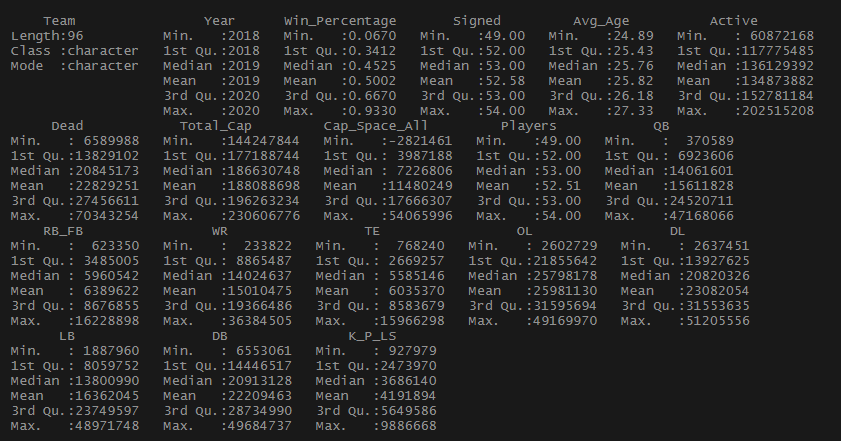
NFL Salary Analysis

The roster of an NFL football team my go through significant changes every year throughout the offseason. Teams may spend more money in different positions than other teams in the league. Throughout the league’s history, different ideologies have been made to reflect how teams won in the past. These ideologies vary from things like “defense win games”, “the league is a passing league, and the Quarterback position is the most important role”, “teams need to win while maximizing rookie contracts.” These are only a few ideas of all the many different takes on NFL football. This analysis will look at how the different ways of spending cap space impacts the number of games a team wins in a season.

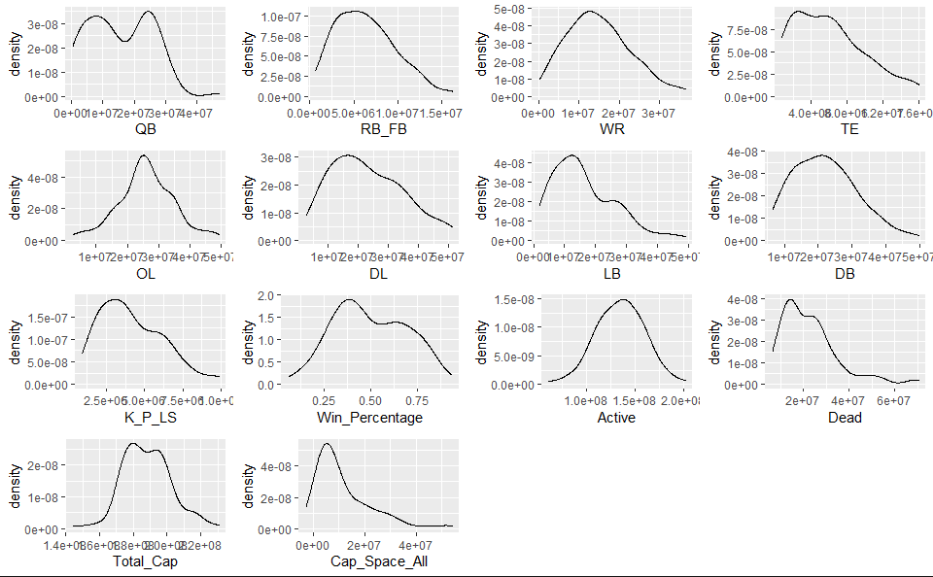
Looking at the summary statistics, the data shows the team’s overall money spent and how much money was spent for each position on the roster. The data shows the different linemen on each side of the field as a combined whole, being “OL” for offensive linemen and “DL” for defensive linemen. For the season 2018 – 2020, The average team spends about $188M for the whole team in a span of 1 season. Teams on average spend the least amount of money on the running back/full position, tight end position, and special teams’ players. On average, the bulk of a team’s spending goes to the offensive line, defensive line, and defensive backs. For most teams, money used on quarterbacks, wide receivers, and linebacker are in the middle.



