Final Project - Samuel Choiniere

Offensive Tendicies in the NFL

Introduction

Football is a continously evolving sprot and offensive coordinater in the NFL are alwyas coming up with new ideas to gain an edge on opponents and give there team the best shot of winning the Super Bowl. This past season the Baltimore Ravens and Lamar Jackson produced a record setting rushing attack even though the league was appearing to go towards more pass heavy schemes. My goal with this data analysis to an analyze the trends the NFL is seeing with rushing versus passing offenses as well as seeing how a teams offensive production can be used as a predictor in that team making the playoffs. To do this analysis I will be using a data set consisting of data from each teams offensive production from 2010-2019 as well as data concerning the red zone, third down and fourth down conversion rates of the teams. I also added in a boolean column that represents weather or not that team reached the playoffs for that year. The data was collected from https://www.pro-football-reference.com/.

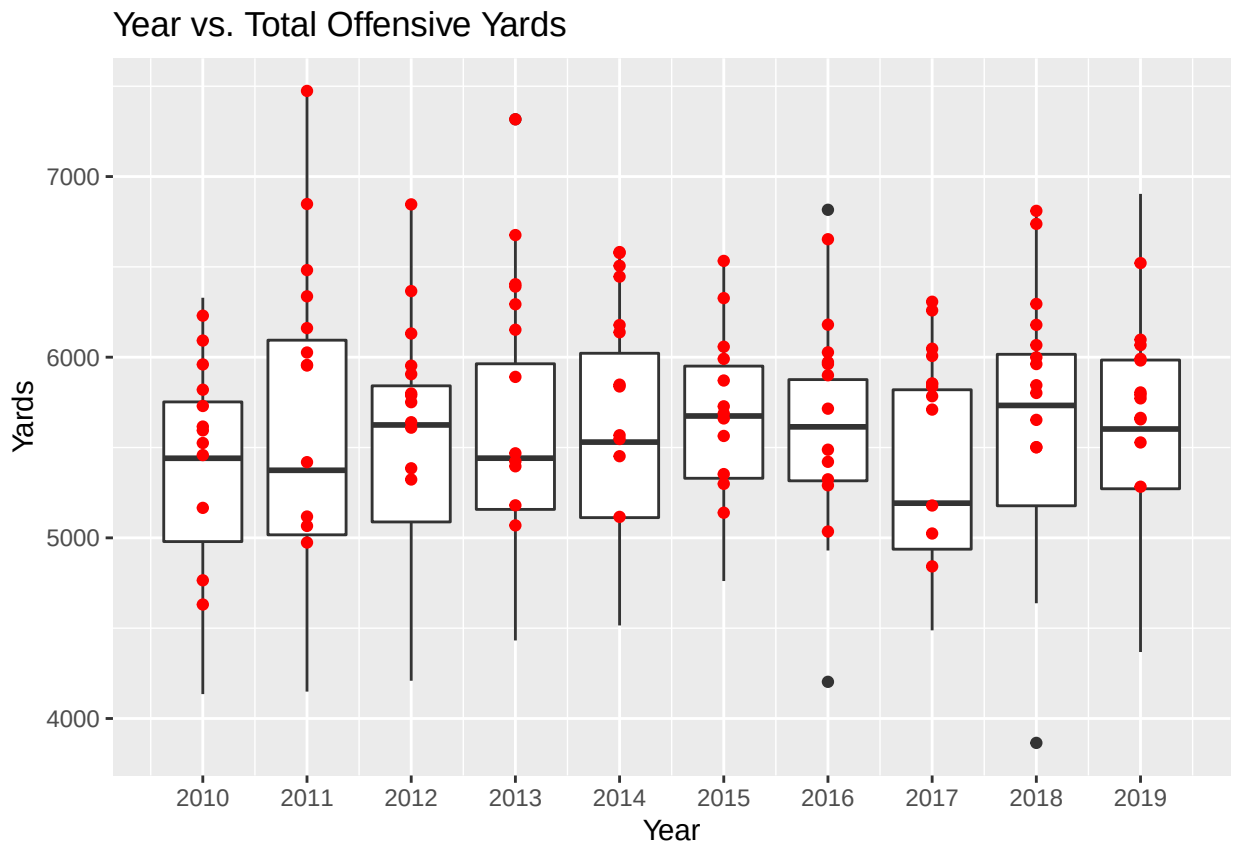
```
stats %>% head()
```

```
##   Rk                  Tm  G  PF  Yds  Ply Y.P TO FL X1stD Cmp PassAtt PassYds
## 1  1 New England Patriots 16 518 5820  986 5.9 10  5   335 331     507    3847
## 2  2   San Diego Chargers 16 441 6329 1039 6.1 29 16   357 359     544    4519
## 3  3  Philadelphia Eagles 16 439 6230 1038 6.0 25 12   322 348     561    3906
## 4  4   Indianapolis Colts 16 435 6092 1088 