Report

Kaiyu Yan

December 4, 2018

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 and table 1.2.

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 : Running Back 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

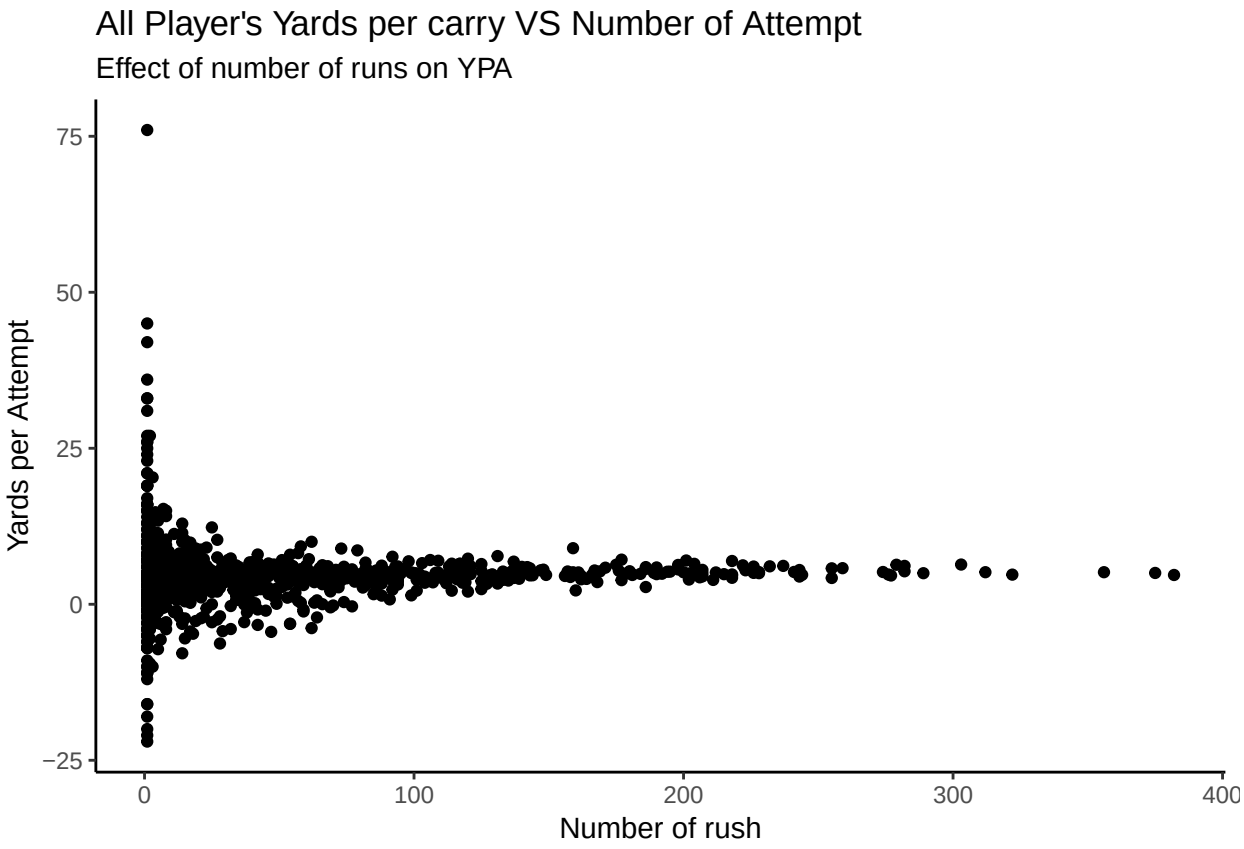
The data used in this project is real statistic data of NCAA College Footbal for the 2012 season. The data include the following information: 1) Attempt: The total number of a player attempt to carry a ball to run in the season.
2) Yards: The total yards of a player gained in the season. 3) YPA: Yards 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 deffense efficiency. 20) Opp.Overall.Eff: Opponent team overall efficiency.

