

# Report

*Kaiyu Yan*

*December 4, 2018*

## Abstract

## Introduction

## Background

Trying to determine who was the best quarterback or running back through out the year is usually an interesting topic for media and college football fans. Also, to win a game, coach always need to put his best players on the field. Therefore, identifying and projecting the performance for players is particularly important for a team. People like to take one of the most basic statistic to measure performance. For example, yards per attempt(YPA), we like to use this measurement to determine the performance of a quarterback or running back's performance. However, for most time, when we look at the leaders in yards per attempt, we will notice that the statistical data is not useful. Because the highest yards per attempt always dependent on the lowest number of attempts as shown in table 1.1.

Attempt	Yards	YPA	Position	Fullname
1	76	76.000000	WR	Robert Woods
1	45	45.000000	WR	Jeric Magnant
1	42	42.000000	K	Anthony Melchiori
1	36	36.000000	S	Bubba Poueu-Luna
1	33	33.000000	PK	Jamie Boyle
1	33	33.000000	WR	Thomas Johnson
1	31	31.000000	DE	Nate Terhune
2	54	27.000000	RB	James Potts

Table 1.1 : Leaders in Yards Per Attempt for the 2012 season

Attempt	Yards	YPA	Position	Fullname
2	54	27.000000	RB	James Potts
27	279	10.333333	RB	Tim Gay
62	621	10.016129	RB	Melvin Gordon
2	20	10.000000	RB	R.J. Robinson
1	10	10.000000	RB	Isiah Willis

Table 1.2 : Leaders in Yards Per Attempt for the 2012 season

## Objective

The main objective of this project is trying to build a model that will generalize the most unbiased information to help us to determine the best performed player based on certain measurement. The potential implication of this project could provide suggestion for team on how to pick best performance as starter.

## Data and Method

### Data Source and Description

The data are from two different sites. Some of them are from Kaggle.com while some data such as power-index was scraped from espn.com. All the data are real and published on the website. I've also used several other site such as sports-reference.com to compare the data realness.

### Description of data

The data used in this project is real statistic data of NCAA College Football for the 2012 season. The data include the following information: 1) Attempt: The total number of a player attempt to carry a ball to run in the season.

2) Yards: The total yards of a player gained in the season. 3) YPA: Yards per attempt of a player gained in the season.

4) Position: Player's position of the field. 5) Fullname: Player's name.

6) TEAM: Team in NCAA FBS Division.

7) FPI: Football power index of each team.

8) OFFENSE: Team offense efficiency index.

9) DEFENSE: Team defense efficiency index.

10) OVERALL: Team overall efficiency index. 11) Rush.Att: Team total rush attempted in season 2012. 12) Rush.Yard: Team total rush yards in season 2012. 13) YPC: Team yard gained per carry in season 2012.

14) Opp.Att.Allowed: Total number of rushing attempt allowed for opponent team. 15) Opp.Yds.Allowed: Total rushing yards allowed for opponent team. 16) Opp.Ypc.Allowed: Rushing yards per carry allowed for opponent team. 17) Opp.Ypg.Allowed: Rushing yards per game allowed for opponent team. 18) Opp.FPI: Opponent power index.