5.6 25  8   366 450     679    4609
## 5  5      Atlanta Falcons 16 414 5458 1097 5.0 17  8   353 361     577    3567
## 6  6      Oakland Raiders 16 410 5674 1039 5.5 26 10   306 279     491    3180
##   PassTD Int NY.A Pass1stD RushAtt RushYds RushTD RushY.A Rush1stD Pen Yds.1
## 1     37   5  7.2      196     454    1973     19     4.3      119  83   766
## 2     30  13  7.8      236     457    1810     18     4.0       94  84   677
## 3     28  13  6.4      189     428    2324     18     5.4      116 129  1101
## 4     33  17  6.6      253     393    1483     13     3.8       87  79   709
## 5     28   9  5.9      200     497    1891     14     3.8      111  58   598
## 6     18  16  5.9      159     504    2494     19     4.9      113 148  1276
##   X1stPy  Sc.  TO.     EXP Playoffs year X3DAtt X3DConv  X3D. X4DAtt X4DConv
## 1     20 47.0  5.4   73.64     TRUE 2010    197      95 48.2%     14       7
## 2     27 40.3 14.5   27.47    FALSE 2010    205      92 44.9%     13       5
## 3     17 39.8 11.7  -11.00     TRUE 2010    214      85 39.7%     13       9
## 4     26 40.2 13.4   28.65     TRUE 2010    204      91 44.6%      8       5
## 5     42 39.3  9.0  -39.55     TRUE 2010    240     112 46.7%     15      11
## 6     34 34.5 11.8 -110.80    FALSE 2010    220      75 34.1%     20       8
##    X4D. RZAtt RZTD RZPct
## 1 50.0%    67   42 62.7%
## 2 38.5%    59   33 55.9%
## 3 69.2%    58   31 53.4%
## 4 62.5%    56   38 67.9%
## 5 73.3%    58   35 60.3%
## 6 40.0%    48   25 52.1%
```

