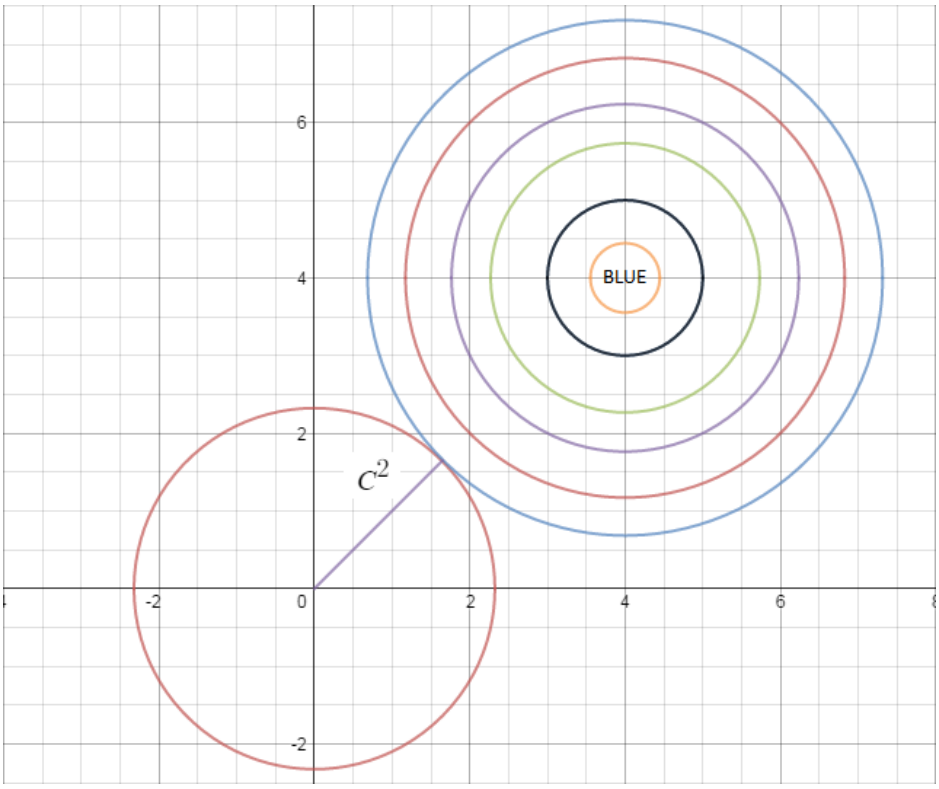


Figure 4: Ridge Regression Topology

The $C^2$ in the above graph is the value of the restriction and will correspond to the lambda below. The axis will be labeled as $\beta_0, \beta_1, ...\beta_n$ because ridge regression is of no use in the linear case. In other words, the topological representation of ridge regression will actually be in multiple dimensions ($n \geq 3$ *dimensions*). Figure 3 is shown for the reader to easily grasp the fundamental idea of what is going on but this 2 dimensional case actually

does not exist. As described earlier, the "BLUE" is where we would find the best parameter estimators provided we did not lift the restriction of unbiasedness. Formally, our equation will be written as follows:

$$\min \vec{\beta} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 \cdot x_i - ... - \beta_n \cdot x_i)^2$$

$\ni$:

$$\beta_0^2 + \beta_1^2 + ... + \beta_n^2 \leq C^2$$

This is of course, now a minimization problem with a constraint and can be solved using lagrangan multipliers set up as:

$$\text{F } (\beta_1, \beta_2, ..., \beta_n) =$$

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 \cdot x_i - ... - \beta_n \cdot x_i)^2 + \lambda \, (\beta_0^2 + \beta_1^2 + ... + \beta_n^2)$$

In the scope of our Model 2, we can rewrite the above as:

$$\text{F (CorsiFor, FenwickFor, SavePercent)} =$$

$$\sum_{i=1}^{n} (Y_i - \beta_0 - CorsiFor \cdot x_i - FenwickFor \cdot x_i - SavePercent \cdot x_i)$$

$$+ \lambda \, (\beta_0^2 + CorsiFor^2 + FenwickFor^2 + SavePercent^2)$$

Normally, the way this problem would be solved is that partial derivatives are taken with respect to each parameter that we have in our model and then each of these functions is minimized[13]. After which, the system of equations is solved. However, a problem arises in this constraint problem because there are not just one set of X's given to us, but rather multiple observations. So, this problem rather then is solved by a computer program after some careful coding.

[13] $\frac{\partial F}{\partial \beta_0} = 0$, $\frac{\partial F}{\partial \beta_1} = 0$, $\frac{\partial F}{\partial \beta_2} = 0$, $\frac{\partial F}{\partial \beta_3} = 0$, $\frac{\partial F}{\partial \lambda} = 0$

To do so, I used the package of R called MASS which contains a function called lm.ridge. This is a very useful function, however I was then tasked with finding a suitable lambda to figure out the rest of the estimates. As an aside, some interesting information on ridge regression is available in section [10]. The method used to find lambda are outlined below:

- First, a function was created such that it could easily create linear models for me with the argument of lambda
- Second, the data set used for this second Model had an observation removed to allow way for cross-validation
- Third, many ridge models were generated for lambdas valued from $(-5, -4, ..., 10)$[14]
- Next, taking the values of corsi, fenwick, and save percentage for the observation I took out, I calculated the expected goals against for each model. This was then compared with the actual result that we wanted
- Because of this previous step, I was able to determine where approximately my lambda would lie (between -3 and -4)
- A function was written that takes into consideration all the possible lambdas such that $|257 - ModelResult|$ is minimized
- Finally, this was run through the R function optim[15] which then finds us this value of lambda

Ultimately, this resulted in a lambda that fit right into the interval where we expected it to fall. Some other things to note, lm.ridge is also not compatible with any of the default R methods for linear models. Because of this, I had to manually write code to calculate residuals.

So, this finally gave way to our parameter estimates which were as follows:

[14] At the suggestion of Dr. Jiguo Cao

[15] similar to a procedure called gradient descent

$\hat{\beta}_0 = 2187.588679$
$\hat{\beta}_1 = -13.971753$
$\hat{\beta}_2 = 9.361983$
$\hat{\beta}_3 = -19039766$

And this results in our model now becoming:

Model 2:

GoalsAgainst =
2187.588679
-13.971753·$CorsiFor$
+ 9.361983·$FenwickFor$
-19.039766·$SavePercentage$[a]

[a]Format explanation: LaTeX being difficult

The next step was now to actually use this model to make predictions on how teams will perform next year as far as the number of goals they allow is concerned. So, in order to this, I had to take a reasonable estimate of what the expected save percentage, corsiFor, and fenwickFor will be. For the goalie save percentage, I just took an average of the goalies expected to start for each team for last season and made small adjustments depending on number of games played. For corsiFor and fenwickFor however, these observations for each team was a little more difficult to get and as such, the estimates could be due for some deviation from what estimates I could have gotten with a more sufficient amount of time. This was remedied a bit by taking using data I found that took an average of corsi and fenwick for each team (this average coming from the players within that team) over the past 2 seasons and used these numbers as my observations. While at first, this may seem to give some obviously skewed numbers, it can easily be adjusted by anyone using the model when the proper rosters for each team for next year have been set. Then, it is only a (tedious) task of finding averages for each player for each team and then ultimately computing the expected goals against using this model. So, to summarize, the predictions made in this paper are relevant only at this immediate time period and are subject to change particularly in the off season. Any problems caused by these changes however, are easily

fixed by just computing new observations.

Finally, after getting the observations as above, the goals against for each team was predicted and is available in the next section. Now, we are ready to move on to the final model.

### *2.3. Model 3 - GOAL DIFFERENTIALS*

Now, in this last subsection, we arrive at the discussion of the last model for this paper in which we will combine the results of our previous two models to predict the winning percentages for each team. This model will be a simple linear regression. However, it will prove to be a very powerful simple linear model that will put into wonderful use the results of our previous two models.

The basic idea for this model has been outlined previously in the introduction[16], but of course, I had to reproduce the results myself to confirm that the strength of the model indeed still stands. Since this is a simple linear model, I had only 2 variables: the response and the regressor. These were chosen based on a problem that I had to deal with which was that the results I would get combining the previous models would result in positive AND negative numbers. Taking something like an absolute value for this number is not useful[17] and simply squaring the number results in a similar problem.

The solution to this problem was to instead look at percentages. For example, instead of taking the difference between goals against and goals for, it would be easier to work with if we compute, of the total number of goals scored while this team played, how many of those goals were theirs. This is summarized in the following formula:

$$\text{Goals For Percentage} = \frac{GoalsFor}{N} \cdot 100,$$

$$\text{where N} = GoalsFor + GoalsAgainst$$

[16] Figure 1

[17] This is because if a team has a -45 goal differential for example, then | -45 | would imply that this team actually did amazing, which is far from what we expect the truth to be

Similarly, to keep the scales the same and to have a comparison that would make sense, we will use a similar formula for winning percentage as follows:

$$\text{Winning Percentage} = \frac{Wins}{N} \cdot 100,$$

where[a] N = 82

[a]The number of games played in a regular NHL season by each team is 82

Thus, we had a suitable scale to make a comparison between the results of our previous two models and the model we will create here. For this model, I found data for each season from the 2007-2008 season to the 2014-2015 season and extracted information on the total goals scored by and against a team, the number of wins by each team, and the total games played by each team. This data was then read into R and using derived variables, I mapped the observations into something we can work with as above. Then a model was fitted which took winning percentage as the response variable and goals for percentage as the predictor. The following picture (figure 5) shows the powerful relationship between these two variables:

The parameter estimates we got then are:

$$\hat{\beta}_0 = \text{-47.851}$$
$$\hat{\beta}_1 = 1.956$$

And, as a result, our model becomes:

$$\text{Winning Percentage} = \text{-47.851} + 1.956 \cdot \text{Goals For Percentage}$$

In section 7, you can find some diagnostic plots to further cement your belief in this model. The next step in this process was putting it all together to get the expected winning percentages for each of the teams in the NHL. After predicting this, we were able to get a relative measure of where each team will end up. There was however a slight problem which does not affect our results, but will however limit our ability to make some inferences. These