**Summary statistics of combined data**

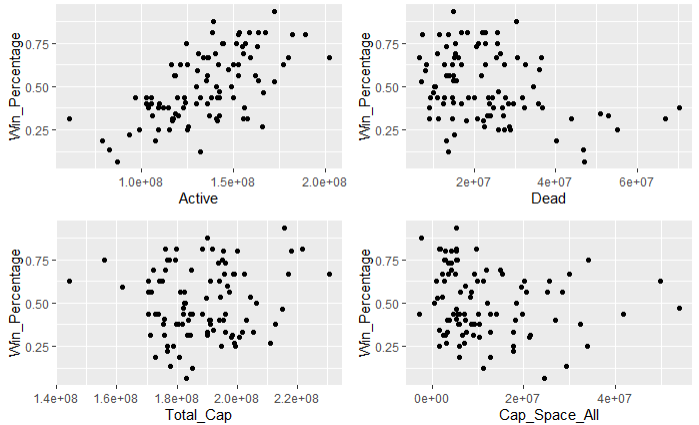
Creating a univariate density plot will show the distribution of how much money NFL teams usually spend on a position.

This plot shows that most of the data looks normally distributed or otherwise skewed to the left.



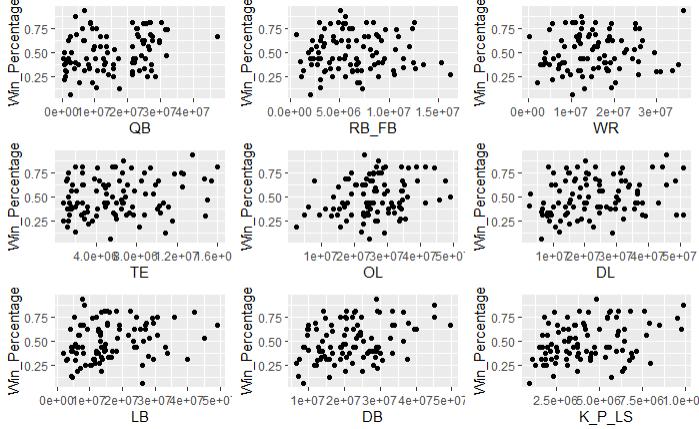
**Univariate Plot**

This bivariate plot represents the team’s overall spending. The plot, indicates that a team’s active money may have linear correlation to the win percentage in a season. As the amount of active money increases, the win percentage also increases. The other variables do not seem to have any significant visual correlation to win percentage.

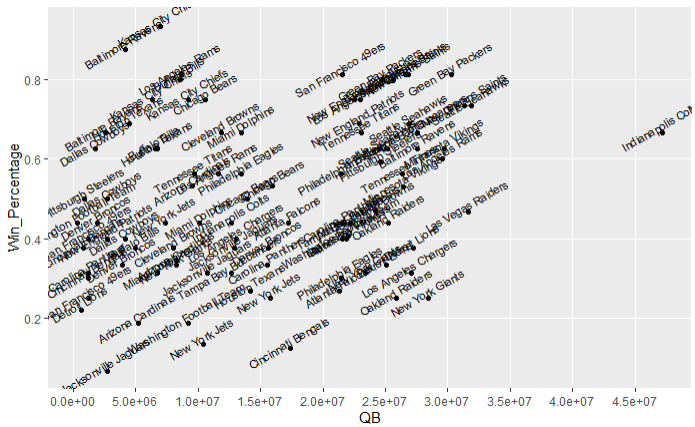


**Bivariate plot of team’s active and dead money**

The next bivariate plot shows the amount of money teams spend on the different positions. Possible linear correlations include the offensive line, defensive backs, and kickers/punters. A noticeable thing is the QB plot has an outlier on the far right of the plot. This spending is almost double of most teams. Doing further research, this point is in the 2020 season for the Indianapolis Colts team. They acquired Phillip Rivers over the offseason and in the prior season before gave Jacoby Brissett a contract extension for being a starter. Whiling keep both QBs on the roster, the amount spent become inflated compared to other teams.