Exploratory Data Analysis

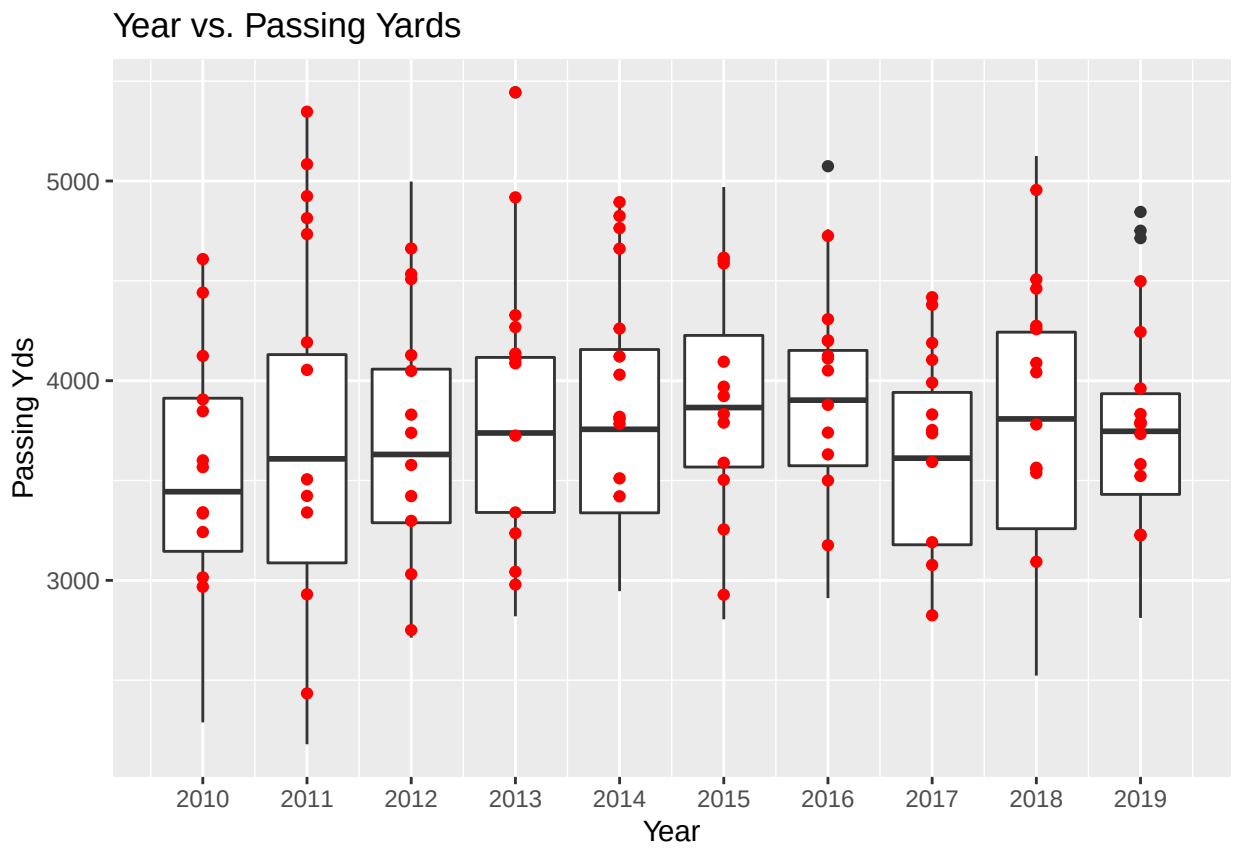
First lets start by plotting the total offensive yards of the teams each year and highlight where the playoff teams lie in relation to the rest of the league. It appears that total offense has a fairly normal distribution each year with a mean of around 5500 yards each season. Also it is worth to note that it appears that a majority of the playoff teams have above the average total yards for each season.

```
stats %>% ggplot(
  mapping = aes(group = year, x = year, y = Yds)) +
  geom_boxplot() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes( group = year, x = year, y = Yds), co
  labs(title="Year vs. Total Offensive Yards",
         x = "Year",
         y = "Yards") +
  scale_x_continuous(breaks = years)
```