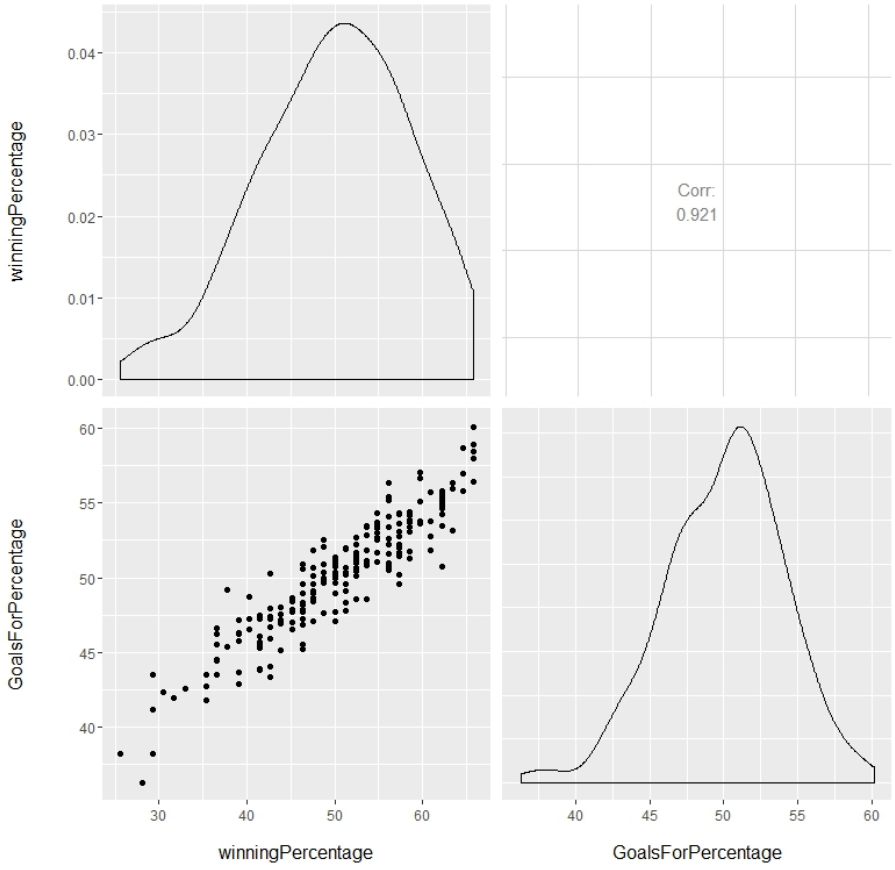


Figure 5: Model 3 Correlation Matrix

problems will be discussed in more detail in section 5.

Thus, we come to a conclusion on this section of the paper. The following section discusses the actual results that the models yield.

## 3. Results

### *3.1. Model 1 - Goals For - Results*

For this model, the results were actually quite well behaved in the end and it resulted in a good model with a nice percentage of the response being

explained by the predictor. What you want to look at in particular here is to see that the predictor variables here are all significant with the response variable.

```
Call:
lm(formula = goalsPerGame ~ Shot + shotPercentage, data = nhlDataLockout)

Residuals:
     Min       1Q   Median       3Q      Max
-0.16756 -0.02911 -0.00639  0.02258  0.37815

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.072e-01  1.515e-03  -70.73   <2e-16 ***
Shot            1.039e-03  9.675e-06  107.44   <2e-16 ***
shotPercentage  1.721e-02  1.464e-04  117.56   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04186 on 5111 degrees of freedom
  (1030 observations deleted due to missingness)
Multiple R-squared:  0.8814,    Adjusted R-squared:  0.8813
F-statistic: 1.899e+04 on 2 and 5111 DF,  p-value: < 2.2e-16
```

Figure 6: Summary Results for Model 1

Another thing to note is the $R^2$ which, as mentioned earlier, says that this model explains 88 percent of our predicted responses and in fact, as you will see in the next figure, the other 12 percent is lost mostly in the upper tail.

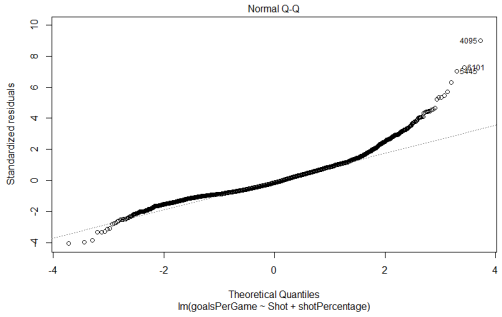


Figure 7: Normal QQ Plot: Model 1

Here, you can see that this model has a very nice fit through most of the

center of the distribution, but has a heavy upper tail. This indicates that there are more values at the right end of the normal distribution then you would normally expect. In other words, there is a positive skew. This is due to the fact that there are some players who are very skilled and these players are usually found at the far end of the distribution. Because these players (snipers) have a better ability to shoot and are given the opportunity to play more in situations such as power plays, they end up with more goals than we would predict them to. Since I did not take into consideration such a statistic (power play time just being one of them), my model suffers from this deviation from normality!

Continuing on in our model 1 results, the next thing to account for was multicollinearity between the variables. The results and appropriate plots for this can be found in section 7, but to summarize, after doing some due diligence to account for an oddball data, it turned out that none of the data was particularly correlated in any way. This was evident already in the information from figure 2, but to make it more concrete, ANOVA results will be available as well.

Looking at the residuals in figure 8, we can see that there is a slight curvature which could be a little alarming, however, again it is explained by what I said earlier. There is a systematic effect on the data going on caused by coaching decisions that was not taken into account. In the end, it does not matter particularly because even if the model underestimates the results of top end players, the significance of this underestimate is not significant in the overall results because there are so few players in the league who score as such an elite pace.

Now, on to the actual goals predicted. This was done as described earlier in the previous section and the results were quite interesting. In fact, it seems that the opposite happened as to what was expected and the goals total for each team was more than expected. This of course, is not a problem because of mappings made that made these overestimations irrelevant, however it did speak a lot to how much room there is for an improved model in future iterations. Some other teams though, did exactly around what you would expect them to do via your instinct. These caused quite a bit of difference between the upper end teams and the lower end teams. More of these possible improvements will be explored in section 5 of this paper.

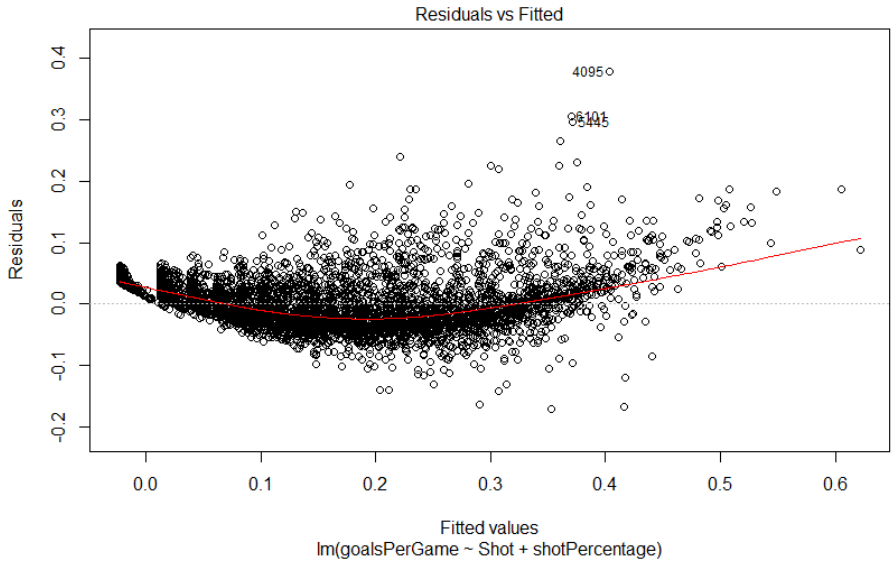


Figure 8: Residuals: Model 1

Finally, you can see the results below in figure 9. The actual goal numbers for each team is available in section 7.

Some things to consider that make these results a bit more plausible is that they are rather consistent with how teams are doing today in the NHL. In fact, the top 5 teams are the general consensus for cup favorites and bottom 5 are all in the race for the NHL draft lottery. This is quite reassuring that the model indeed has given us results with some degree of accuracy because the model considered the rosters of these teams today and so the results you would expect would be similar. It is also quite easy to adjust these results for next year. It's as simple as an analysis of an updated roster.

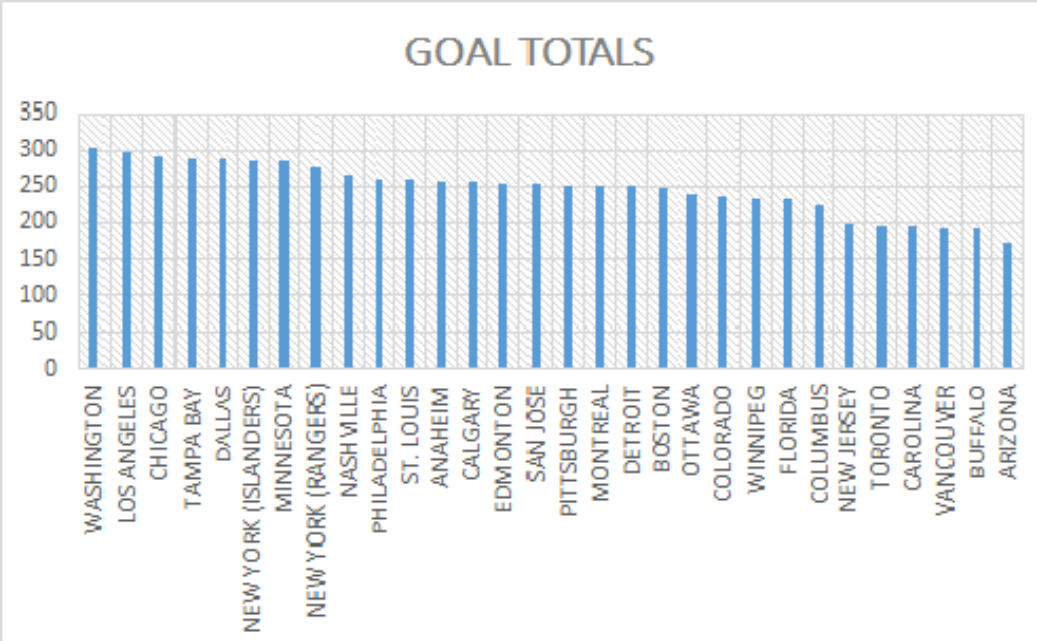


Figure 9: Goals For

### *3.2. Model 2 - Goals Against - Results*

The second model, because of the smaller data set, caused less problems ultimately by the end because it was just easier to troubleshoot if any inconsistencies showed up. First, just to have visual representation of what our optimal lambda would look like graphically, we can have a look at the ridge trace in figure 10. Even though we used an R function (optim) to figure out the value of lambda, it is still nice to see that the graphical results are consistent as well.