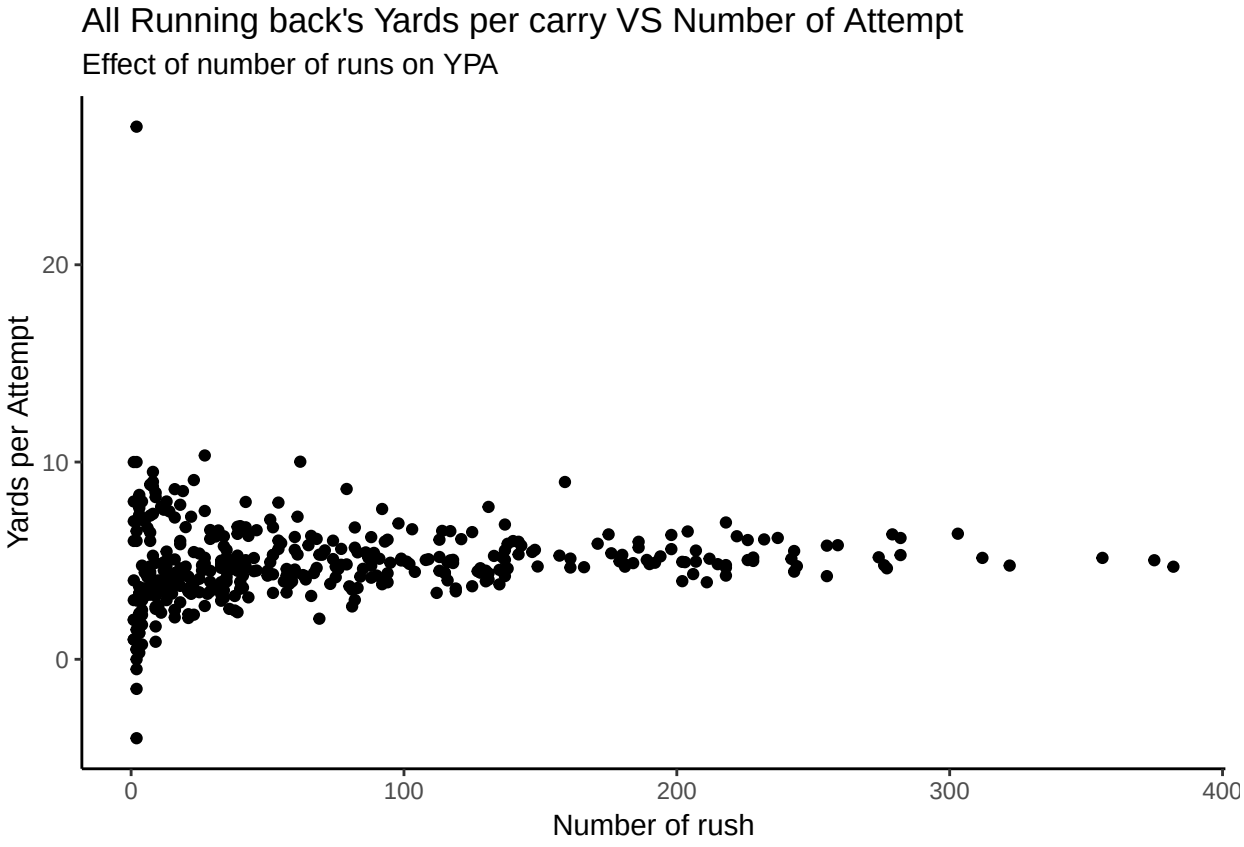
EDA and Result

EDA

As a NCAA college football fan, I have some knowledge of football stats and I also have some expection 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.

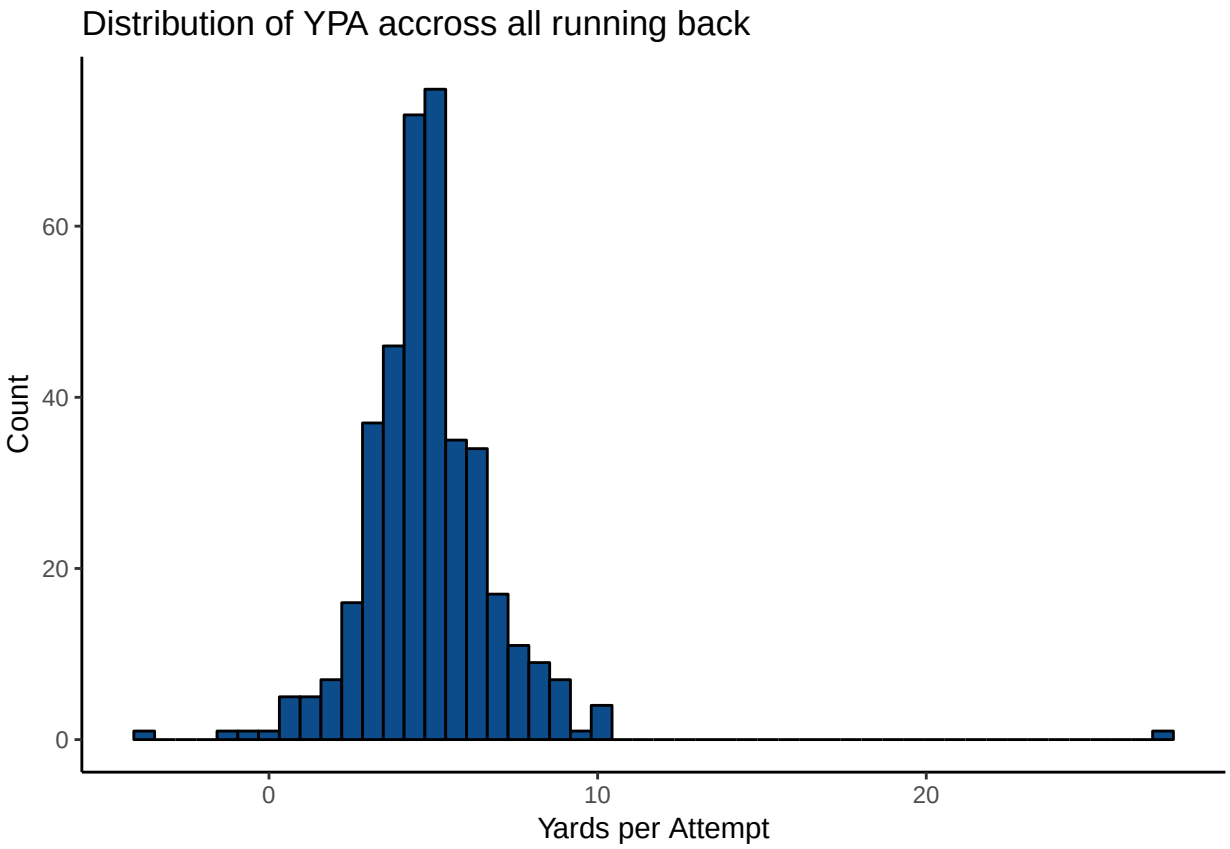
1. Yards per Carry vs. Number of Attempt