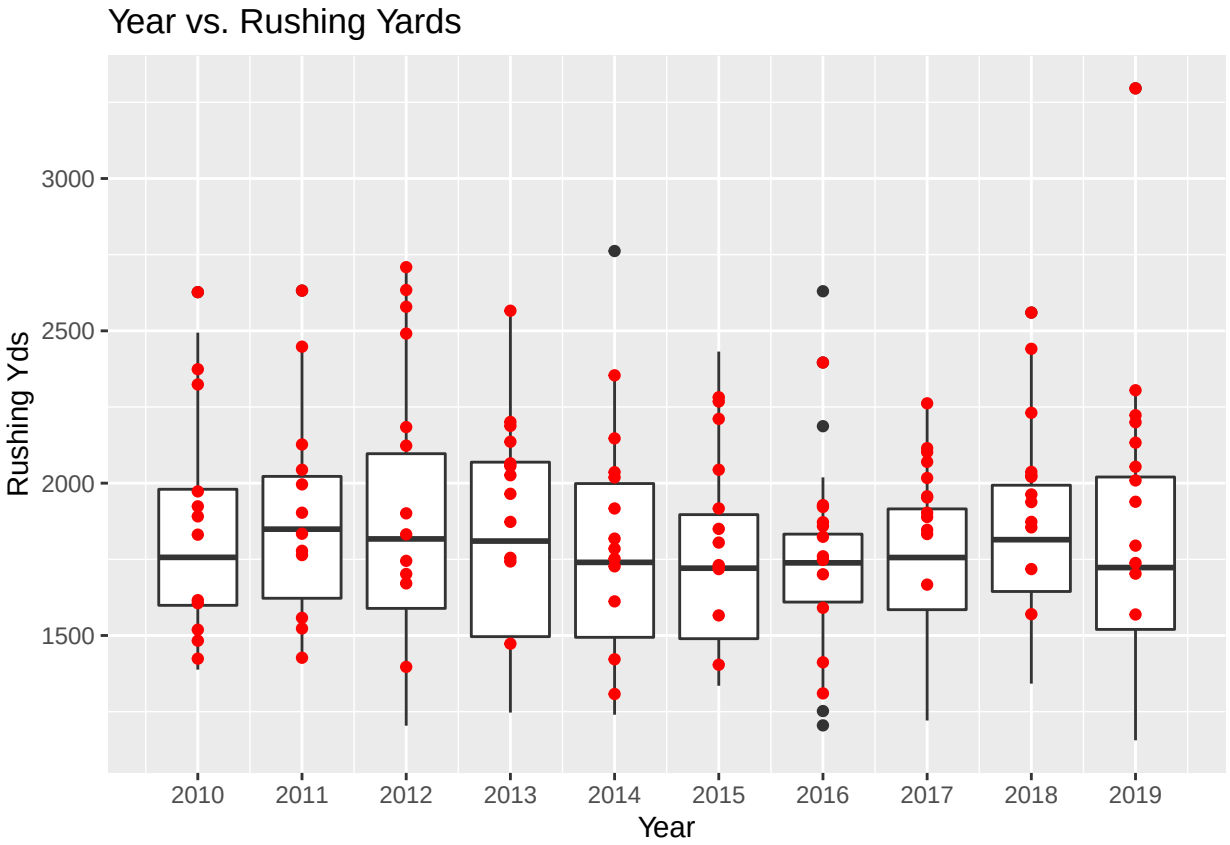


Next lets split the offense up into passing and tushing yards and compare the trends.

```
stats %>% ggplot(
  mapping = aes(group = year, x = year, y = PassYds)) +
  geom_boxplot() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes( group = year, x = year, y = PassYds
  labs(title="Year vs. Passing Yards",
         x = "Year",
         y = "Passing Yds") +