Continuing on, figure 12 and figure 13 (page 25) show the residuals and normal QQ plot for our model. While, the residuals do not seem to be spread out VERY nicely, they do maintain some level of variance when you consider all the points. However, there is clearly room for improvement, which will be discussed more in depth in section [5].

As far as the model is concerned in terms of it's ability to accurately predict our results, an $R^2$ of around 84 percent was achieved, however this

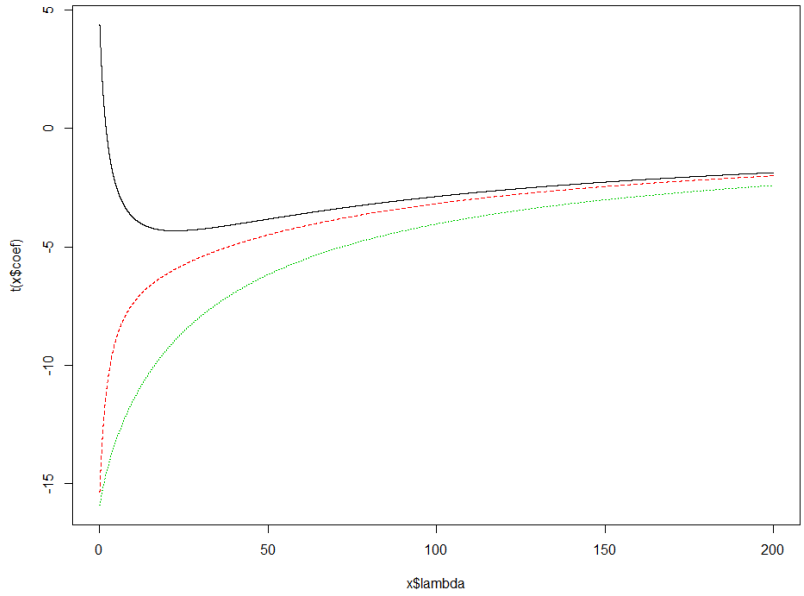


Figure 10: Ridge Trace

was done without the ridge model and we would expect that despite their being bias, we would get a better predictive ability from our model. In fact, the bias is kind of a irrelevant point because ultimately, every team will be affected by this same bias and what we are looking for in the end is ordinal rankings. The winning percentages can be scaled relative to previous years to account for anything lost in the bias.

Finally, to have a look at what our results produced, we can see them in the table on page 24. Some interesting things to note is that relative to model 1, our results here are going to produce quite a big differential between the goals for and goals against for some teams. Because of this, some of our results are going to produce numbers that don't necessarily add up. This has no effect on the ordinal rankings as mentioned in the abstract, but the cardinal ones will need to be scaled. As far as just looking at the relativity of the results, since information was used for recent rosters of teams, we got an expected result overall.

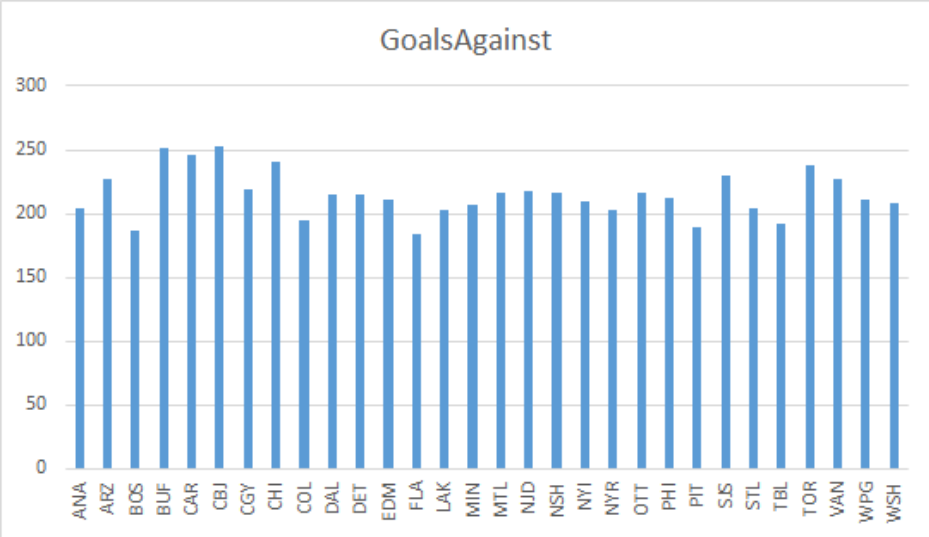


Figure 11: Results - Model 2

There will be more charts and graphs of importance available in the later sections corresponding to each of these sections here. Now, we move on to the results for our final section.

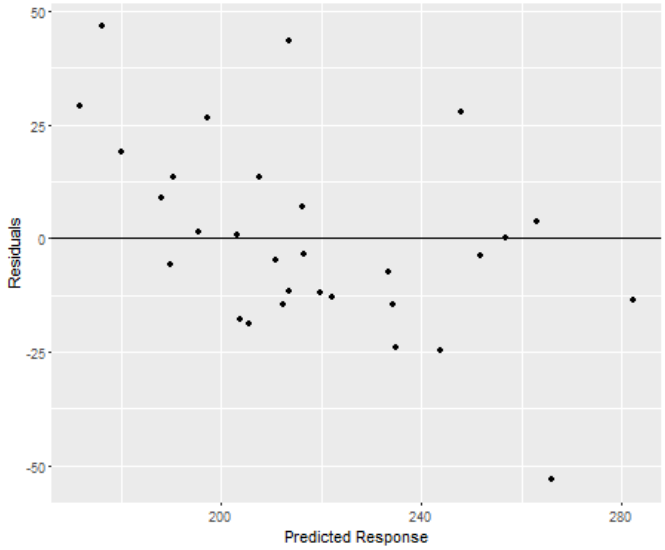


Figure 12: Residuals - Model 2

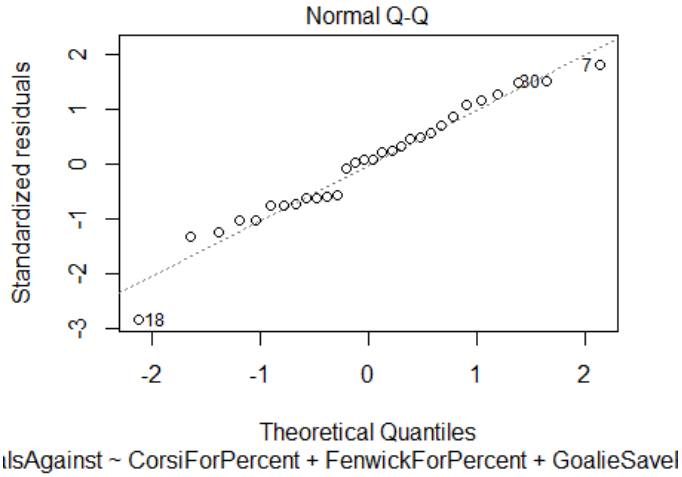


Figure 13: Normal QQ - Model 2

*3.3. Model 3 - Goal Differentials - Results*

This last model, which focused on whether winning percentage is influenced by goals for percentage[18], had a very strong relationship. The following are some diagnostic plots and the strength of the relationship (as measured by $R^2$ and the residuals).

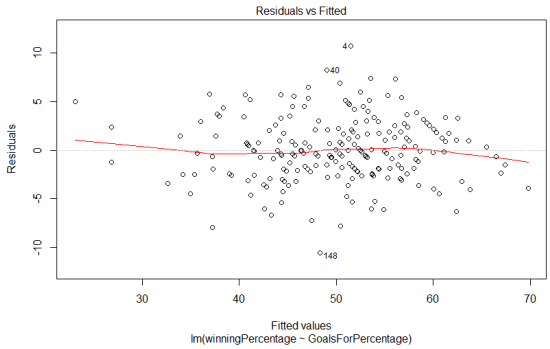


Figure 14: Residuals - Model 3

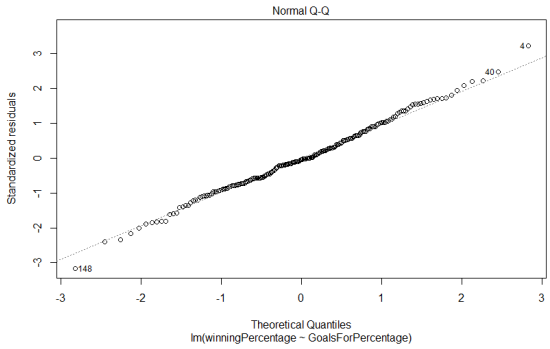


Figure 15: Normal QQ - Model 3

[18]Which is just a modification of the combination of our previous two results

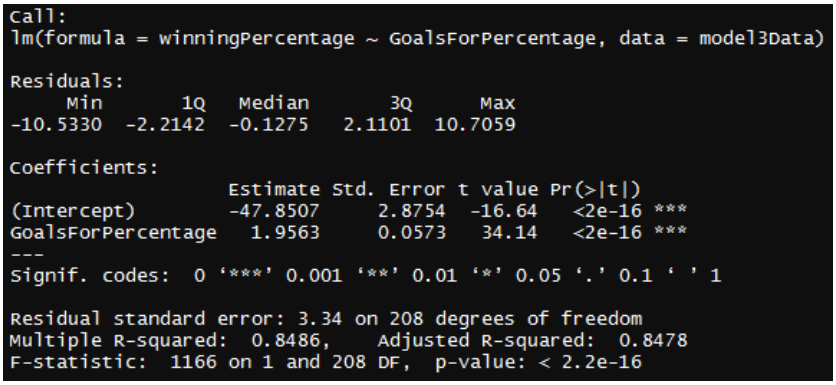


Figure 16: Summary - Model 3

As you can see in figure 14, 15, and 16, there is clearly evidence for a linear relationship between these two variables. The residuals in figure 14 show that they are seemingly random with no distinct pattern without too much variance. In figure 15, the points fit nearly perfectly which is evidence of normality thus satisfying the conditions for the error of a multiple linear regression model[19]. And finally, figure 16 gives us reassurance in our model by just looking at the $R^2$ to see that indeed most of our values of winning percentage can be explained by the goals for percentage of a team.