19) Opp.Def.Eff: Opponent team defense efficiency. 20) Opp.Overall.Eff: Opponent team overall efficiency.

```
## -- Attaching packages -----
## v ggplot2 3.0.0      v purrr  0.2.5
## v tibble  1.4.2      v dplyr  0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

## EDA and Result

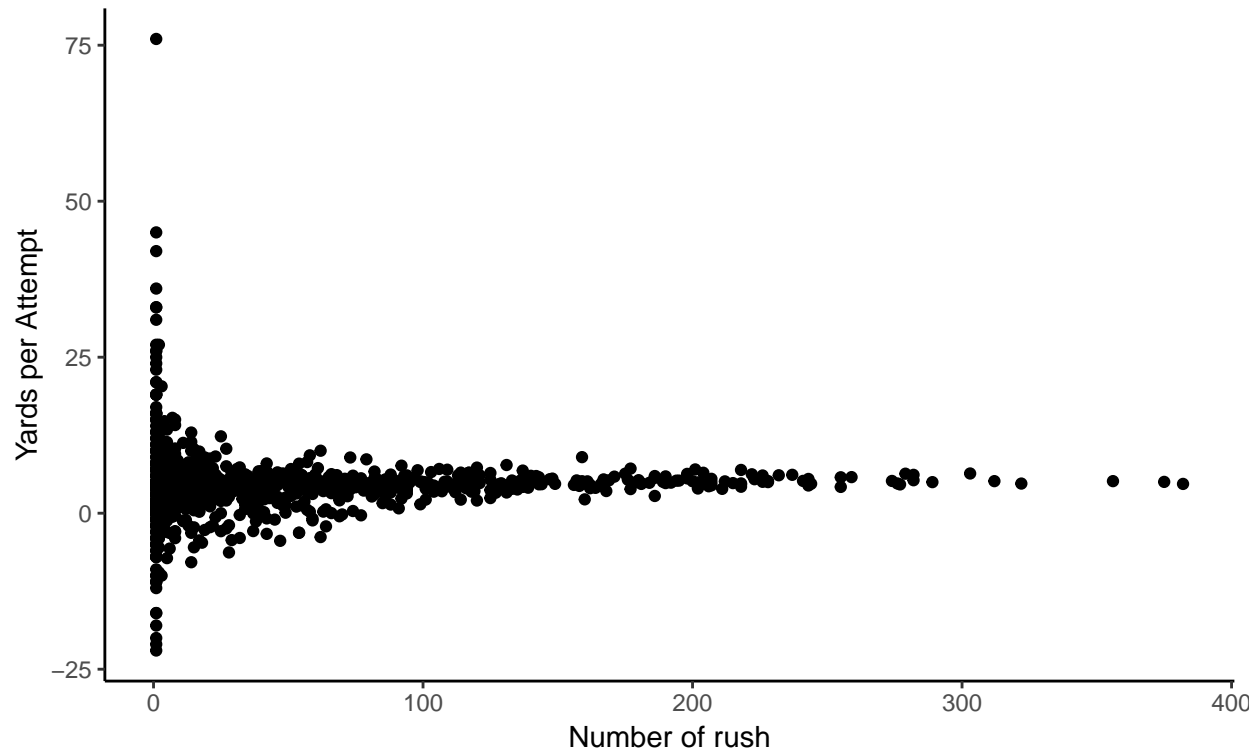
### EDA

As a NCAA college football fan, I have some knowledge of football stats and I also have some expectation of my data exploration. I'm expecting a lower number of players runs the ball more time. I'm also expecting a player will have higher YPA when his team are better and the opponent FPI or opponent defense efficiency are low.

## 1. Yards per Carry vs. Number of Attempt

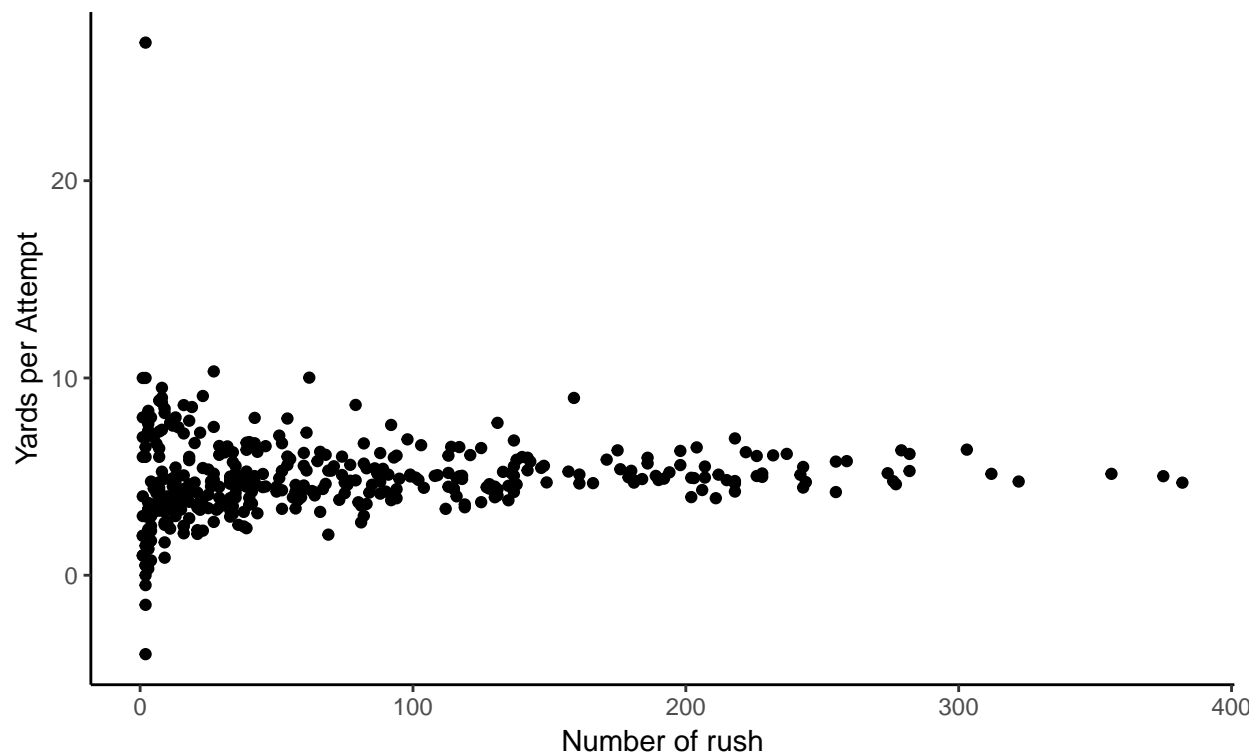
All Player's Yards per carry VS Number of Attempt

Effect of number of runs on YPA



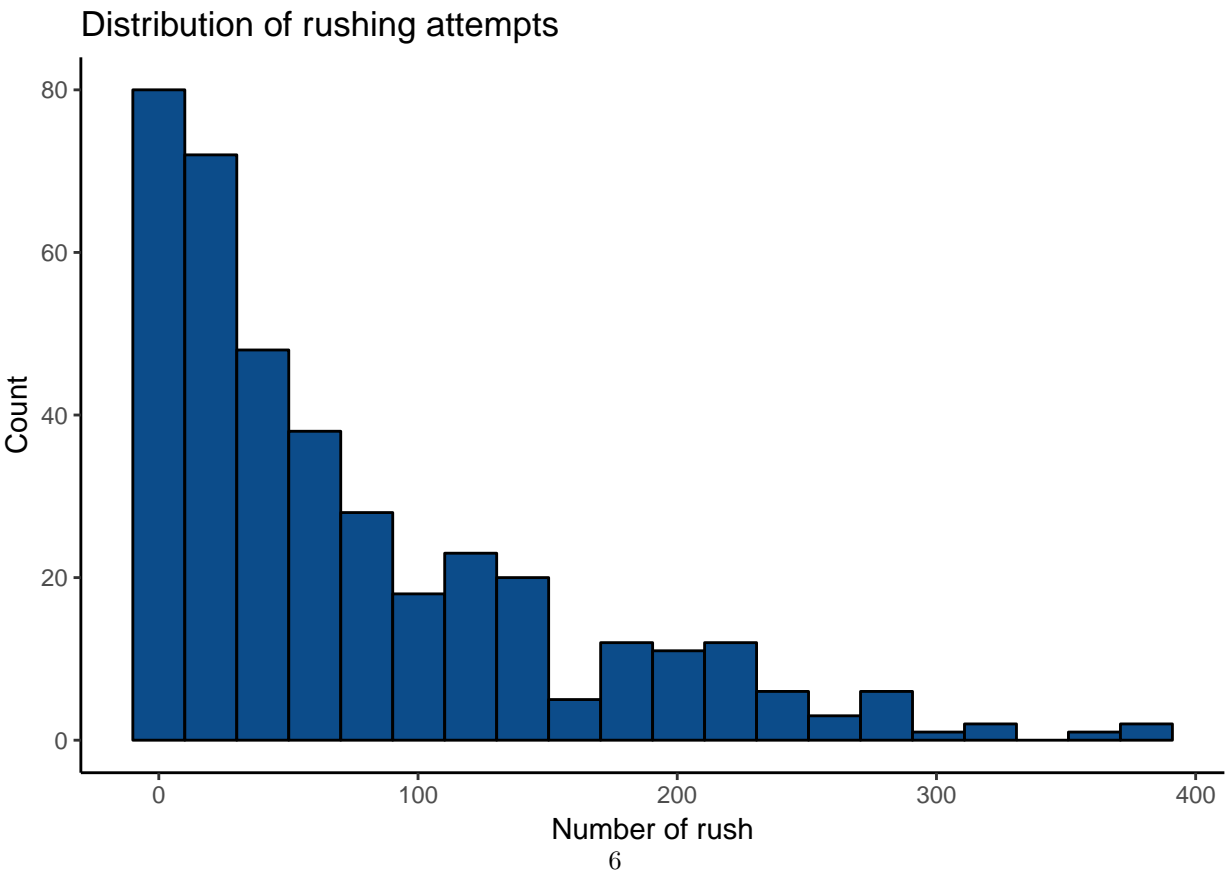
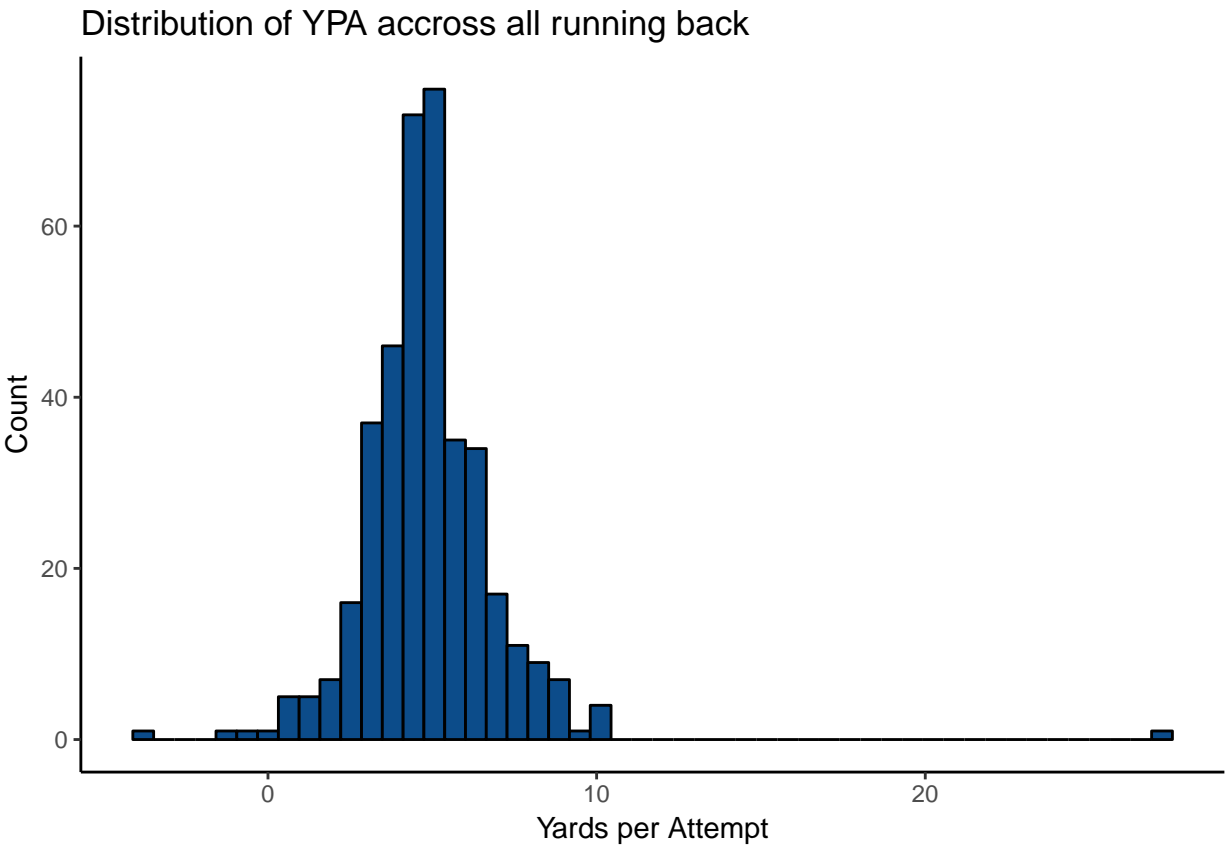