  scale_x_continuous(breaks = years)
```



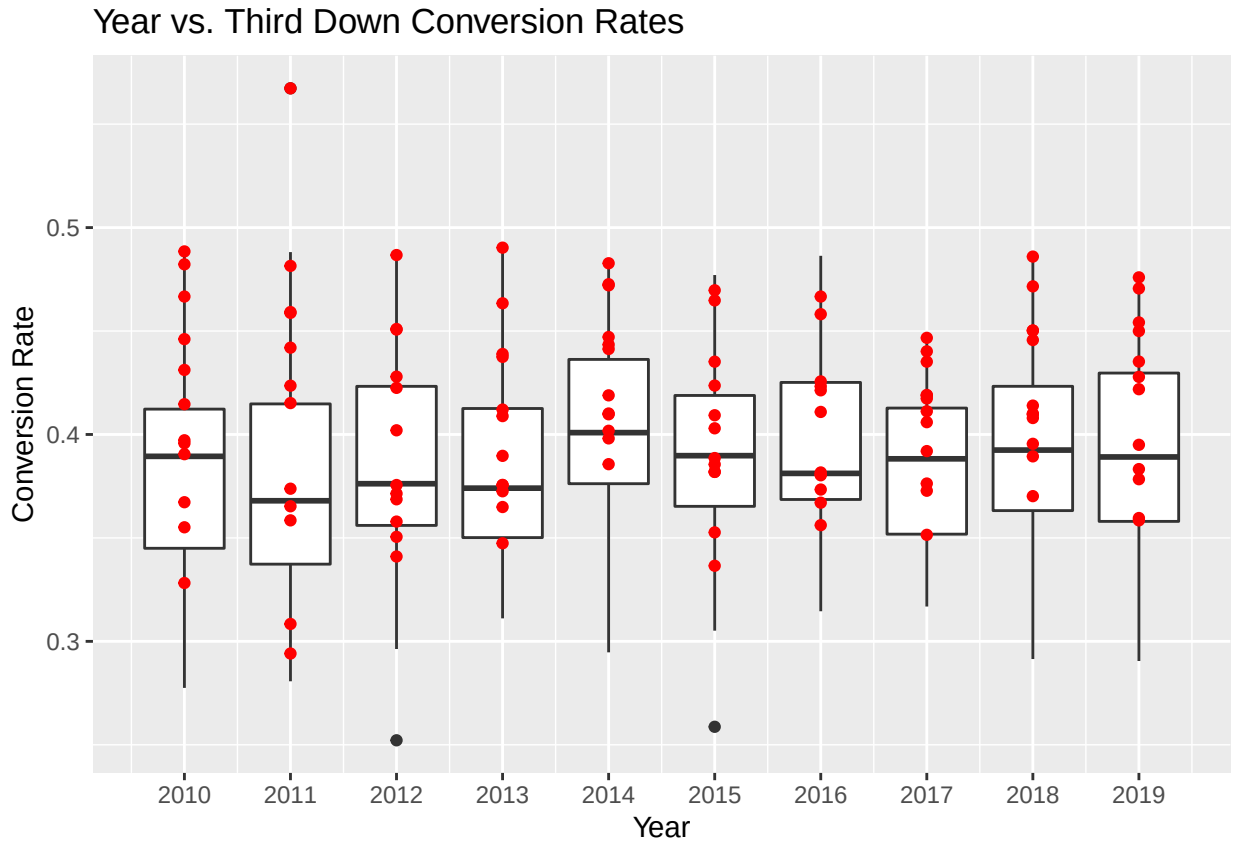

```
stats %>% ggplot(
  mapping = aes(group = year, x = year, y = RushYds)) +
  geom_boxplot() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes( group = year, x = year, y = RushYds
  labs(title="Year vs. Rushing Yards",
         x = "Year",
         y = "Rushing Yds") +
  scale_x_continuous(breaks = years)
```