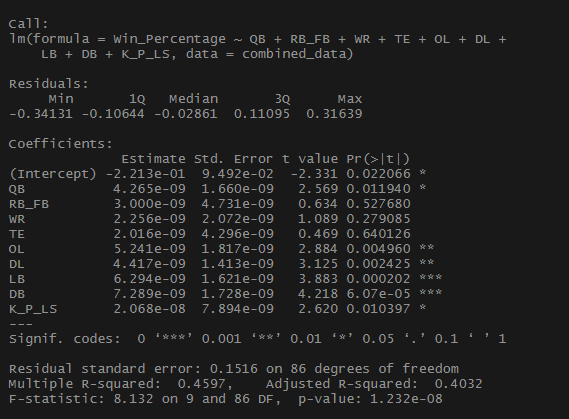
**Bivariate Positional Plot**



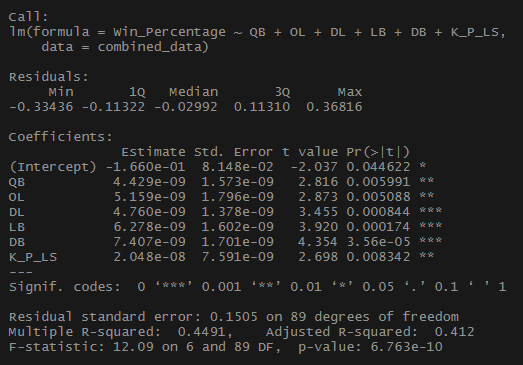
**Explaining the Outlier**

Linear Regression Model:

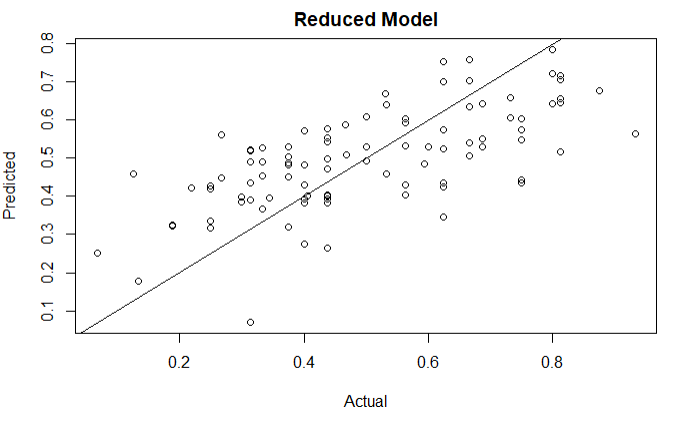
For this analysis, the goal is to create a linear regression model, calculating the win percentage based on the money a NFL team spends on each position. The initial model includes all the positions on the roster. After taking a look at the summary, a reduced model was created using only the statistically significant variables in the initial model. The multiple linear regression model has a R-squared value of 0.44 and a residual standard error of 0.15 on 89 degrees of freedom. The residuals show that the model is a mediocre fit for the data due to the variance it does not cover.