All Running back's Yards per carry VS Number of Attempt

Effect of number of runs on YPA



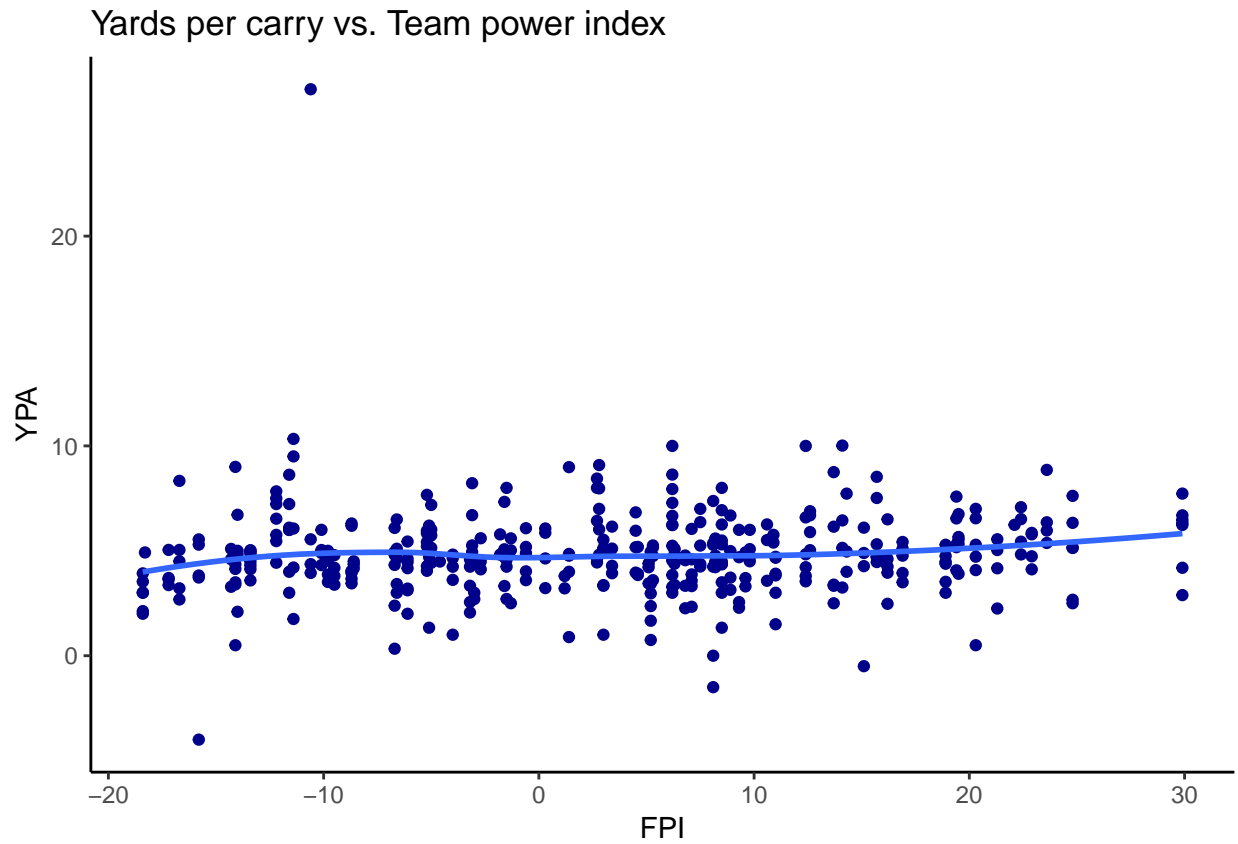
For those two plots, I displayed the relationship and effect between yards per attempt and the number of attempt. it doesn't account for the quality of opposition faced or the strength of own team. You can find that for all player position and running back player, they have similar trend for yards per attempt and number of rush.

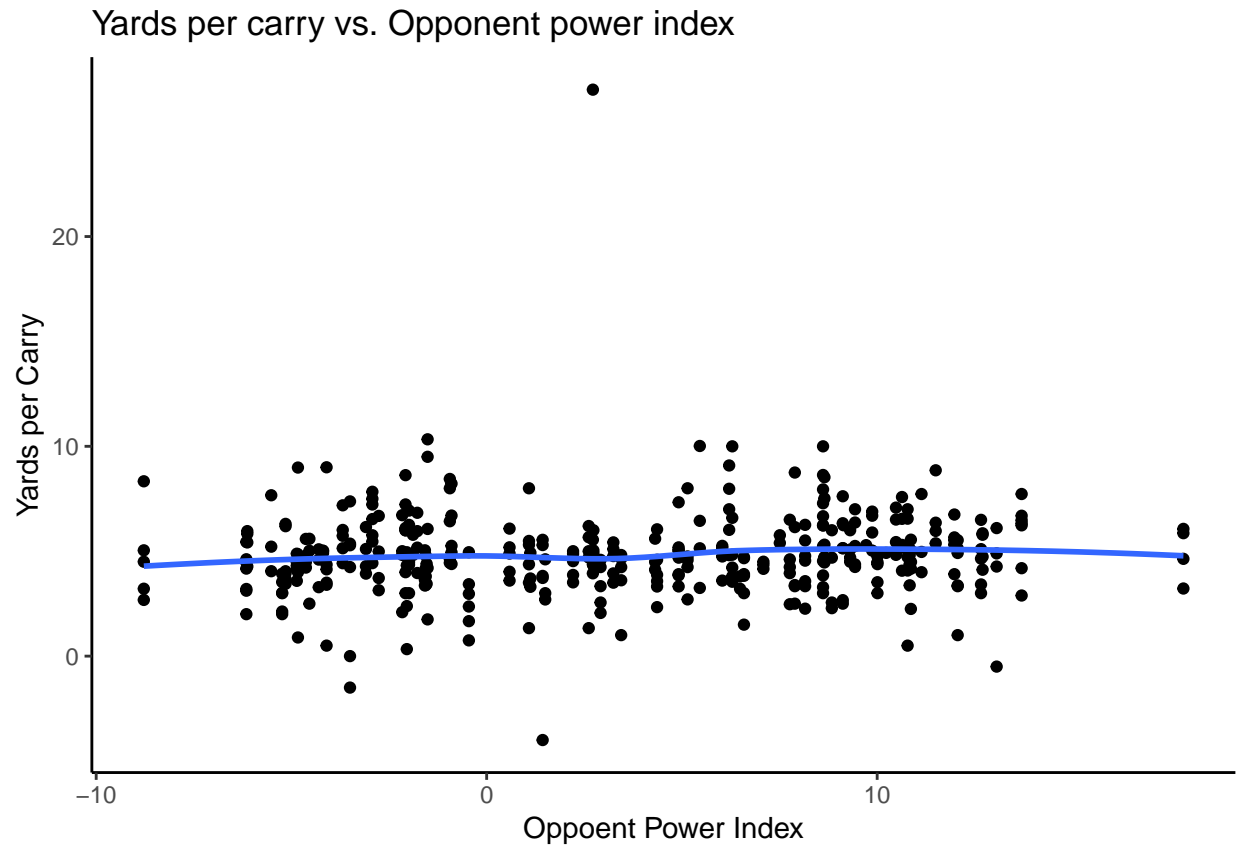
2. Distribution of Rushing Attempt and Yards per Carry



Those two graphs above shows the distribution of YPA and rushing attempts. From both plots I can say that it consist with my expectation. The distribution of YPA is knid of a normal distribution with some outliers and the count of rushing attempts will decreasing with more attempts.

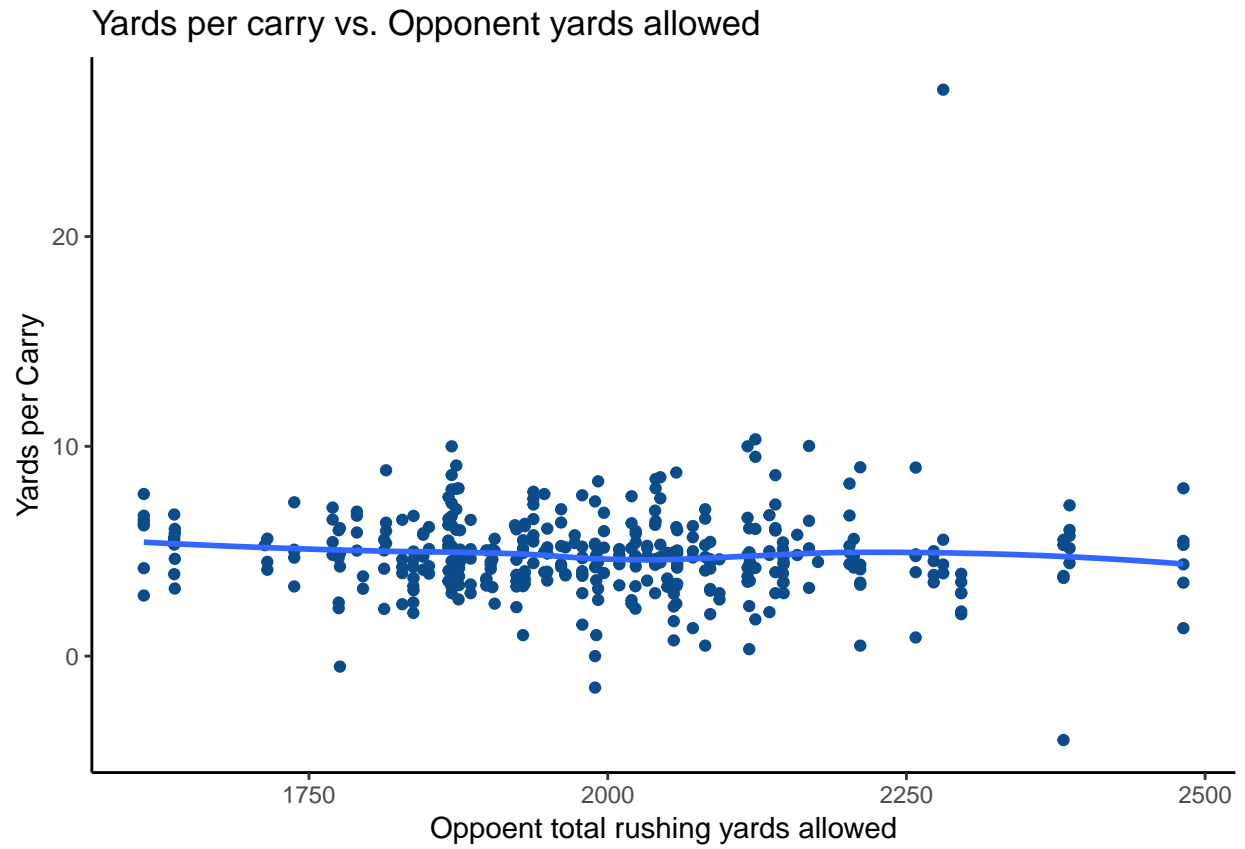