Moving on, using our model 3, I predicted the winning percentages for each team in accordance to our mapped goals for percentages. The results are shown in figure 17. This is ultimately what we desired. For exact percentages, you can check section [9]. These are of course, subject to change and all of that will be dealt with according to day by day results of what happens next season, but as of now, this seems to be a good representation of what we should expect!

[19]These assumptions being: $\mathrm{E}(\epsilon) = 0$, $\epsilon$ Normal, and that $\mathrm{Var}(\epsilon) = \sigma^2 * \mathrm{I}$

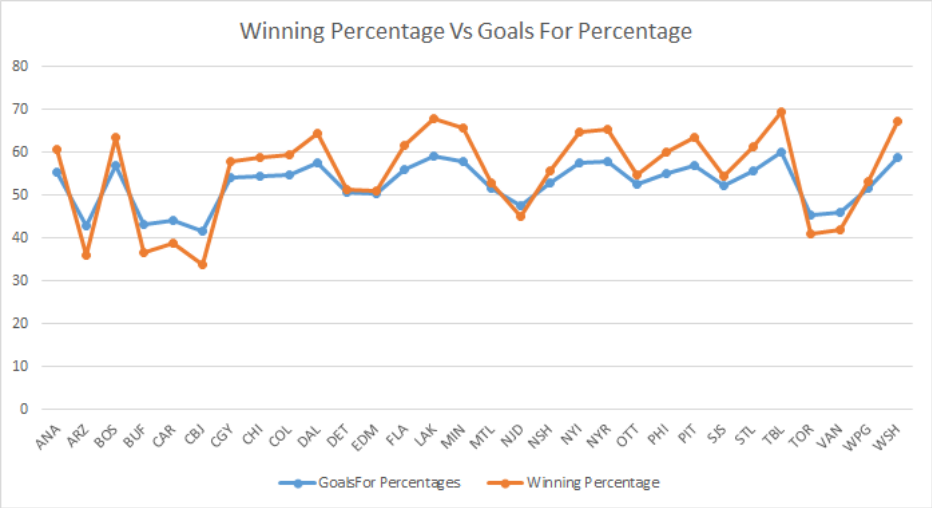


Figure 17: Expected Winning Percentages

Thus, we conclude the results section of this paper.

## 4. Conclusions

As far as my results were concerned, I achieved exactly what I set out to do. The strength of each individual model showed to be quite powerful and as a result, we can say we with confidence that there is at least some respectable level of merit with our final outcomes. An interesting couple of notes to talk on could be that the play offs teams, as outlined in figure 18, left out of a couple of key teams that you would expect to be in while also having some strange rankings as far as where teams would place. For example, you can see that Chicago, a very elite team, would come into the playoffs as one of the lower seeds. This could be concerning, but as another example, the LA Kings in 2012 came in as the 8th seed and won the stanley cup. The point being that it does not matter in some respects to our results whether they are 5th or 15th because either way, they are in. However, a concern will also be discussed in the next section regarding this.

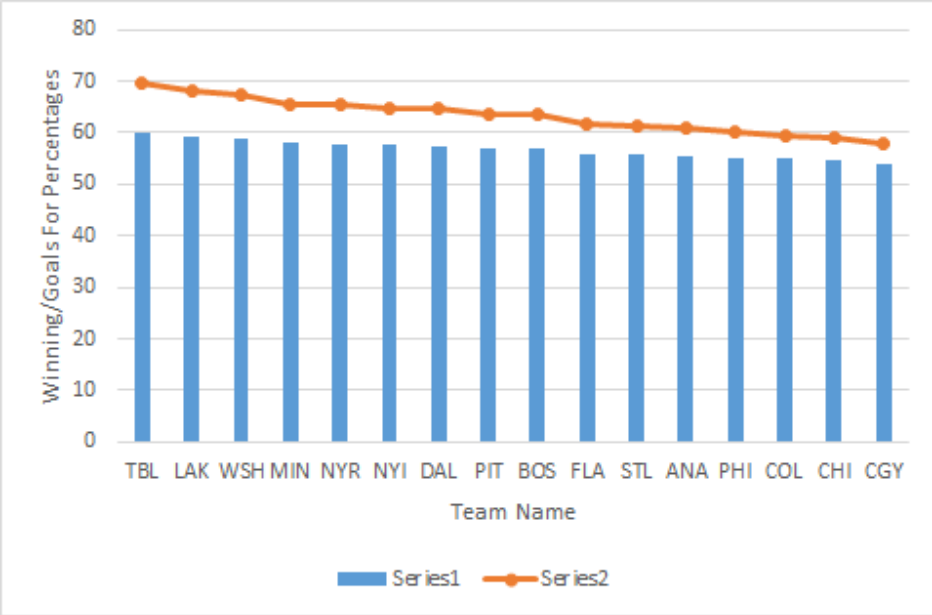


Figure 18: Playoff Teams

Despite these anomalies, in all other respects, the models seem to be well constructed in terms of statistical methodology. The accuracy of the predictions however, will not be known until a later time.

## 5. Discussions

There are quite a few interesting discussions to be had. I'll start off with what was mentioned in the previous section. One of the other goals of this paper was to predict the winning percentage for a team in order to get cardinal rankings. In the example of Chicago, we discussed why the ordinal case does not matter, but indeed the cardinal one does. This is because in the cardinal case, we are getting an exact result, i.e winning percentage. If a person was interested in making bets based on this model over a season, then there are some kinks that would need to be fixed to make this work. In fact, the probably seems to stem from Chicago's ability to stop goals from going in. Then, does there exist some problem in the model? Maybe the ridge model is not the best? How about a principal component analysis instead? That would be quite an intriguing analysis to attempt and maybe looking at

things retrospectively, we can figure out whether or not a PCA is a better way of running the model.

Another discussion to be had is that of the variables and data that were used for the first analysis. First of all, the data set was large and had lots of strange statistics inside that were accounted for to a large degree, but I am quite certain that more could have been done to make the results even better. For example, new variables could have been created that took into consideration the goals scored by players who were scoring at outlier-esque paces. These results interfere with the models ability to be a good predictor! The other thing to discuss is the methods. For hockey, sometimes it can be difficult to NOT find the data, but rather make it work. In fact, it can be quite tedious to sift through data to find what you want and put it into a new file. This has a high opportunity cost and one that I could not afford. As a consequence of this, the number of variables that predicted the goals per game had to be limited to what was available. Other variables that I would have considered would have been things such as power play time for each, time on ice, the situations that players play in and even something like the division of the player. These are all variables that would be considered for future papers.

There are many other things that I can hypothesize on, but I will end this section here on the assumption that enough has been said!

## 6. References


- "Hockey Analytics." Hockey Analytics. N.p., n.d. Web. Mar. 2016.
- "Hockey-Reference.com." Hockey-Reference.com. N.p., n.d. Web. Mar. 2016.
- "Hockey Abstract." Hockey Abstract. N.p., n.d. Web. Mar. 2016


## 7. Model 1 Plots/Graphs

The first few figures here showcase the multicollinearity. I know age was dropped as a variable, but I ended up doing some analysis on it anyway. For anyone curious, you can see those results as well.

```
Call:
lm(formula = Shot ~ Age, data = nhlData1967.2014)

Residuals:
   Min     1Q Median     3Q    Max
-83.75 -51.28 -13.57  37.53 414.53

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 115.63305    5.32063  21.733   <2e-16 ***
Age          -0.09399    0.19095  -0.492    0.623
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 64.17 on 5343 degrees of freedom
  (2509 observations deleted due to missingness)
Multiple R-squared:  4.535e-05, Adjusted R-squared:  -0.0001418
F-statistic: 0.2423 on 1 and 5343 DF,  p-value: 0.6226
```

Figure 19: Multicollinearity: Shots Vs Age

```
Call:
lm(formula = ShotPercentage ~ Age, data = nhlDataLockout)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8057 -1.9839 -0.2289  1.7809  5.2221

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.721359   0.259738  41.278   <2e-16 ***
Age          0.002829   0.009337   0.303    0.762
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.387 on 2999 degrees of freedom
  (3143 observations deleted due to missingness)
Multiple R-squared:  3.061e-05, Adjusted R-squared:  -0.0003028
F-statistic: 0.09182 on 1 and 2999 DF,  p-value: 0.7619
```

Figure 20: Multicollinearity: Shot Percentage Vs Age

```
Call:
lm(formula = ShotPercentage ~ Shot, data = nhlDataLockout)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0855 -1.9116 -0.1059  1.7868  4.9573

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.178615   1.161683   8.762   <2e-16 ***
Shot         0.004937   0.006706   0.736    0.462
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.351 on 583 degrees of freedom
  (5559 observations deleted due to missingness)
Multiple R-squared:  0.000929,  Adjusted R-squared:  -0.0007847
F-statistic: 0.5421 on 1 and 583 DF,  p-value: 0.4619
```

Figure 21: Multicollinearity: Shots Vs Shot Percentage

Next is a look at a graph plotting shot versus shot percentage (figure 22). Clearly, you can see that there is no discernible pattern here.