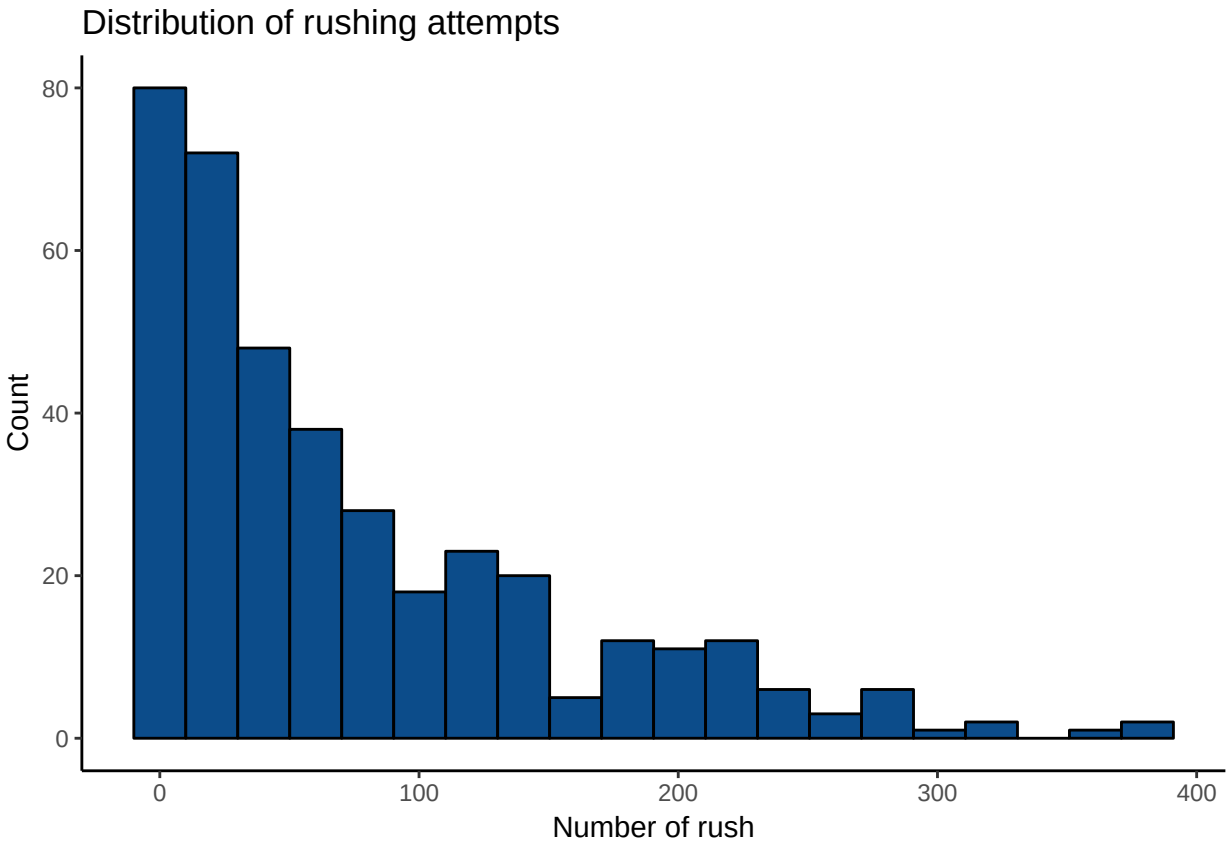




For those two plots, I displayed the relationship and effect between yards per attempt and the number of attempt. it doesn't account for the quality of opposition faced or the strength of own team. You can find that for all player position and running back player, they have similer trend for yards per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

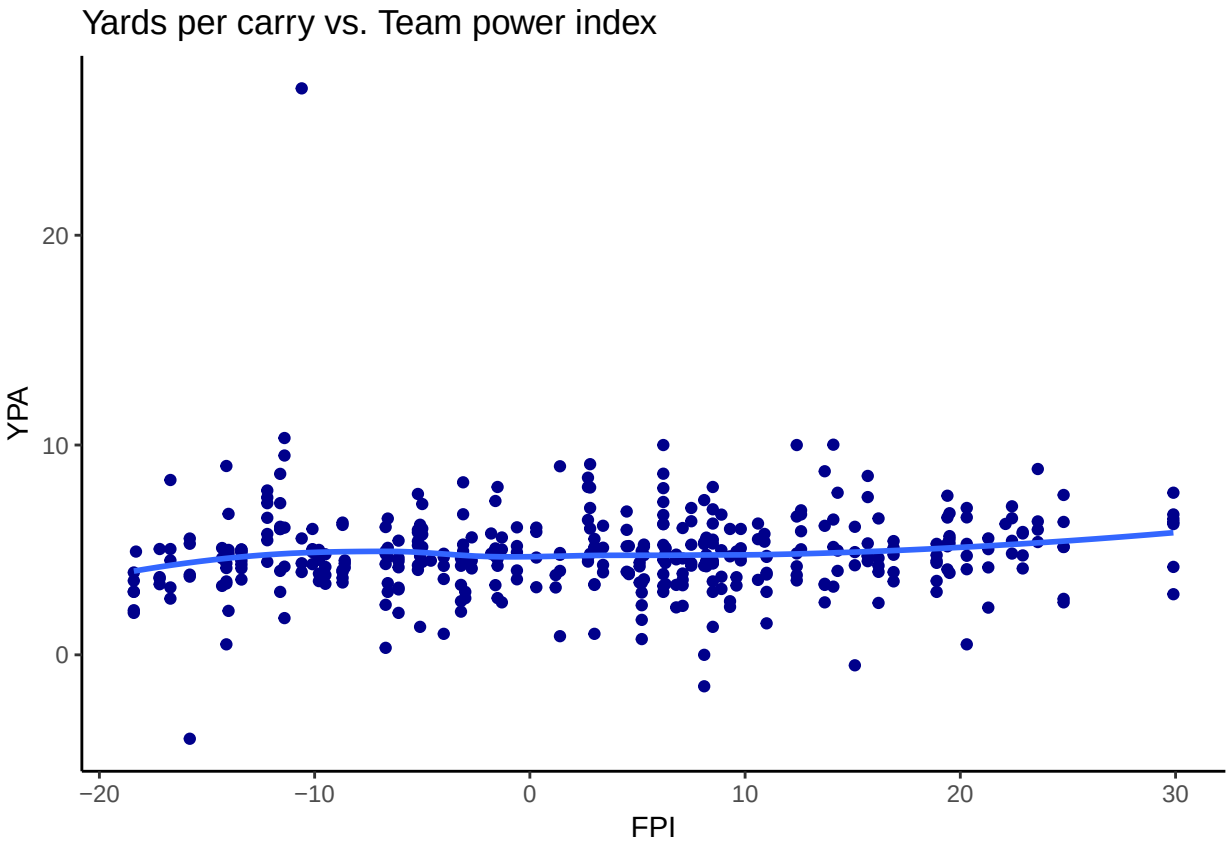
2. Distribution of Rushing Attempt and Yards per Carry