### 3. Relationship between YPA and other variables.

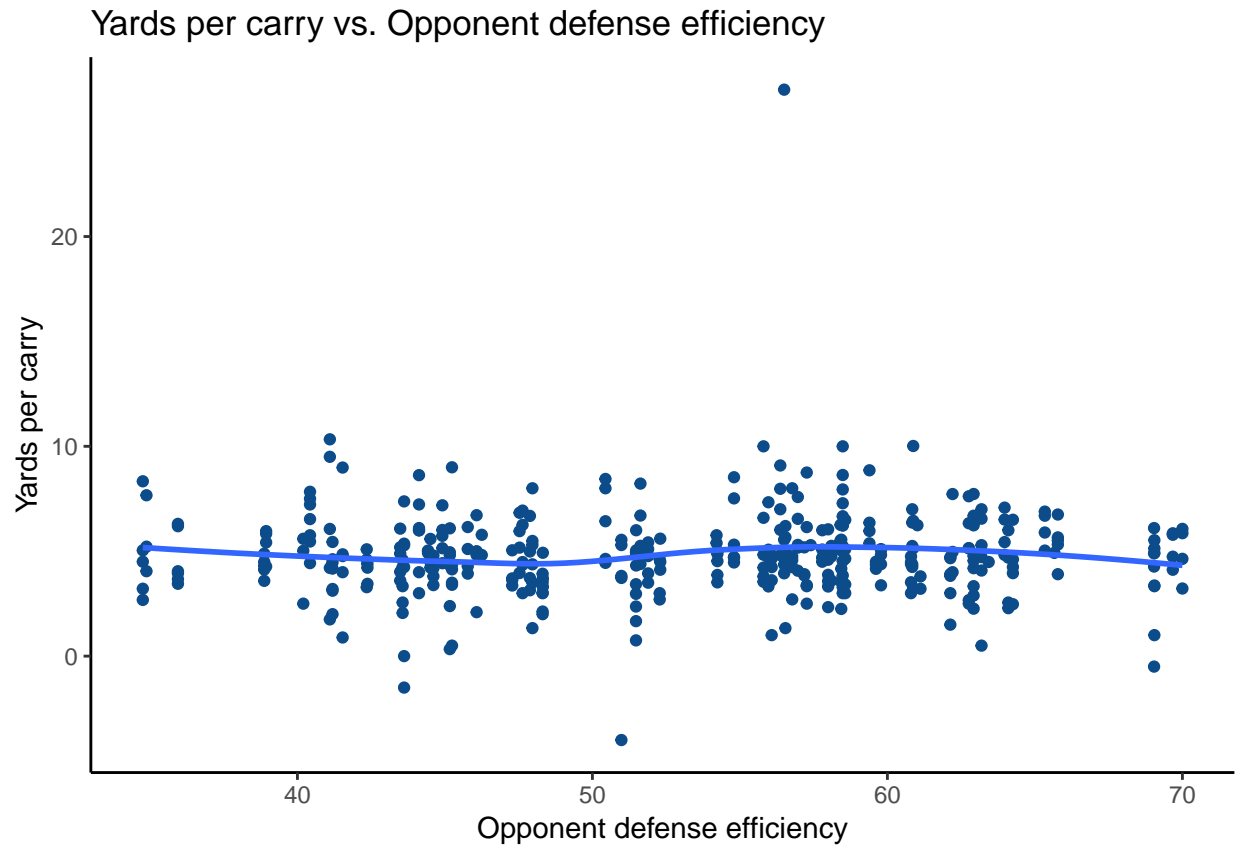




I've made those two plot to explore the relationship between YPA of a player with the strength of his own team and the opponent team. Although it is not very clear, but figure 2.5 shows that a player will have higher YPA when his team has a higher power index.







For Figure 2.7, we can find that a player will have lower YPA when his opponent has a higher total rushing yards allowed. Figure 2.8 shows that a higher opponent defense efficiency will leads to a lower YPA.

## Model Used

Models selected to explain the YPA variable are: – Linear Model – Multilevel Mixed effect Model

## Model Choice

## Interpretation

## Model Checking

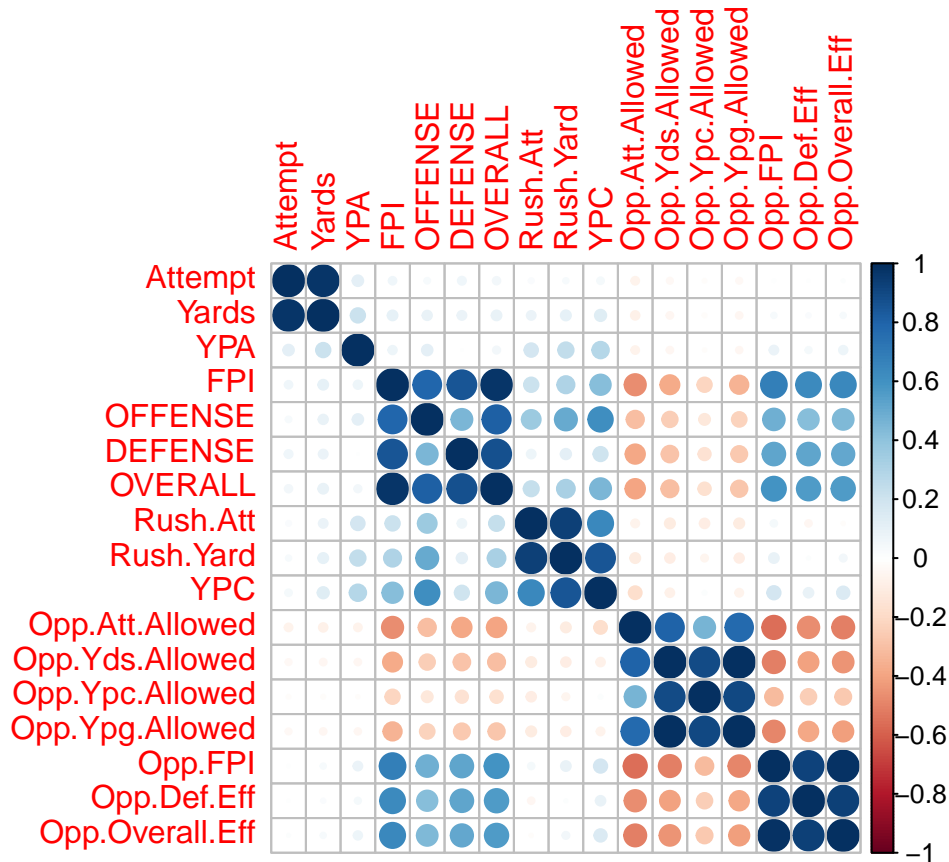
## Discussion

## Limitation

## Future Direction

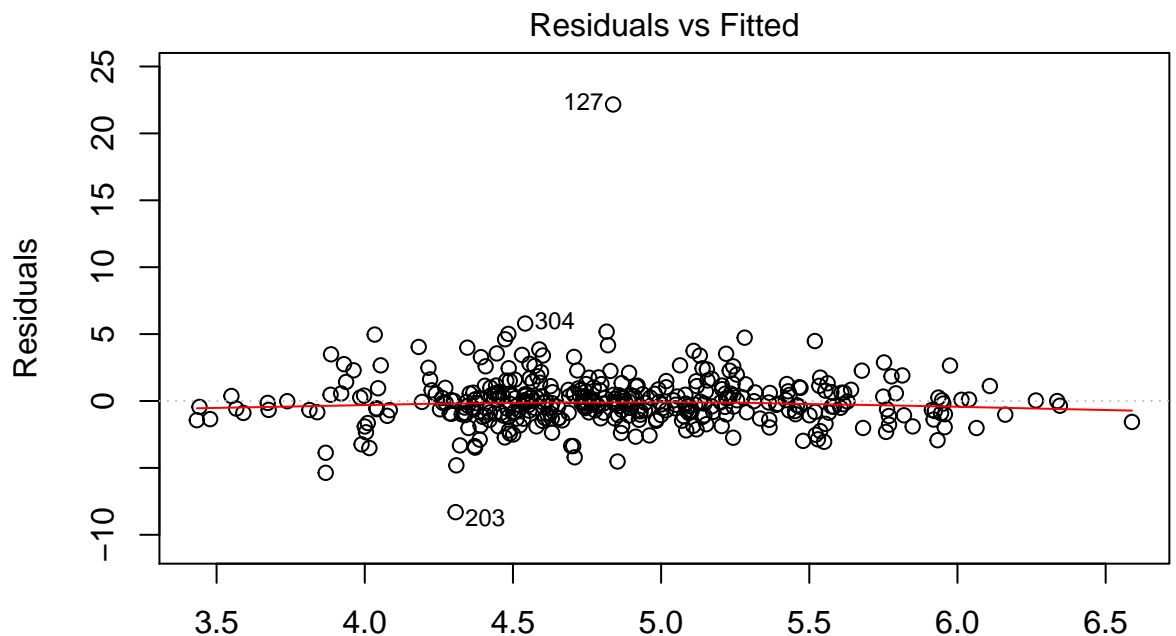
```
## corplot 0.84 loaded
```