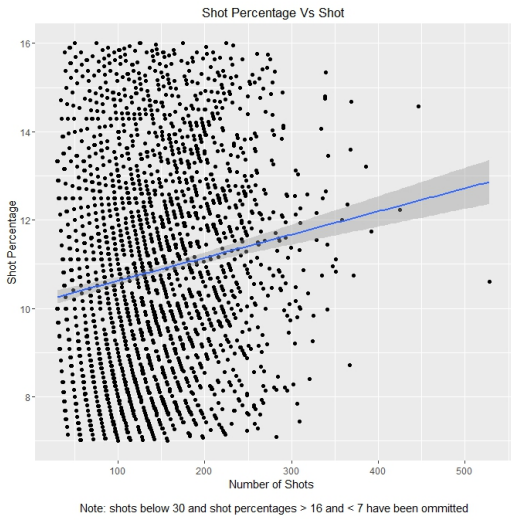


Figure 22: Plot: Shot Vs Shot Percentage

And finally, there are some diagnostic results for this model and interesting 3D plots. Some of them may be repeated from earlier. They are all present here for completeness and convenience. In the end of it all, you will find the actual goals for that I predicted for each team and that will conclude this section.

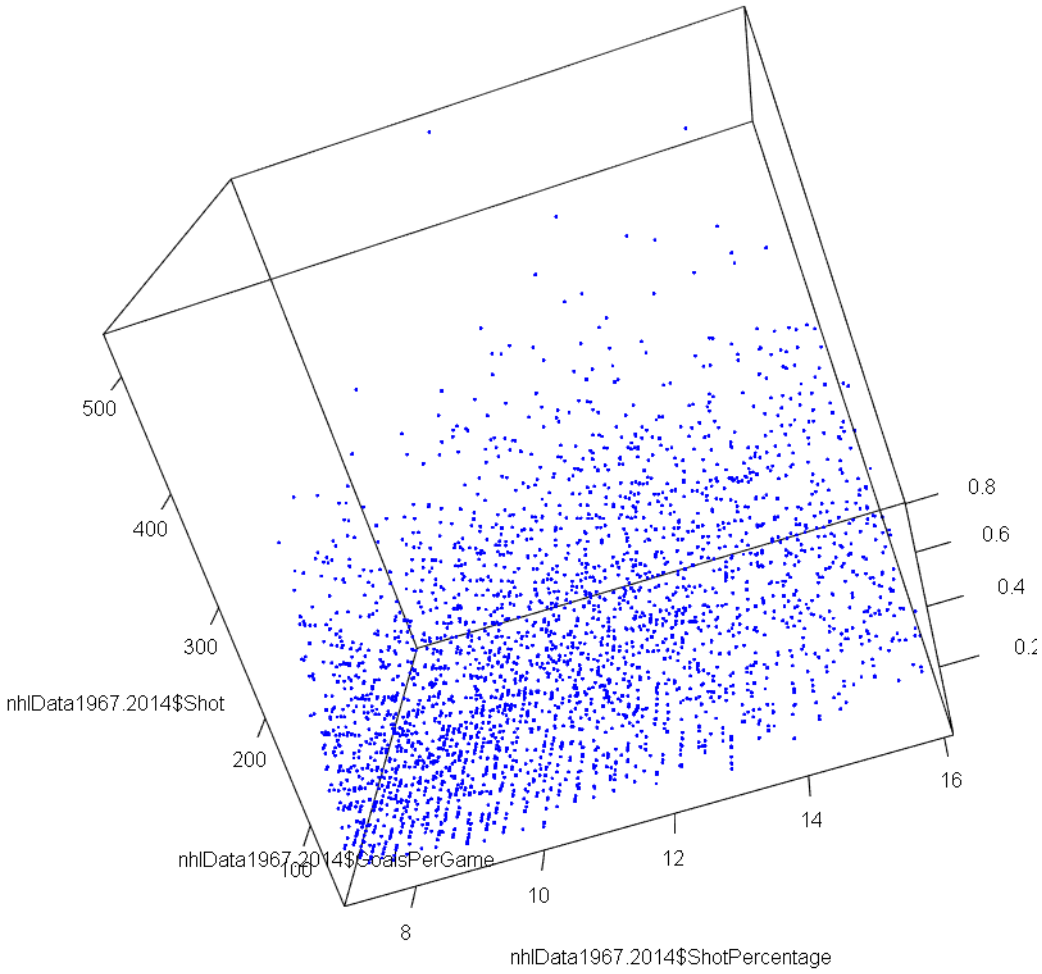


Figure 23: Shots and ShotPercentage

This figure looks at the points plotted of the shot percentage versus the shots. As you can see, there is clearly not any real discernible pattern between these two variables. This is consistent with the correlation we got for these two variables. The next two are similar in nature.