The plots show that there was a steady increase in passing yards from 2010 to 2016 but saw a small decrease following the 2016 season. It also appears that a good amount of playoff teams actually have below average passing yards which I did not expect. In terms of rushing yards the league seems to be fairly steady in terms of average and spread except for a few outliers including the 2019 Baltimore Ravens that excceded 3000 rushing yards. I also appears that there are less playoff teams with below average rushing yards than playoff teams with below average passing yards. This could imply that a good rushing attack can increase a teams chance of making the playoff.

Another aspect of an offense is how efficient an offense is and there ability to convert third downs in order to stay on the field and increase there scoring potential. So next we will look at the third down conversion rates of playoff teams in comparision to to the rest of the league.

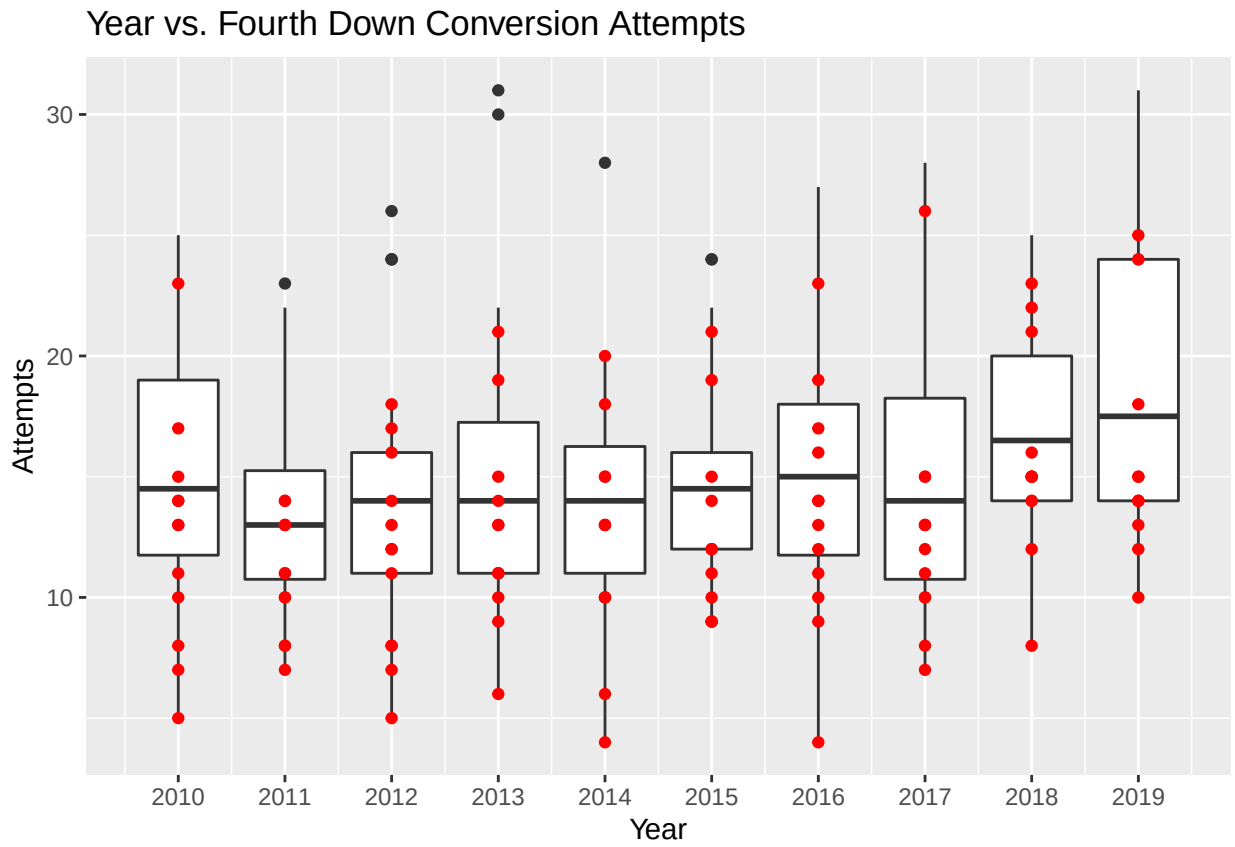
```
stats$ConvRate <- stats$`X3DConv`/stats$`X3DAtt`
stats %>% ggplot(
  mapping = aes(group = year, x = year, y = ConvRate)) +
  geom_boxplot() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes( group = year, x = year, y = ConvRat
  labs(title="Year vs. Third Down Conversion Rates",
         x = "Year",
         y = "Conversion Rate") +
  scale_x_continuous(breaks = years)
```



Again it appears that playoff teams have above average conversion rates on thrid down in comparison to the rest of the league.

Finally lets look at the trends for fourth down attempts as teams that are more aggressive on offense may have an increased chance of making the playoffs.

```
stats %>% ggplot(
  mapping = aes(group = year, x = year, y = `X4DAtt`)) +
  geom_boxplot() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes( group = year, x = year, y = `X4DAtt
  labs(title="Year vs. Fourth Down Conversion Attempts",
         x = "Year",
         y = "Attempts") +
  scale_x_continuous(breaks = years)
```