```
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
## The following object is masked from 'package:purrr':
##
##     some
```

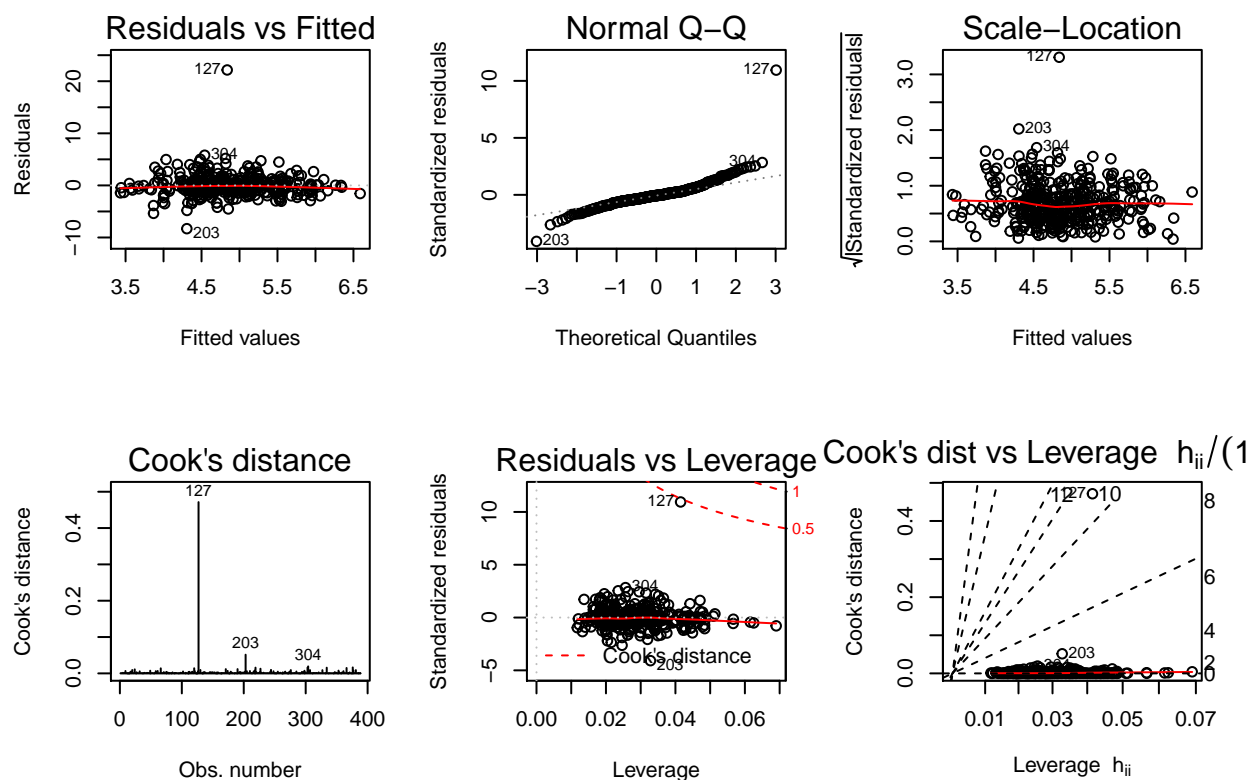