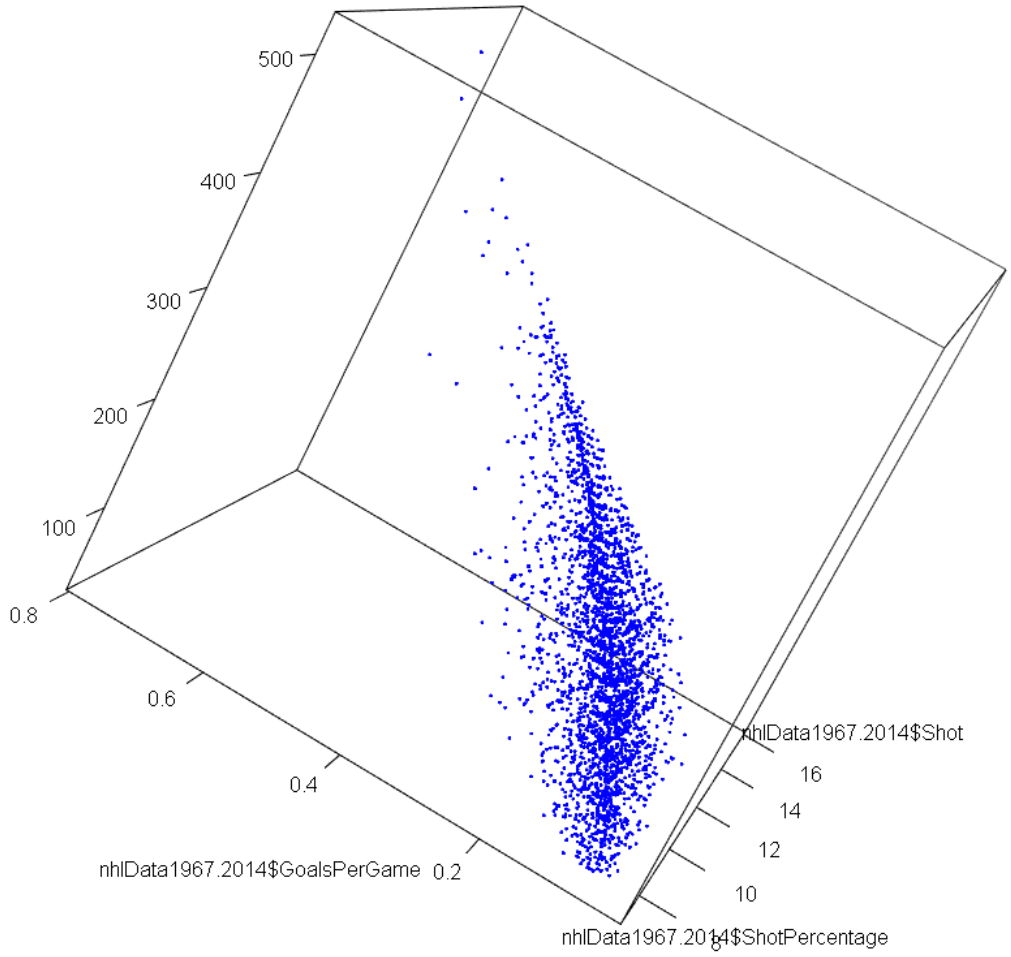


Figure 24: Shots Vs Goals Per Game

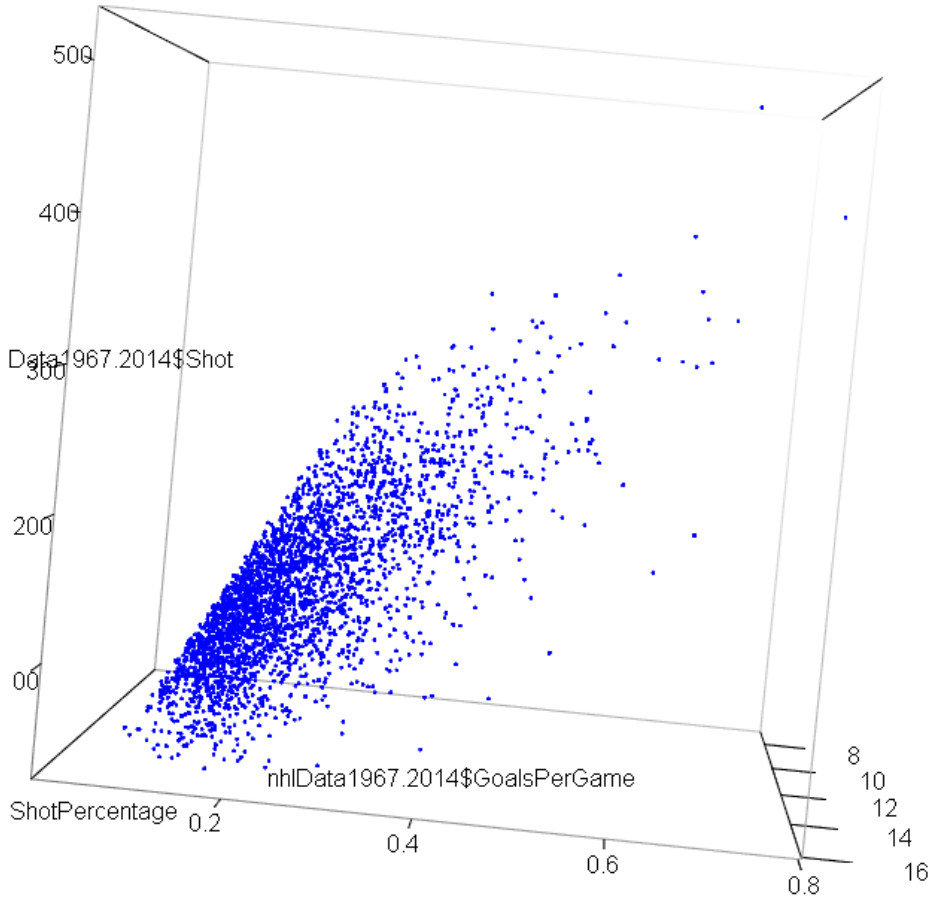


Figure 25: Shot Percentage Vs Goals Per Game

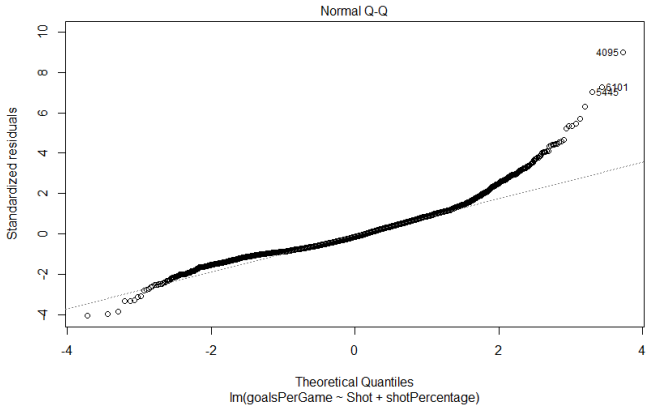


Figure 26: Normal QQ

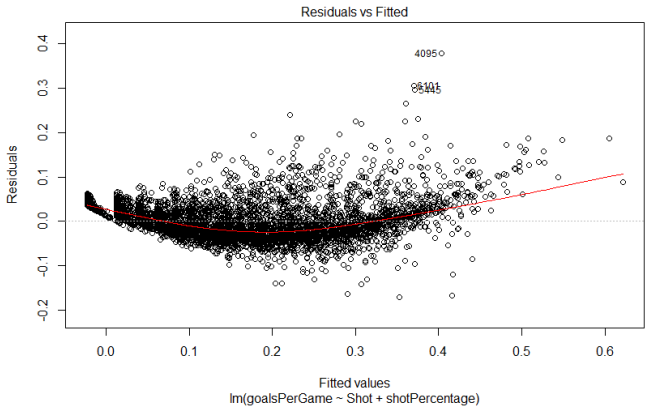


Figure 27: Residuals

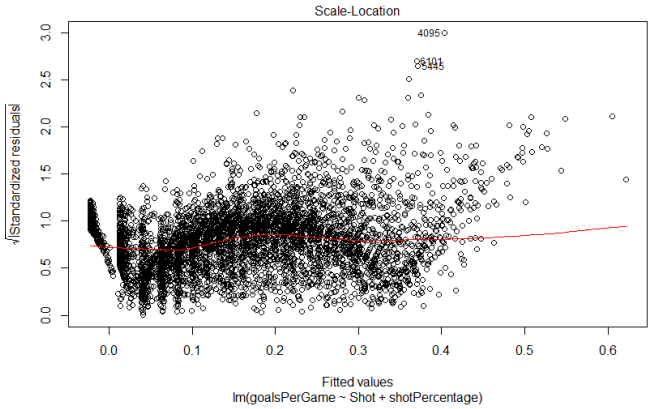


Figure 28: Scale

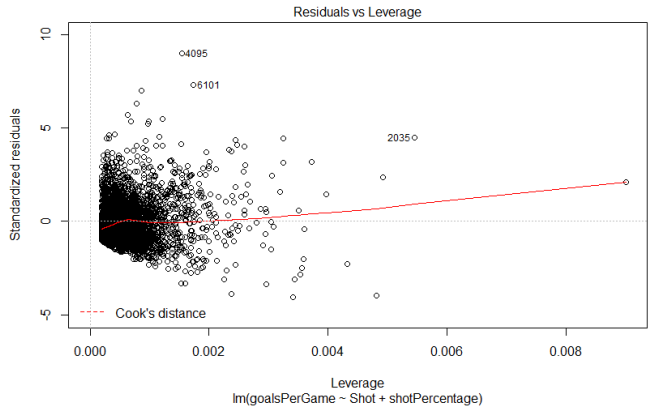


Figure 29: Leverage

| TEAM NAME | GOALS TOTAL |
|---|---|
| WASHINGTON | 302.2576 |
| LOS ANGELES | 298.7253 |
| CHICAGO | 292.0194 |
| TAMPA BAY | 289.0998 |
| DALLAS | 287.8628 |
| NEW YORK (ISLANDERS) | 287.0131 |
| MINNESOTA | 286.6679 |
| NEW YORK (RANGERS) | 278.3809 |
| NASHVILLE | 264.3797 |
| PHILADELPHIA | 261.174 |
| ST. LOUIS | 259.9754 |
| ANAHEIM | 257.5363 |
| CALGARY | 255.4695 |
| EDMONTON | 255.2873 |
| SAN JOSE | 253.0385 |
| PITTSBURGH | 252.3376 |
| MONTREAL | 252.096 |
| DETROIT | 251.1424 |
| BOSTON | 246.6236 |
| OTTAWA | 238.9788 |
| COLORADO | 237.5999 |
| WINNIPEG | 234.7603 |
| FLORIDA | 234.127 |
| COLUMBUS | 223.4844 |
| NEW JERSEY | 197.5006 |
| TORONTO | 196.6678 |
| CAROLINA | 195.102 |
| VANCOUVER | 193.5533 |
| BUFFALO | 192.3312 |
| ARIZONA | 170.5993 |

Figure 30: Predicted Goals For

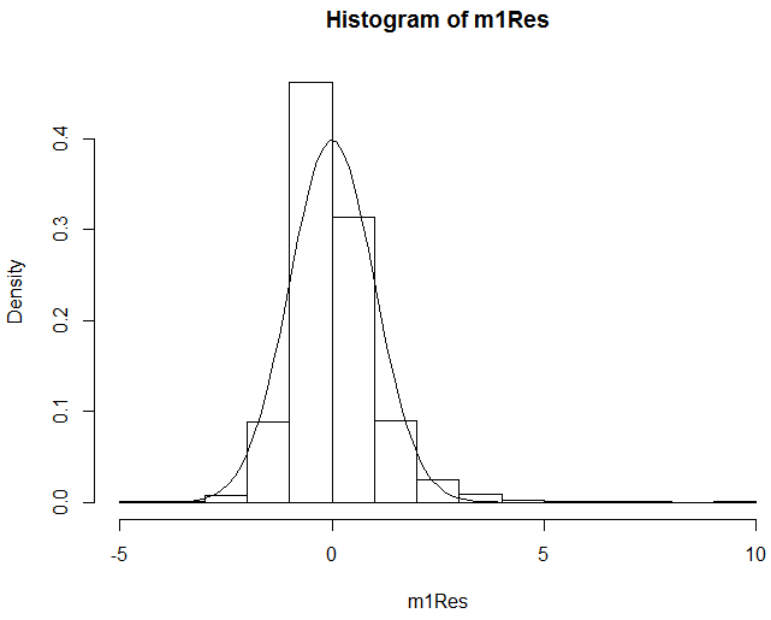


Figure 31: Histogram Of Residuals - Model 1

```
Start:  AIC=-25482.55
goalsPerGame ~ 1

                  Df Sum of Sq    RSS    AIC
+ Shot             1    46.146 33.584 -27556
+ ShotPercentage   1    44.400 35.330 -27282
+ GP               1     4.006 75.724 -23149
+ Age              1     0.114 79.617 -22877
<none>                         79.730 -22871

Step:  AIC=-28228.67
goalsPerGame ~ Shot

                  Df Sum of Sq    RSS    AIC
+ ShotPercentage   1   23.2977 10.286 -33969
+ GP               1    7.1681 26.416 -28856
+ Age              1    0.0728 33.511 -27566
<none>                         33.584 -27556

Step:  AIC=-32877.09
goalsPerGame ~ Shot + ShotPercentage
```

Figure 32: AIC

## 8. Model 2 Plots/Graphs

Similar to the structure of the previous section, we begin with the multicollinearities between the variables.

```
Call:
lm(formula = GoalieSavePercentage ~ CorsiForPercent, data = goalsAgainstData)

Residuals:
     Min       1Q   Median       3Q      Max
-1.31449 -0.42980 -0.01404  0.31602  1.39432

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     83.68730    2.56404  32.639  < 2e-16 ***
CorsiForPercent  0.15046    0.05104   2.948  0.00639 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6704 on 28 degrees of freedom
Multiple R-squared:  0.2368,    Adjusted R-squared:  0.2096
F-statistic: 8.689 on 1 and 28 DF,  p-value: 0.006391
```

Figure 33: Multicollinearity: Corsi Vs Save Percentage

```
Call:
lm(formula = FenwickForPercent ~ GoalieSavePercentage, data = goalsAgainstData)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2516 -0.9525  0.1214  1.1789  3.8147

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          -64.5740    47.0671  -1.372   0.1810
GoalieSavePercentage   1.2567     0.5159   2.436   0.0215 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.095 on 28 degrees of freedom
Multiple R-squared:  0.1749,    Adjusted R-squared:  0.1454
F-statistic: 5.934 on 1 and 28 DF,  p-value: 0.02147
```

Figure 34: Multicollinearity: Fenwick Vs Save Percentage

```
Call:
lm(formula = FenwickForPercent ~ CorsiForPercent, data = goalsAgainstData)

Residuals:
     Min       1Q   Median       3Q      Max
-2.20033 -0.48049  0.07259  0.42364  2.40034

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.80813    3.71850    2.10   0.0449 *
CorsiForPercent  0.84246    0.07402   11.38 5.14e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9723 on 28 degrees of freedom
Multiple R-squared:  0.8223,    Adjusted R-squared:  0.8159
F-statistic: 129.5 on 1 and 28 DF,  p-value: 5.142e-12
```

Figure 35: Multicollinearity: Fenwick Vs Corsi

Next, we have the results for our lambda results which we used to confirm the lambda we got through optimization and some diagnostic plots.