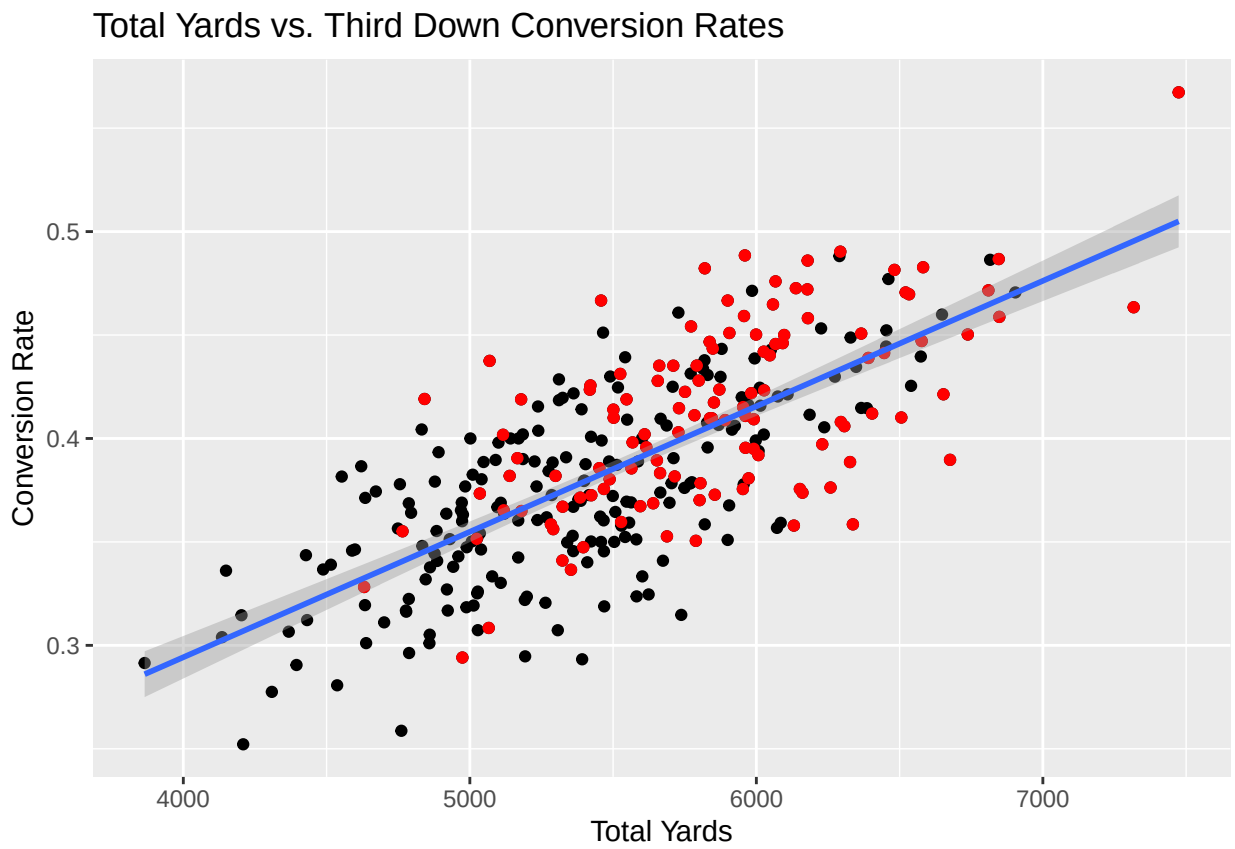


There does not seem to be much correlation between fourth down attempts and playoff teams here, however it is worth noting the steady increase on the overall number of attempts across the league in the past few years. This is mostly due to a rise in analytics use by football teams and realizing it pays off to be more aggressive on fourth down that previously thought. More information about that can be found here as it is out of the scope of this report, https://www.advancedfootballanalytics.com/index.php/home/research/game-strategy/120-4th-down-study.

Linear Regression

Based on our exploratory data analysis it seems that there is a higher probability of a team with above average total yards and third down conversion rate making the playoffs. So lets plot a teams total yards vs. thier thrid down conversion rate and highlight the playoff teams to see their relationship.

```
stats %>% ggplot(
  mapping = aes(x = Yds, y = ConvRate)) +
  geom_point() +
  geom_point(data = subset(stats, Playoffs == TRUE), mapping = aes(x = Yds, y = ConvRate), color = "red
  geom_smooth(method = lm) +
  labs(title="Total Yards vs. Third Down Conversion Rates",
         x = "Total Yards",
         y = "Conversion Rate")
```



Thier is a positive relation ship between a teams total yards and third down conversion rates. This makes sense if a team has a higher rate of conversion they are on the field more often and then having more opportunity to generate more yardage. It also it mushc clearer now that the playoff teams are above the non-playoff teams in terms of both total yards and conversion rate. We can now fit a linear regression to predict the relationship between total yards and conversion rate.

```
linear_fit <- lm(ConvRate~Yds, data = stats)
linear_fit %>% tidy()
```