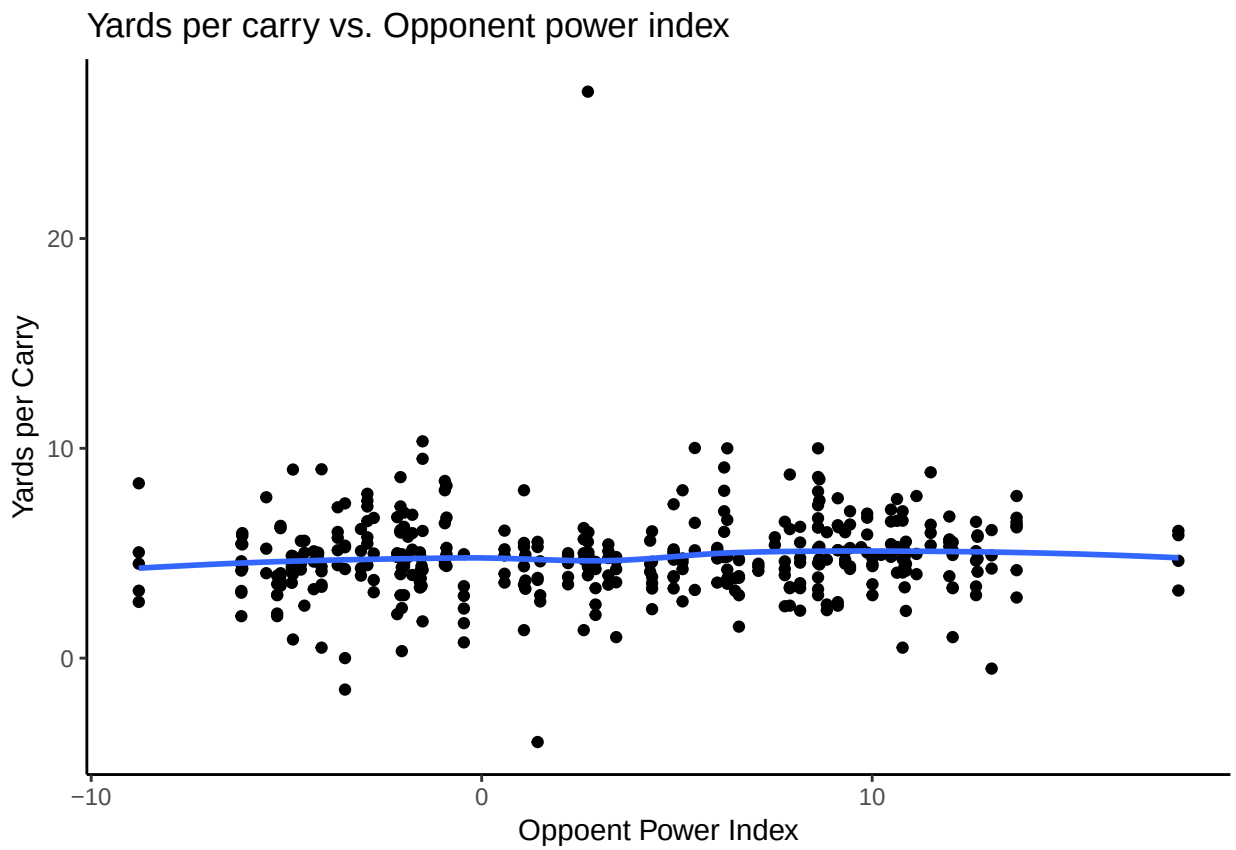




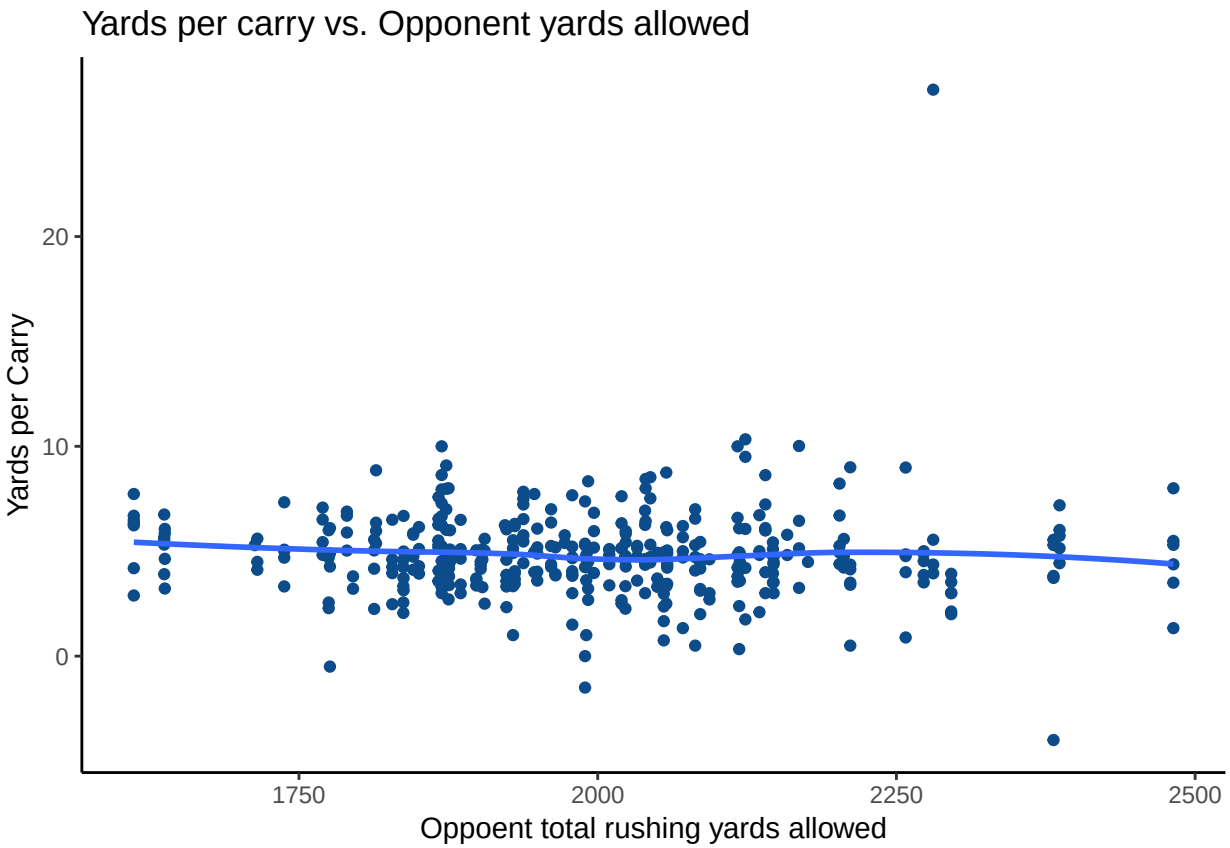
Those two graphs above shows the distribution of YPA and rushing attempts. From both plots I can say that it consist with my expectation. The distribution of YPA is knid of a normal distribution with some outliers and the count of rushing attempts will decreasing with more attempts.