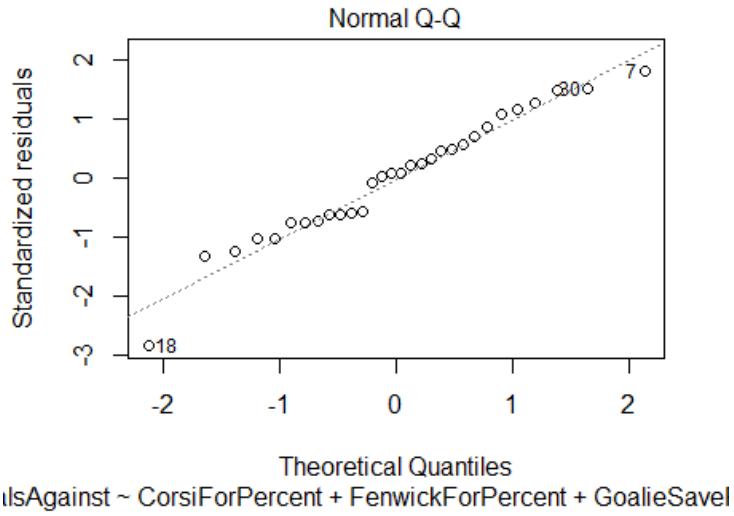


Figure 36: Normal QQ

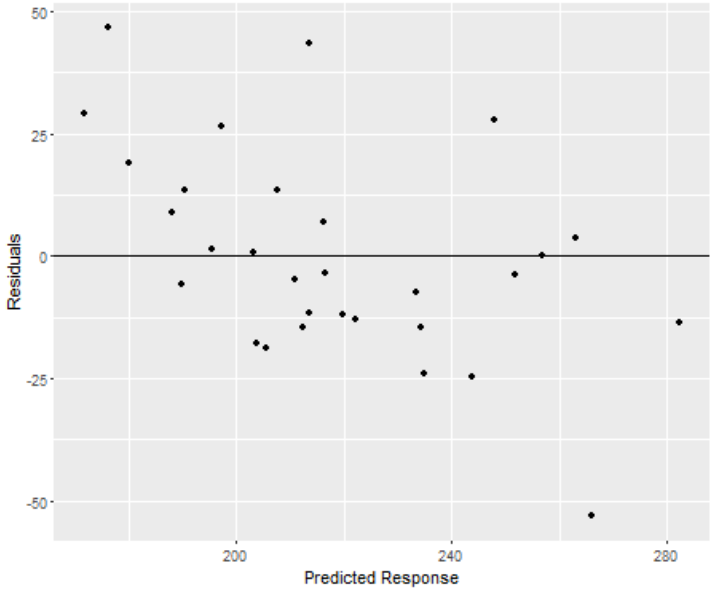


Figure 37: Residuals

And the $R^2$ results

```
Response: GoalsAgainst
                   Df Sum Sq Mean Sq F value    Pr(>F)
CorsiForPercent     1 8968.9  8968.9 79.6175 2.152e-09 **
FenwickForPercent   1  939.5   939.5  8.3399  0.007713 **
GoalieSavePercentage  1 5777.0  5777.0 51.2824 1.322e-07 **
Residuals          26 2928.9   112.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
Residual standard error: 10.78 on 25 degrees of freedom
Multiple R-squared:  0.8439,    Adjusted R-squared:  0.819
F-statistic:  33.8 on 4 and 25 DF,  p-value: 9.514e-10
```

Figure 38: Summary: Model 2

```
# Testing r1
predictRidge(r1, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 253.8226

# Testing r2
predictRidge(r2, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 254.4329

# Testing r3
predictRidge(r3, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 265.1881

# Testing r4
predictRidge(r4, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 234.8937

# Testing r5
predictRidge(r5, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 241.4532

# Testing r6
predictRidge(r6, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 242.2154

# Testing r7
predictRidge(r7, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 242.0963

# Testing r8
predictRidge(r8, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 241.7144

# Testing r9
predictRidge(r9, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 241.2394

# Testing r10
predictRidge(r10, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 240.7388

# Testing r11
predictRidge(r11, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 240.2247

# Testing r12
predictRidge(r12, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 239.2393

# Testing r13
predictRidge(r13, corsiDallas, fenwickDallas, savePercentDallas)
# Result: 238.7714
```

Figure 39: Ridge Lambda Predictions

And now finally, the results for goals against for each team:

| Team | GoalsAgainst |
|---|---|
| ANA | 204.542307 |
| ARZ | 227.733563 |
| BOS | 186.594526 |
| BUF | 250.862408 |
| CAR | 246.107287 |
| CBJ | 252.427099 |
| CGY | 218.431486 |
| CHI | 241.007289 |
| COL | 195.009803 |
| DAL | 215.104329 |
| DET | 214.270324 |
| EDM | 210.338086 |
| FLA | 183.613011 |
| LAK | 203.155254 |
| MIN | 206.637772 |
| MTL | 216.750839 |
| NJD | 217.338039 |
| NSH | 215.732443 |
| NYI | 210.029605 |
| NYR | 203.208784 |
| OTT | 215.95248 |
| PHI | 211.611427 |
| PIT | 189.485353 |
| SJS | 229.980654 |
| STL | 203.617605 |
| TBL | 191.657378 |
| TOR | 237.460083 |
| VAN | 227.066059 |
| WPG | 211.442673 |
| WSH | 208.808303 |

Figure 40: Goals Against Predictions

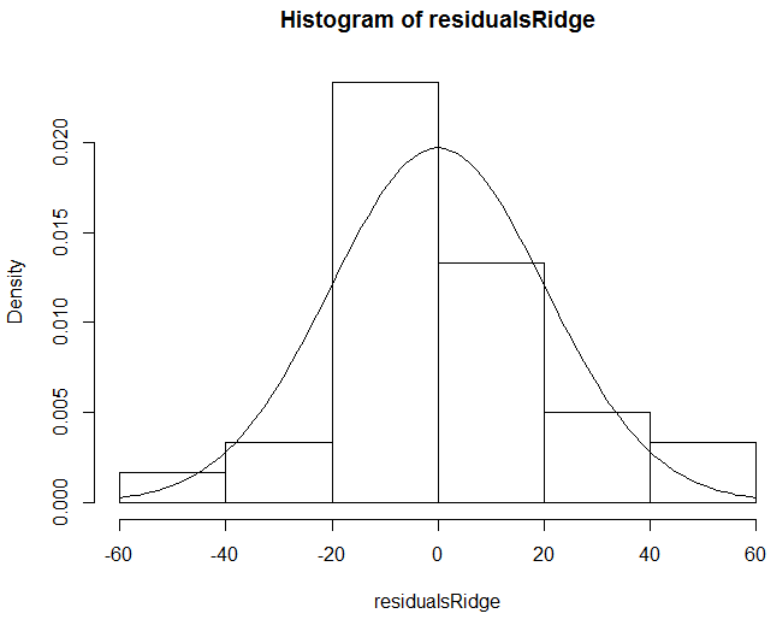


Figure 41: Histogram Of Residuals - Model 2

## 9. Model 3 Plots/Graphs

This model was a simple linear model, so we did not have any multicollinearity going on. The following figures are some diagnostic plots for this model and some extras for you to look at for inferences.