```
##
## Call:
## lm(formula = YPA ~ Attempt + FPI + OFFENSE + OVERALL + Rush.Att +
##     Opp.Ypc.Allowed + Opp.Ypg.Allowed + Opp.FPI + Opp.Def.Eff +
##     Opp.Overall.Eff, data = Player_rush1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3067 -0.9217 -0.1336  0.6572 22.1619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.025824   4.175310   0.246  0.80606
## Attempt       0.002991   0.001350   2.216  0.02729 *
## FPI           0.063055   0.048608   1.297  0.19535
```

```
## OFFENSE      0.013751  0.010097  1.362  0.17406
## OVERALL     -0.047160  0.026411 -1.786  0.07496 .
## Rush.Att     0.003371  0.001090  3.094  0.00212 **
## Opp.Ypc.Allowed 0.785875  1.101208  0.714  0.47589
## Opp.Ypg.Allowed -0.009988  0.019856 -0.503  0.61526
## Opp.FPI      -0.047041  0.130403 -0.361  0.71850
## Opp.Def.Eff   -0.028087  0.033954 -0.827  0.40863
## Opp.Overall.Eff 0.064004  0.079261  0.808  0.41989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.068 on 377 degrees of freedom
## Multiple R-squared:  0.06925,    Adjusted R-squared:  0.04457
## F-statistic: 2.805 on 10 and 377 DF,  p-value: 0.002299
```

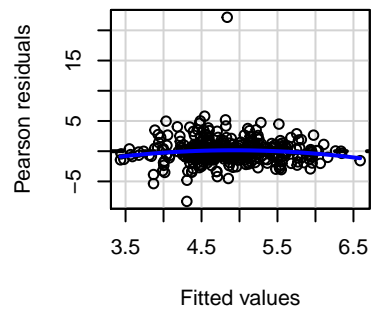


Fitted values  
lm(YPA ~ Attempt + FPI + OFFENSE + OVERALL + Rush.Att + Opp.Ypc.Allowed + O

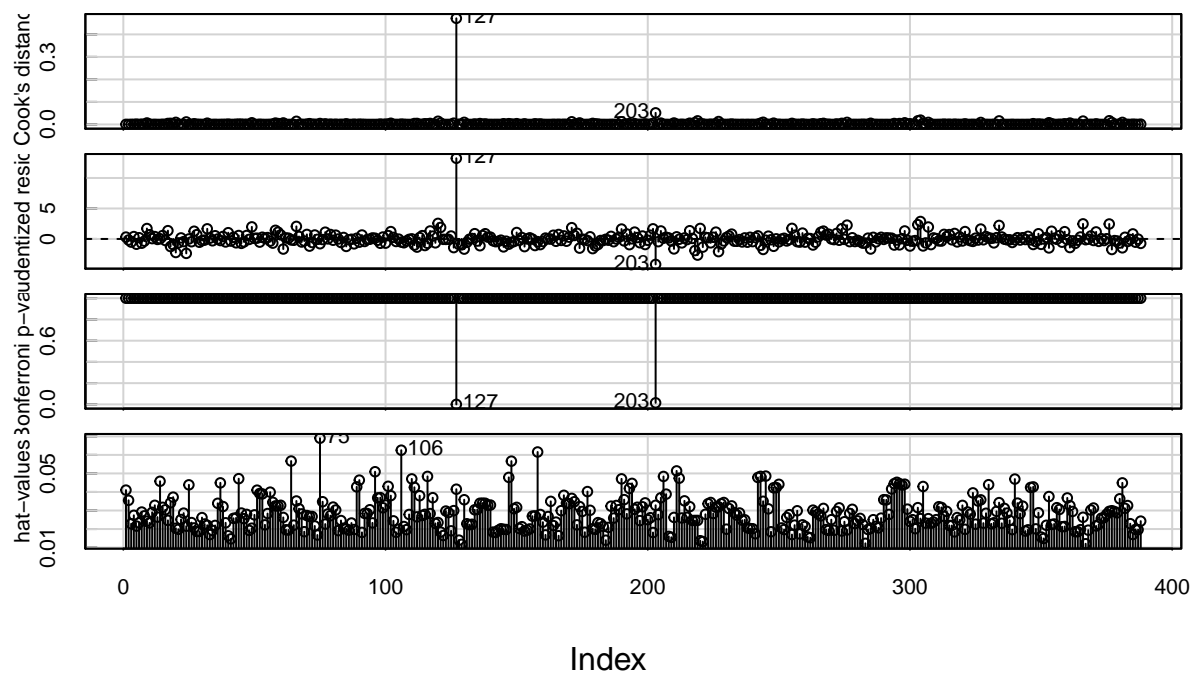