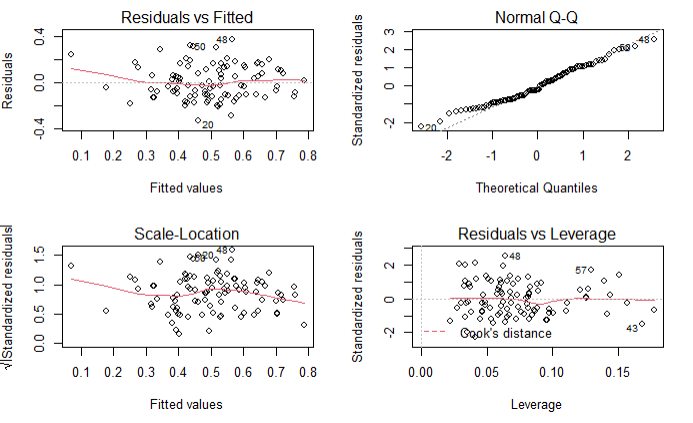
**Initial Model**



**Reduced Model**

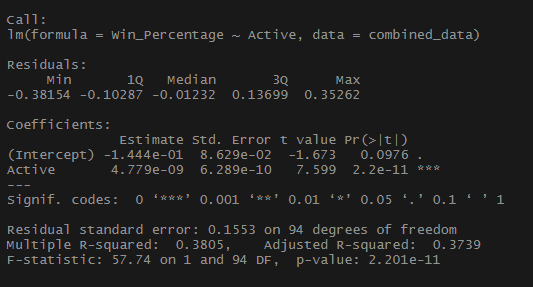


**Actual vs Predicted**

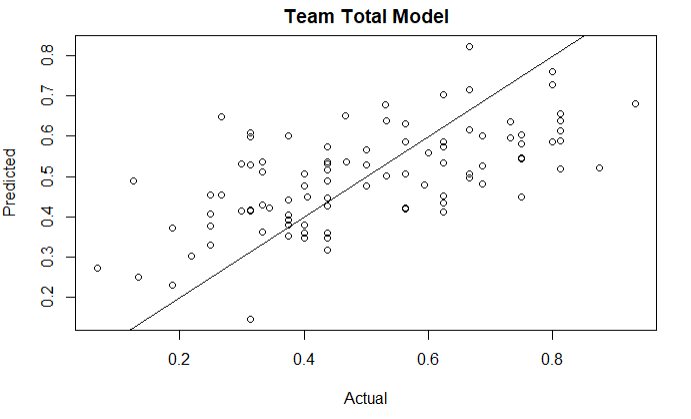


**Reduced Model Residuals**

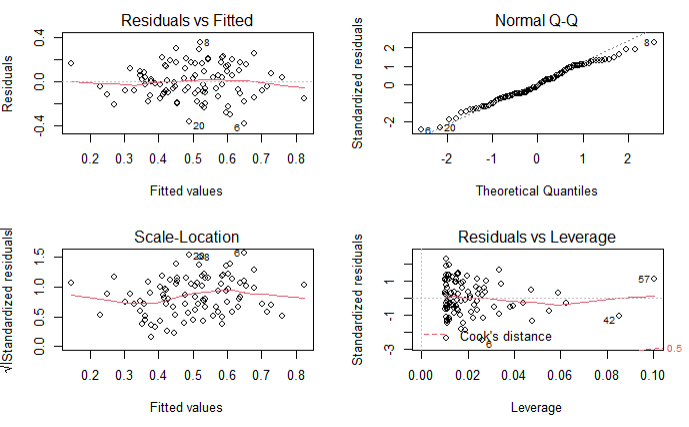
A second linear regression model is created to show the correlation of active money spent on win percentage. The variable Active is statistically significant with win percentage but due the complex data, the model does not cover much of the variance found. This is indicated in the low R-squared value of 0.3805.



**Overall Active Money Model**



**Actual vs Predicted Active Money Model**



**Active Money Model Residuals**

Limitations:

There are a few noticeable limitations to this analysis. The first being that the data does not go into a deep dive on how money is allocated in each unique position. For example, if a team spends $25M on the offensive line for a year, the data does not indicate if most of that money is going to one star player or spread out evenly amongst many players. Secondly the data does not reflect injuries that could have happened throughout the season. Injuries could have a season altering effect on a team’s win percentage. For example, if a team spends $30M on a star QB but he is injured in week 1; and the team has a losing record. This unplanned outcome does not entirely reflect a team’s expectations coming into the season and the data that is historically recorded. The third limitation is the analysis does not account how deep the game really is. In a football game, there are many other factors than the 22 players on the field. This includes coaching, management, referees, and practice. Even with these limitations, this analysis can still be of use.

Application:

The linear models that were created throughout the analysis can still be of use. Teams can use the model to create realistic goals for their teams coming into the season. For example, a team that spends less money may make it their goal to rebuild and find new talent throughout the season. On the other hand, a team that spends a lot of money may make it their goal to reach the playoffs or win a Superbowl. The model may also be useful in off season planning by prioritizing money on statistically significant positions. One course of action is to spend less money on wide receivers and focus spending money on the defensive line. The model is far from perfect, but it still has some useful aspects to it.

Conclusion:

Thank you for reading my analysis. This was a for fun project I wanted to do with a topic that interests me. I am pretty satisfied with the outcome of this project, as this was the first time, I scraped data off the internet and created my own analysis. Please comment any critiques or ideas you may have, and I will be sure to implement them in my next project.