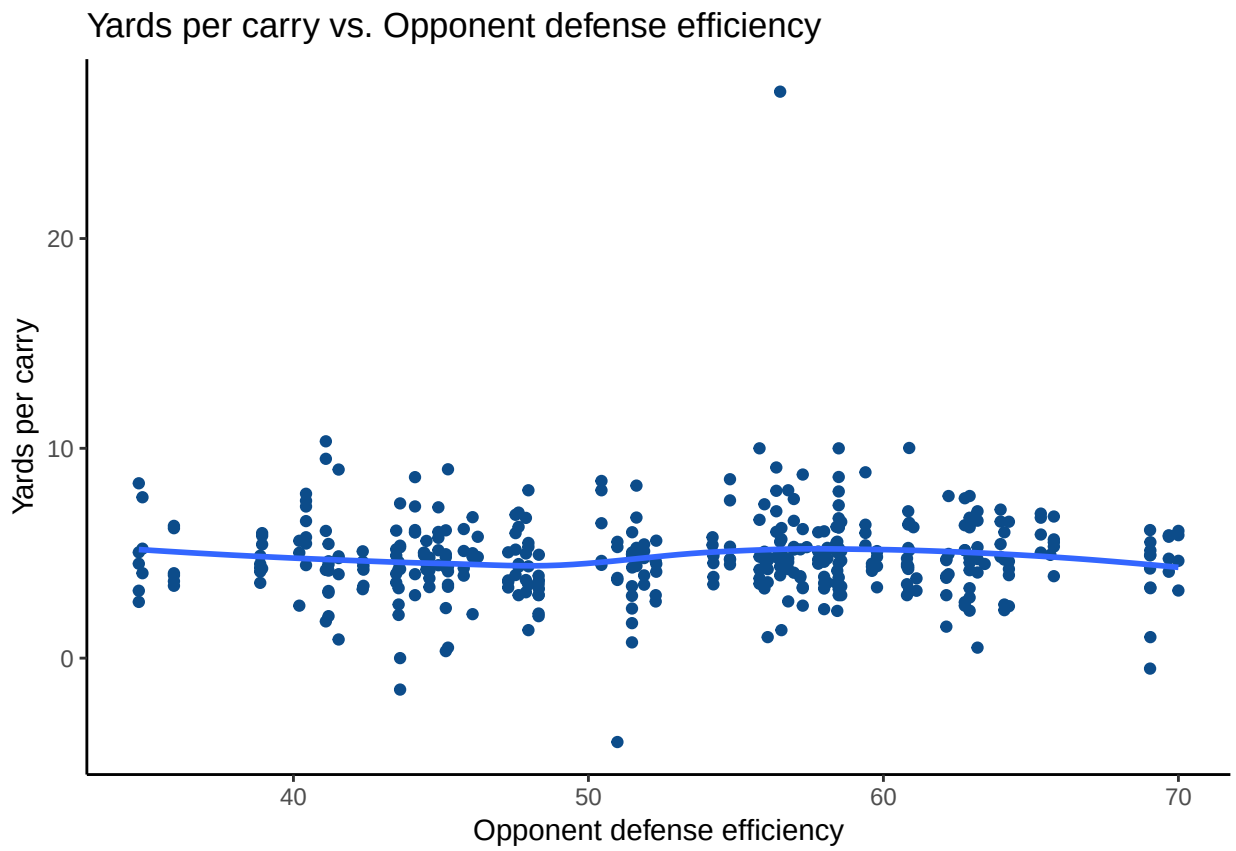
3. Relationship between YPA and other variables.





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.


Yards per carry vs. Opponent yards allowed
Yards per Carry
20
10
0
1750
2000
2250
2500
Oppoent total rushing yards allowed



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:

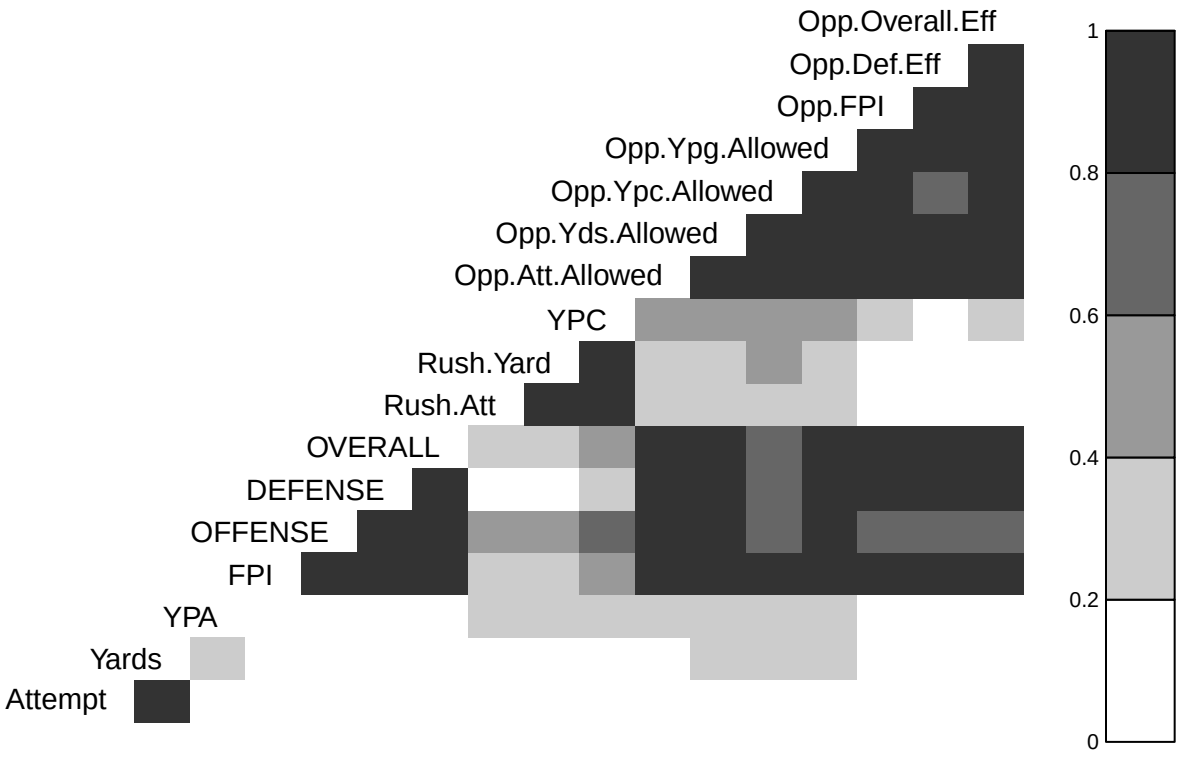
1. Linear Model
2. Multilevel Mixed effect Moderl