```
##          Test stat Pr(>|Test stat|)
## Tukey test  -2.1052      0.03527 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##          rstudent unadjusted p-value Bonferonni p
## 127 13.233846      4.3594e-33    1.6914e-30
## 203 -4.171956      3.7540e-05    1.4566e-02
```

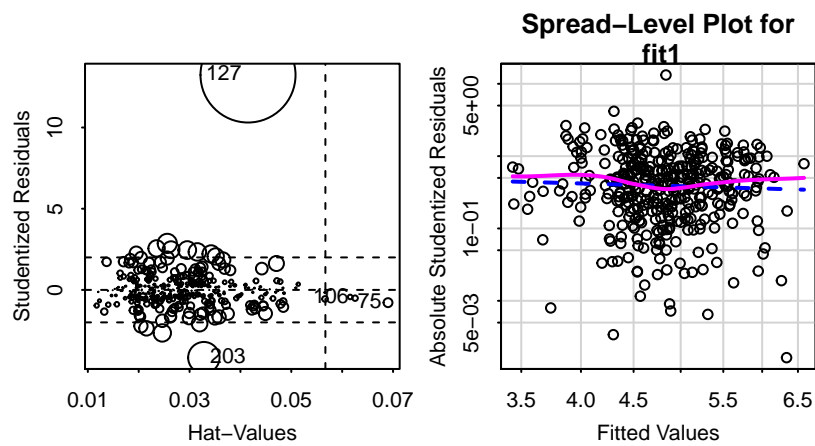


## Diagnostic Plots

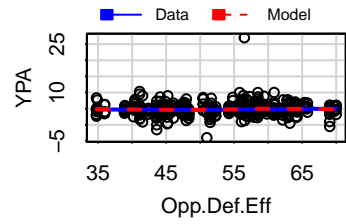
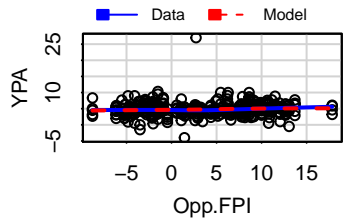
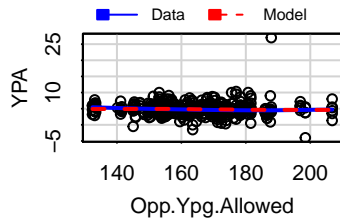
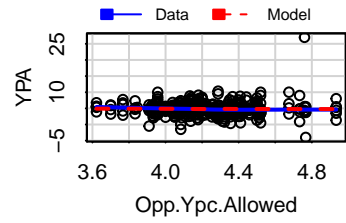
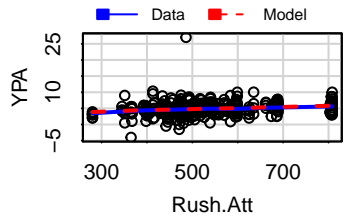
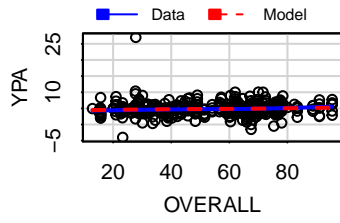
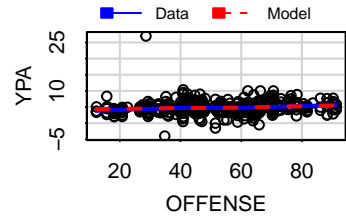
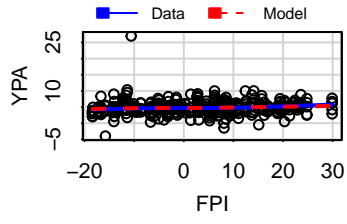
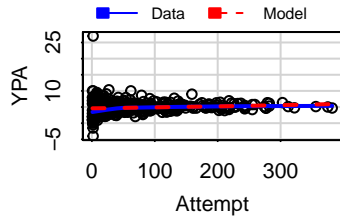