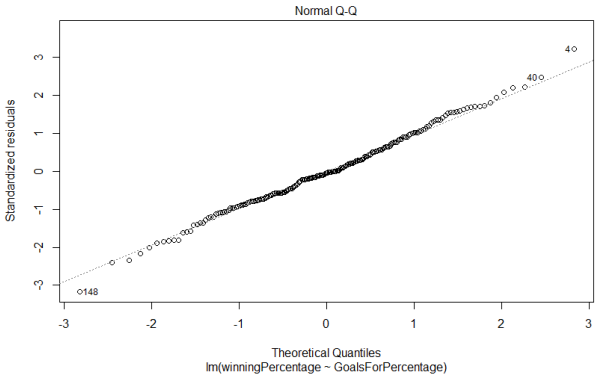


Figure 42: Normal QQ

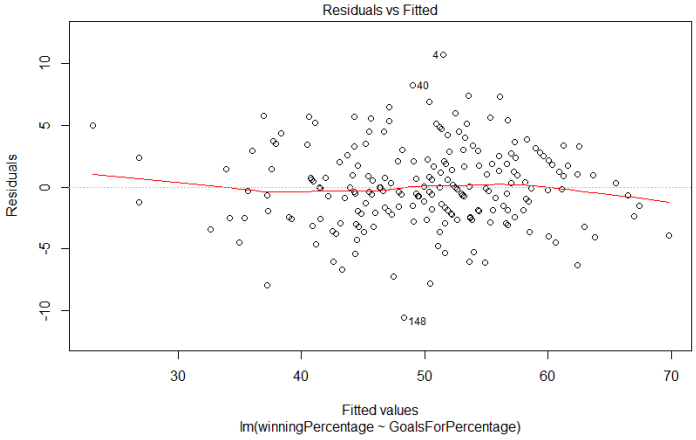


Figure 43: Residuals

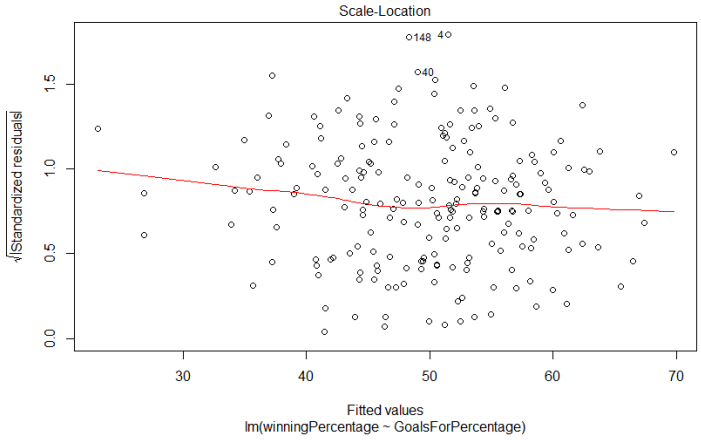


Figure 44: Scale

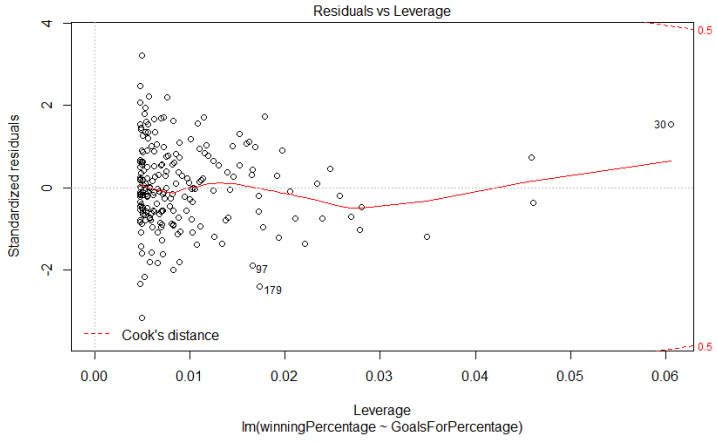


Figure 45: Leverage

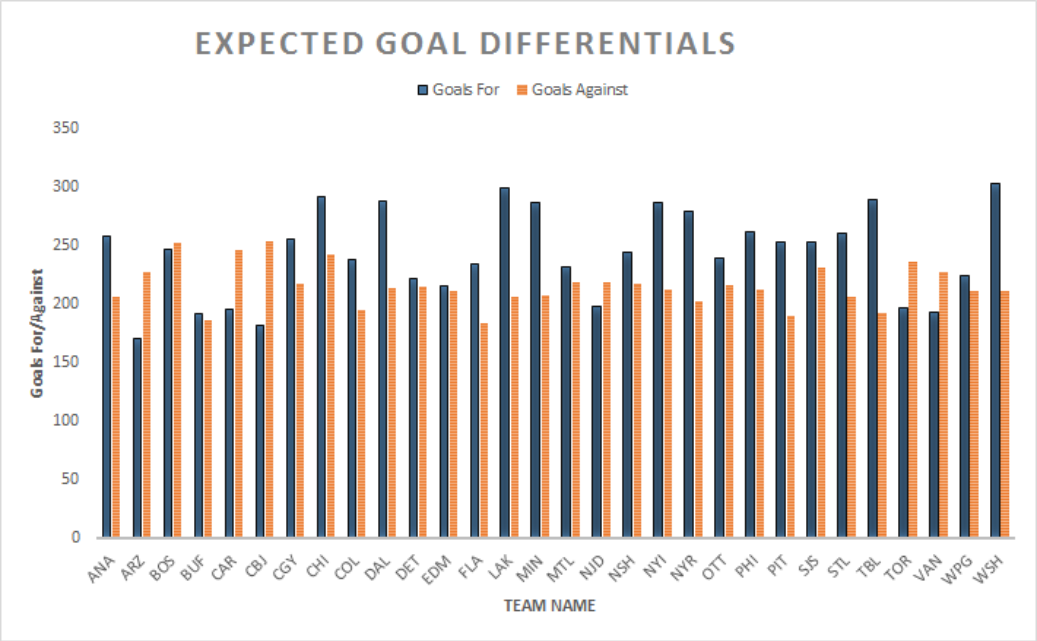


Figure 46: Goal Differentials

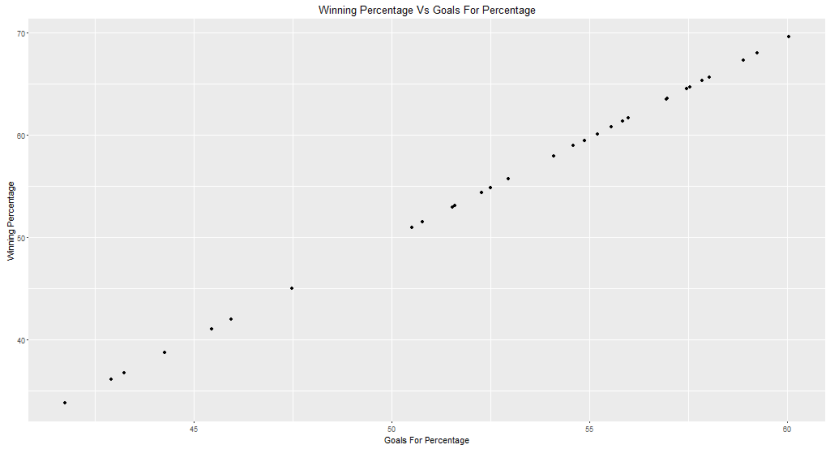


Figure 47: Plot of Predicted Values Using Model 3

| Team | Goals For Percentage | Predicted Winning Percentage |
|---|---|---|
| ANA | 55.56529819 | 60.83472326 |
| ARZ | 42.9210921 | 36.10265614 |
| BOS | 56.94835993 | 63.53999202 |
| BUF | 43.24017755 | 36.7267873 |
| CAR | 44.270931 | 38.74294104 |
| CBJ | 41.75356094 | 33.81896519 |
| CGY | 54.09881672 | 57.9662855 |
| CHI | 54.60189593 | 58.95030844 |
| COL | 54.87781262 | 59.49000148 |
| DAL | 57.45953554 | 64.53985152 |
| DET | 50.79161038 | 51.49738989 |
| EDM | 50.51171408 | 50.94991275 |
| FLA | 55.99052036 | 61.66645782 |
| LAK | 59.2496193 | 68.04125536 |
| MIN | 58.03439981 | 65.66428604 |
| MTL | 51.53482788 | 52.95112333 |
| NJD | 47.48851523 | 45.03653578 |
| NSH | 52.96583733 | 55.75017782 |
| NYI | 57.53849295 | 64.69429221 |
| NYR | 57.86054531 | 65.32422662 |
| OTT | 52.51517356 | 54.86867948 |
| PHI | 55.20180074 | 60.12372224 |
| PIT | 56.98238567 | 63.60654638 |
| SJS | 52.28211968 | 54.4128261 |
| STL | 55.83722925 | 61.36662041 |
| TBL | 60.0537249 | 69.6140859 |
| TOR | 45.45965312 | 41.06808151 |
| VAN | 45.94432196 | 42.01609376 |
| WPG | 51.6007893 | 53.08014386 |
| WSH | 58.90028791 | 67.35796316 |

Figure 48: Predicted Values Using Model 3

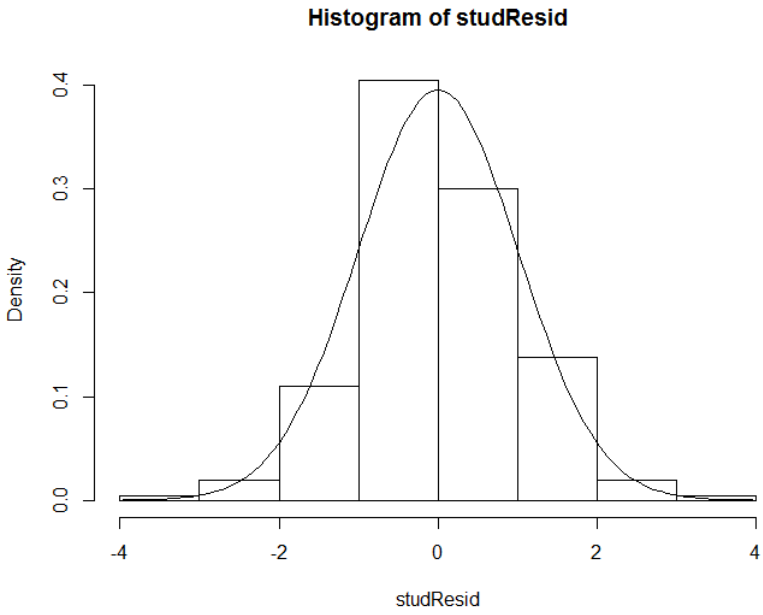


Figure 49: Histogram Of Residuals - Model 3

## 10. APPENDIX

So, how does a ridge regression work? Let's take a look at it in a more thorough manner. So normally, we would have estimators based off of the GAUSS-MARKOV theorem which would have no bias, and of all the unbiased estimators, they would have the least variance i.e best linear unbiased estimators i.e BLUE. And, the way we find this is by minimizing the sum of squared residuals like so:

$$Y = \mathbf{X} \cdot \beta + \epsilon \text{ and SSE} = \Sigma\ \epsilon^2$$

$$\text{Therefore: SSE} = \Sigma\ (Y - \mathbf{X} \cdot \beta)^2$$

And, when this is minimized, you get:

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\ \mathbf{X}^T Y$$

However, with regression, it is different because as mentioned earlier in the paper, you have correlated variables which causes the $(\mathbf{X}^T\mathbf{X})$ determinant to be nearly 0 and so, the inverse of it isn't of much use anymore (it loses accuracy). This results in some very large variances. We fix this by adding a new matrix to our linear model and minimize that instead. What this new matrix does is that it in a way, penalizes larger values of $\beta$ and thus we end up with variances that are not abnormally large and in fact, often times smaller than the GAUSS-MARKOV estimators. This is kind of a elementary explanation, but I'll end it here for now! Below is a proof of the new minimization to find the estimators.

$$\min_{\beta}(Y - \beta^T\mathbf{X})^T(Y - \beta^T\mathbf{X}) + \lambda\beta^T\beta \tag{1}$$

Note:

$$\frac{\partial(Y - \beta^T\mathbf{X}^T(Y - \beta^T\mathbf{X}))}{\partial\beta} = 2\mathbf{X}^T(Y - \beta^T\mathbf{X}) \tag{2}$$

$$\frac{\partial\lambda\beta^T\beta}{\partial\beta} = 2\lambda\beta. \tag{3}$$

Therefore:

$$\mathbf{X}^T Y = \mathbf{X}^T\mathbf{X}\beta + \lambda\beta. \tag{4}$$

And so:

$$\beta = (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T Y. \tag{5}$$

And we are done!

## 11. Closing Comments

Thank you for getting here.