Model Choice

Variable selection

Outcome Variable - YPA

Correlation plot



Classic Linear Regression

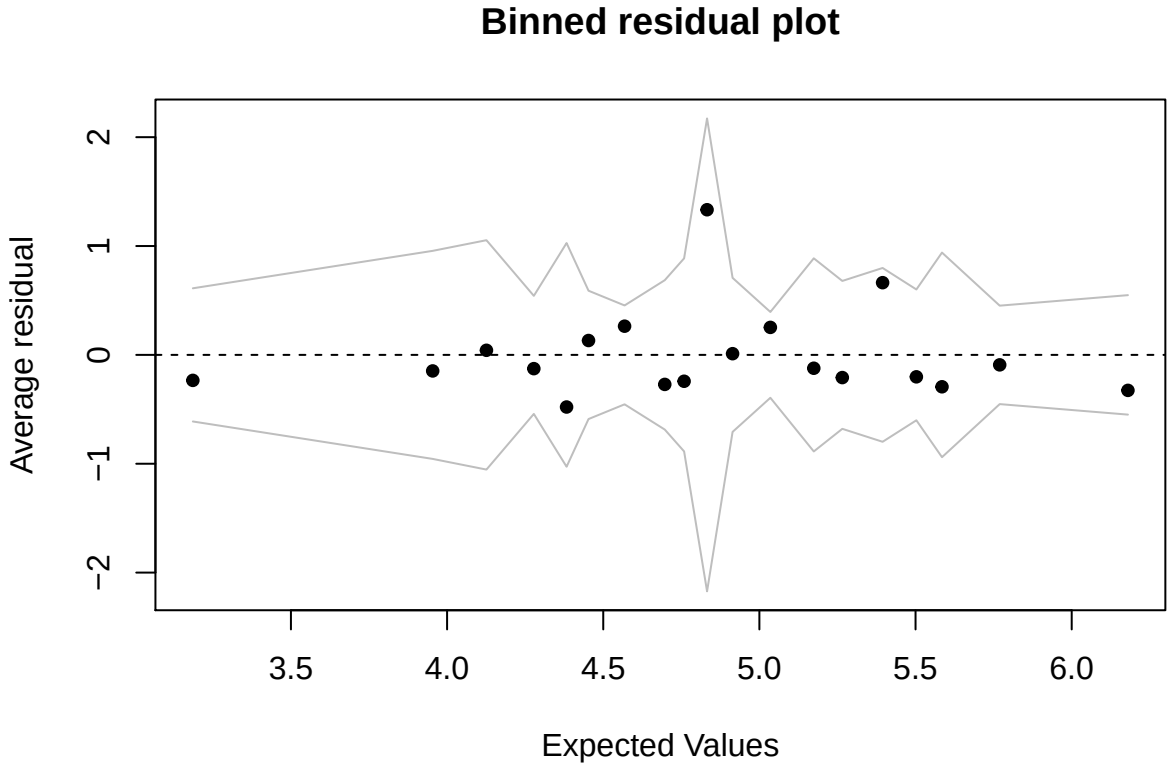
```
FALSE
FALSE Call:
FALSE lm(formula = YPA ~ Attempt + FPI + OVERALL + YPC, data = Player_rush1)
FALSE
FALSE Residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -8.1626 -0.9056 -0.1274  0.6574 22.1335
FALSE
FALSE Coefficients:
FALSE              Estimate Std. Error t value Pr(>|t|)
FALSE (Intercept)  3.646766   1.122536   3.249  0.00126 **
FALSE Attempt      0.002849   0.001304   2.186  0.02942 *
FALSE FPI          0.086571   0.037585   2.303  0.02180 *
FALSE OVERALL     -0.058991   0.021990  -2.683  0.00762 **
FALSE YPC          0.864627   0.142301   6.076 2.98e-09 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE Residual standard error: 2.007 on 383 degrees of freedom
FALSE Multiple R-squared:  0.1096,  Adjusted R-squared:  0.1003
FALSE F-statistic: 11.78 on 4 and 383 DF,  p-value: 4.888e-09

FALSE
```