```
##          StudRes      Hat      CookD
## 75 -0.7850719 0.06900846 0.004157431
## 106 -0.5092579 0.06253954 0.001575936
## 127 13.2338464 0.04148945 0.471415509
## 203 -4.1719556 0.03280861 0.051435541

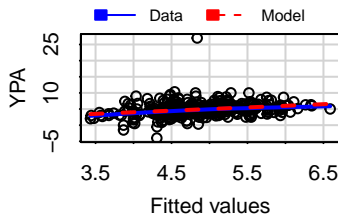
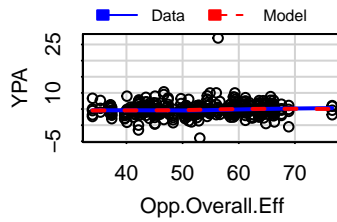
##
## Suggested power transformation: 1.39552
```







## Marginal Model Plots



```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
## collapse

## Linear mixed-effects model fit by REML
## Data: Player_rush1
##      AIC      BIC    logLik
## 1744.991 1796.11 -859.4956
##
## Random effects:
## Formula: ~1 | TEAM
##      (Intercept) Residual
## StdDev:   0.4407102 2.025776
##
## Fixed effects: YPA ~ Attempt + FPI + OFFENSE + OVERALL + Rush.Att + Opp.Ypc.Allowed + Opp.Ypg.A
##
##              Value Std.Error DF   t-value p-value
## (Intercept)   0.8437510  4.534111 301   0.1860896  0.8525
## Attempt       0.0029694  0.001332 301   2.2295292  0.0265
## FPI           0.0658718  0.053026  76   1.2422640  0.2180
## OFFENSE       0.0136492  0.010934  76   1.2482831  0.2158
## OVERALL      -0.0483389  0.028754  76  -1.6811307  0.0968
## Rush.Att      0.0033054  0.001210  76   2.7309134  0.0078
## Opp.Ypc.Allowed 0.8380446  1.196742  76   0.7002718  0.4859
## Opp.Ypg.Allowed -0.0103621  0.021556  76  -0.4807123  0.6321
```

```

## Opp.FPI          -0.0527664  0.141528  76 -0.3728337  0.7103
## Opp.Def.Eff      -0.0266671  0.036879  76 -0.7230928  0.4718
## Opp.Overall.Eff  0.0652108  0.086040  76  0.7579171  0.4508
## Correlation:
##               (Intr) Attmpt FPI      OFFENS OVERAL Rsh.At Opp.Ypc.A
## Attempt                0.020
## FPI                    0.093 -0.039
## OFFENSE                -0.025  0.020  0.000
## OVERALL                -0.095  0.016 -0.930 -0.269
## Rush.Att              -0.359 -0.030  0.054 -0.264 -0.040
## Opp.Ypc.Allowed      -0.462 -0.044 -0.036 -0.030  0.029  0.026
## Opp.Ypg.Allowed      0.415  0.059  0.070  0.002 -0.062 -0.030 -0.897
## Opp.FPI              0.749  0.060 -0.282 -0.064  0.235 -0.243 -0.085
## Opp.Def.Eff          -0.200  0.005  0.064  0.072 -0.113  0.143  0.048
## Opp.Overall.Eff     -0.715 -0.051  0.163  0.034 -0.128  0.222  0.027
##               Opp.Ypg.A Op.FPI Op.D.E
## Attempt
## FPI
## OFFENSE
## OVERALL
## Rush.Att
## Opp.Ypc.Allowed
## Opp.Ypg.Allowed
## Opp.FPI              0.251
## Opp.Def.Eff          -0.067   -0.155
## Opp.Overall.Eff     -0.169   -0.913 -0.207
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.95685683 -0.44864227 -0.04353071  0.33510659 10.52293626
##
## Number of Observations: 388
## Number of Groups: 86

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand

##
## Attaching package: 'lme4'

## The following object is masked from 'package:nlme':
##
##     lmList

## Linear mixed model fit by REML ['lmerMod']
## Formula: YPA ~ FPI + (1 | TEAM)
## Data: Player_rush
##
## REML criterion at convergence: 6852.8
##
## Scaled residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -4.5204 -0.3433  0.0000  0.2646 12.5869
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   TEAM     (Intercept)  0.6515  0.8072
##   Residual                31.2633  5.5914
## Number of obs: 1087, groups: TEAM, 93
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.14272   0.19744  20.983
## FPI          0.02082   0.01617   1.288
##
## Correlation of Fixed Effects:
##      (Intr)
## FPI -0.277
##
## (Intercept)      FPI
##  4.14271567  0.02081795
##
##      (Intercept)
## Air Force      0.29889100
## Akron          -0.60987743
## Alabama        -0.18894827
## Arizona         0.02634157
## Arizona State  -0.07386908
## Arkansas       -0.20664178
##
##      (Intercept)      FPI
## Air Force      4.441607 0.02081795
## Akron          3.532838 0.02081795
## Alabama        3.953767 0.02081795
## Arizona        4.169057 0.02081795
## Arizona State  4.068847 0.02081795
## Arkansas       3.936074 0.02081795

```