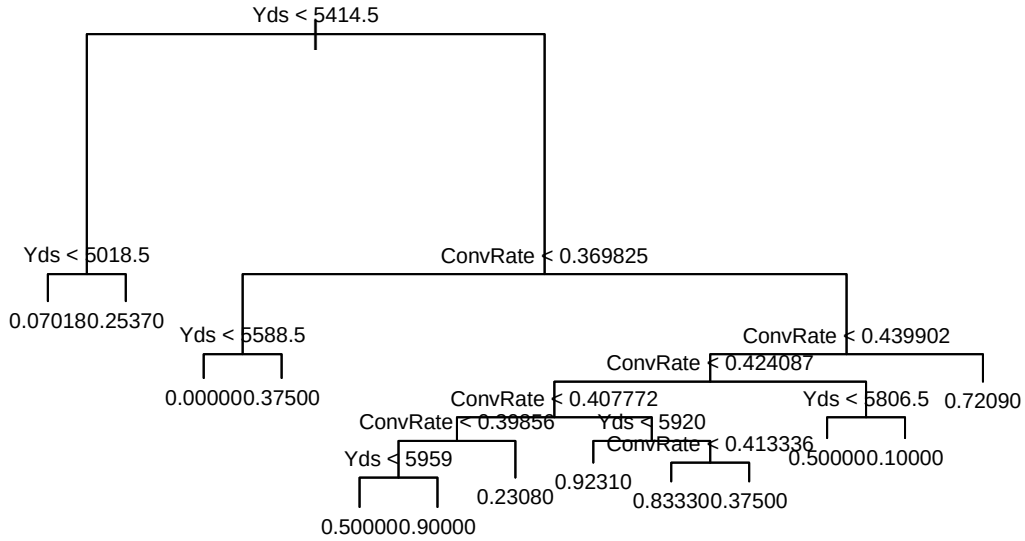
```
## # A tibble: 2 x 5
##   term         estimate  std.error statistic  p.value
##   <chr>           <dbl>      <dbl>     <dbl>    <dbl>
## 1 (Intercept) 0.0517    0.0176          2.93 3.63e- 3
## 2 Yds         0.0000606 0.00000316     19.2  5.16e-55
```

The model projects that for every extra yard a team produces there conversion rate goes up by .00006, or for every 1000 yards a teams conversion rate increases by .06. Since the p-value is low we can accept this assumption. We will then use this model to predict whether or not a team makes the playoffs.

Prediction Method

Due to the seeminly higher probability of making the playoffs with higher total yard and conversion rate, lets make a decision tree to determine the odds a team has in making the playoffs based on third down conversion rate and total yards. We will use the years 2010-2018 to create the model and then test it on the teams from 2019.

```
tree <- tree(Playoffs~ConvRate+Yds, data = subset(stats, year < 2019))
plot(tree)
text(tree,cex = .65, pretty = 0)
```



Based on the tree a team will have the highest probability of making the playoffs by having a offense with more than 5920 total yards and a conversion rate between .4077 and .4241, and team with less than 5424.5 total yards have the smallest chance of making the playoffs. Now using this tree we can calculate the 12 teams with the highest probability of making the playoffs from 2019 and compare them to the teams that actualy did make the playoffs.

```
prediction <- predict(tree, subset(stats, year == 2019))
ordered_prediction <- order(prediction, decreasing = T)[1:12]
mapped_prediction <- vector( , 32)
for(value in ordered_prediction) {
  mapped_prediction[value] <- TRUE
}
playoff_accuracy <- mapped_prediction & subset(stats, year == 2019)$Playoffs
playoff_accuracy <- table(playoff_accuracy)
playoff_accuracy[names(playoff_accuracy) == TRUE]
```

```
## TRUE
##    9
```

The tree surpisingly predicted 9 of the 12 playoff teams from 2019. Thus accurately determining if a team make the playoffs or not 26/32 times. The sample and testing data is fairly small so there is no guarentee the model will always be as accurate but it is interesting to note how accurate it was for 2019.

Conclusion

Based on our findings it seems that third down conversion rates and total yardage a strong predictor in a team chance of making the playoffs. Team that have higher numbers in these categories tend to have a higher change of making the playoff but there is a sweet spot in the middle that results in a higher percentage and just having the highest in both categories is not always the right place to be. However, offense is only half of the game and thus can only determine half of a teams success. Defense and special teams are critical parts of the game that seperate the top teams from the rest. It is interesting to see the impact that a good offense has when excluding the other aspects of the game however. Also I did not see how the teams that make the playoffs actually faired in them and how offenseive production could lead to the Super Bowl.