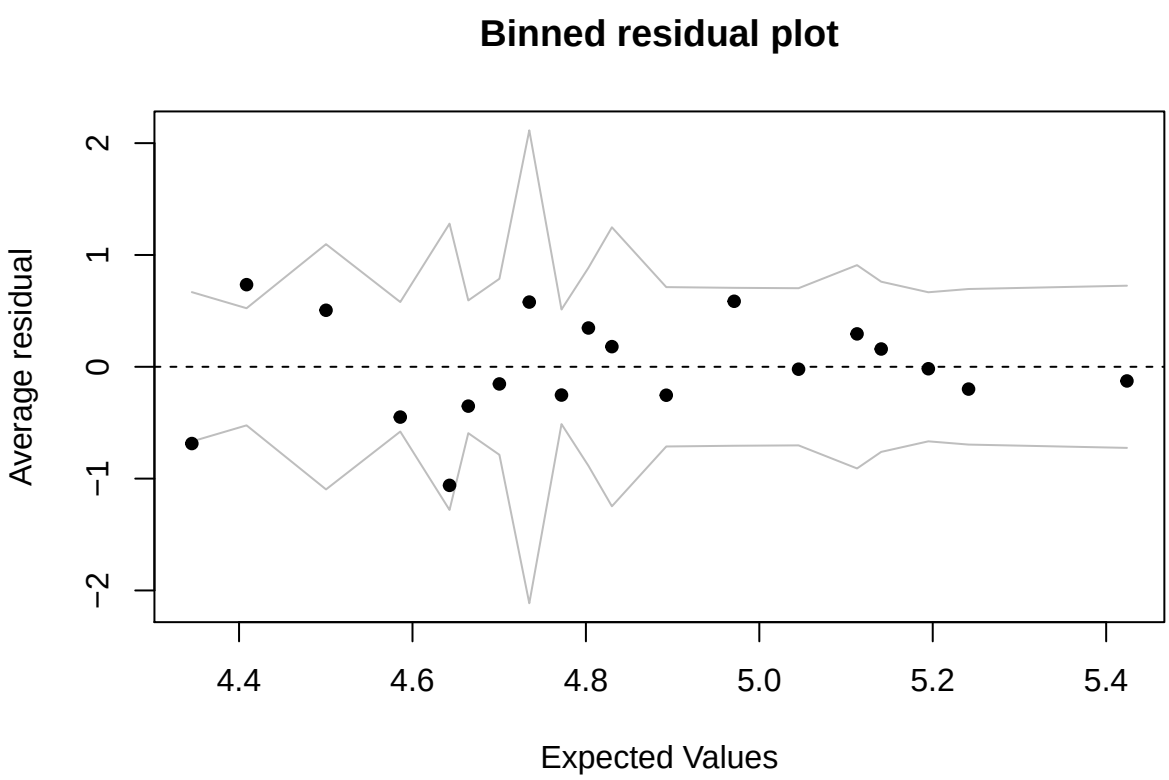
```
FALSE Call:
FALSE lm(formula = YPA ~ Opp.Att.Allowed + Opp.Yds.Allowed + Opp.Ypc.Allowed +
FALSE     Opp.Ypg.Allowed + Opp.FPI + Opp.Def.Eff + Opp.Overall.Eff,
FALSE     data = Player_rush1)
FALSE
FALSE Residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -8.6363 -1.0065 -0.0959  0.8960 22.2809
FALSE
FALSE Coefficients:
FALSE                   Estimate Std. Error t value Pr(>|t|)
FALSE (Intercept)     -2.390870  18.250066  -0.131    0.896
FALSE Opp.Att.Allowed  0.018375   0.037332   0.492    0.623
FALSE Opp.Yds.Allowed  0.002111   0.011749   0.180    0.858
FALSE Opp.Ypc.Allowed  3.050736   4.156220   0.734    0.463
FALSE Opp.Ypg.Allowed -0.086351   0.096288  -0.897    0.370
FALSE Opp.FPI          0.154861   0.126594   1.223    0.222
FALSE Opp.Def.Eff     -0.054461   0.037326  -1.459    0.145
FALSE Opp.Overall.Eff -0.035965   0.077753  -0.463    0.644
FALSE
FALSE Residual standard error: 2.115 on 380 degrees of freedom
FALSE Multiple R-squared:  0.01909, Adjusted R-squared:  0.001022
FALSE F-statistic: 1.057 on 7 and 380 DF,  p-value: 0.3911
```

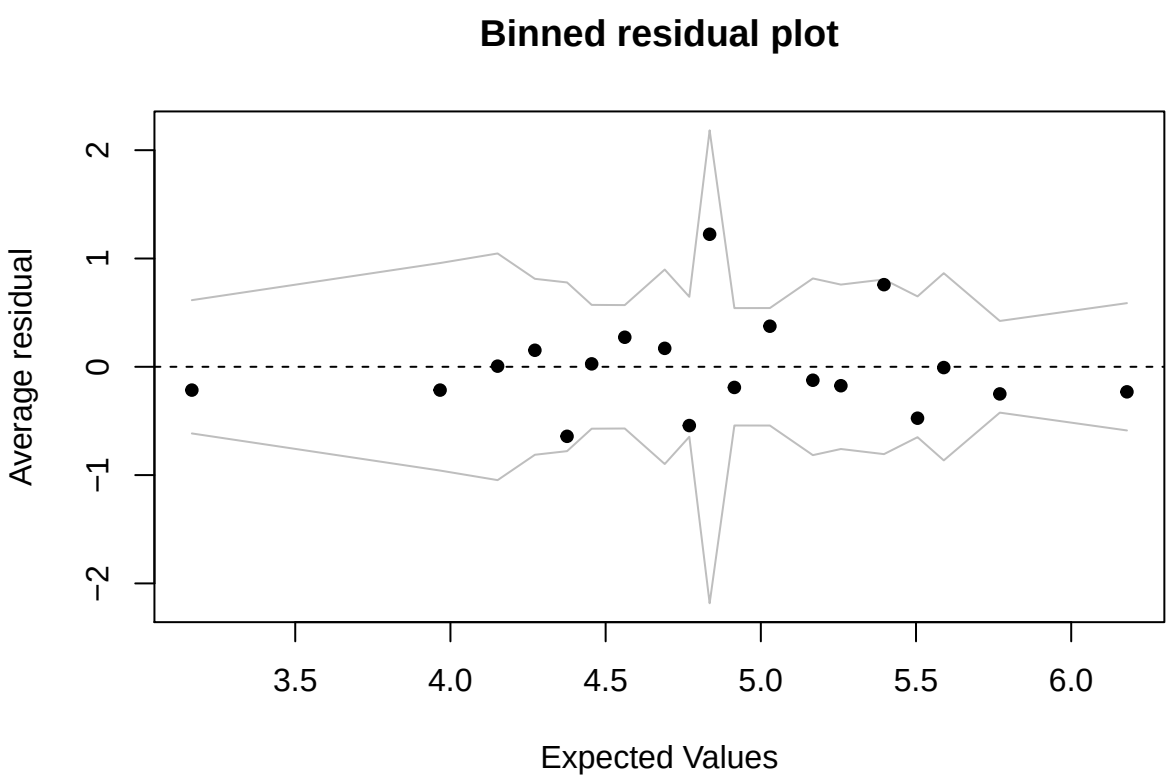
```
FALSE
FALSE Call:
FALSE lm(formula = YPA ~ Attempt + Rush.Att + OFFENSE + YPC + OVERALL +
FALSE     FPI + YPC:Opp.Ypc.Allowed, data = Player_rush1)
FALSE
FALSE Residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -8.1077 -0.9315 -0.1264  0.6650 22.1488
FALSE
FALSE Coefficients:
FALSE                         Estimate Std. Error t value Pr(>|t|)
FALSE (Intercept)          3.563e+00  1.162e+00   3.067  0.00231 **
FALSE Attempt              2.834e-03  1.309e-03   2.165  0.03103 *
FALSE Rush.Att             4.467e-05  1.228e-03   0.036  0.97100
FALSE OFFENSE             -3.368e-03  1.060e-02  -0.318  0.75081
FALSE YPC                  9.684e-01  4.899e-01   1.977  0.04878 *
FALSE OVERALL             -5.636e-02  2.324e-02  -2.425  0.01577 *
FALSE FPI                  8.529e-02  3.893e-02   2.191  0.02909 *
FALSE YPC:Opp.Ypc.Allowed -1.910e-02  9.700e-02  -0.197  0.84401
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE Residual standard error: 2.014 on 380 degrees of freedom
FALSE Multiple R-squared:  0.1099,  Adjusted R-squared:  0.0935
FALSE F-statistic: 6.703 on 7 and 380 DF,  p-value: 1.688e-07
```

```
FALSE      df      AIC
FALSE fit1  6 1648.632
FALSE fit2  9 1692.187
FALSE fit3  9 1654.494
```

```
FALSE Analysis of Variance Table
FALSE
FALSE Model 1: YPA ~ Attempt + FPI + OVERALL + YPC
FALSE Model 2: YPA ~ Opp.Att.Allowed + Opp.Yds.Allowed + Opp.Ypc.Allowed + Opp.Ypg.Allowed +
FALSE     Opp.FPI + Opp.Def.Eff + Opp.Overall.Eff
FALSE Model 3: YPA ~ Attempt + Rush.Att + OFFENSE + YPC + OVERALL + FPI + YPC:Opp.Ypc.Allowed
FALSE   Res.Df    RSS Df Sum of Sq F Pr(>F)
FALSE 1    383 1542.6
FALSE 2    380 1699.4  3   -156.78
FALSE 3    380 1542.1  0    157.33
```




Binned residual plot
Average residual
2
1
0
−1
−2
4.4
4.6
4.8
5.0
5.2
5.4
Expected Values



Based on the output and compare the AIC and residual plot, I think model 3 is better than others. Looking

at the summary output, we can see that Attempt, YPC, OVERALL, and FPI are significant. But I think it is still meaningfull to keep the rest variables.

With each unit increase in Attempt, the YPA (Yard per attempt) increase by 0.003. One unit increase in YPC, the YPA increases by 0.97. One unit increase in OVERALL Efficiency, the YPA decrease by 0.057. With each unit increase in FPI, the YPA would increase by 0.085.

Multilevel Modelling

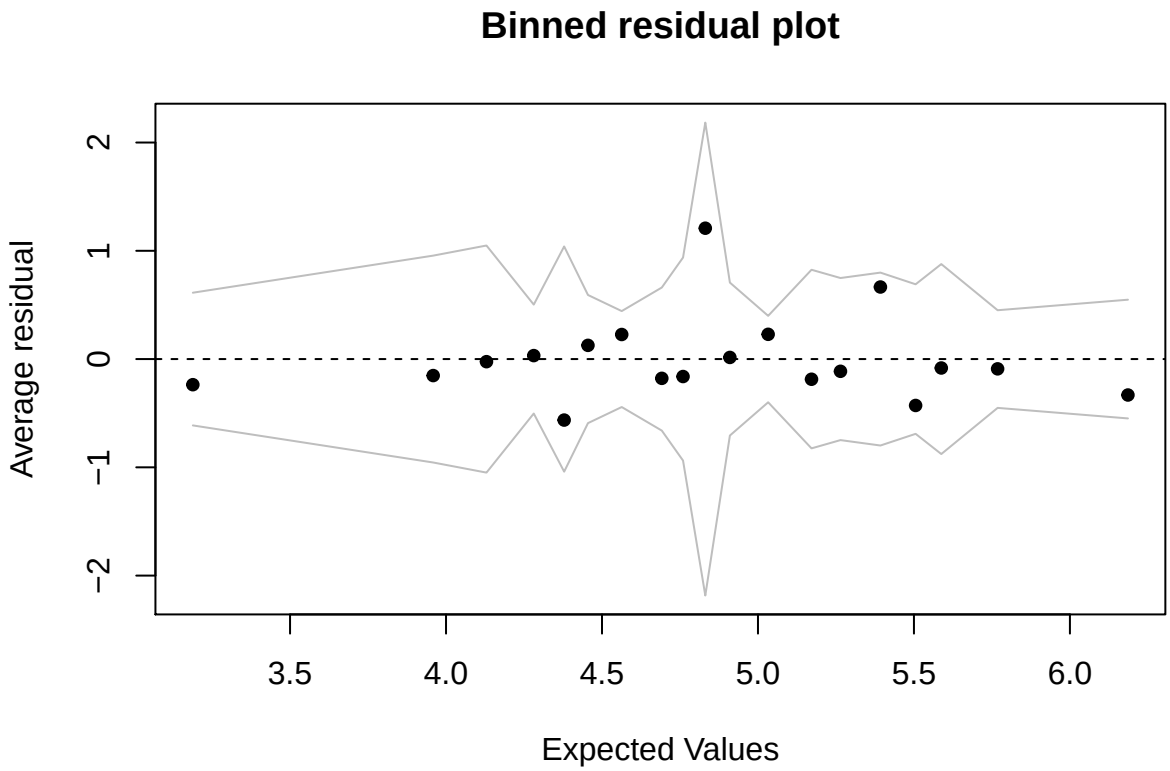
```
FALSE Linear mixed model fit by REML ['lmerMod']
FALSE Formula:
FALSE YPA ~ Attempt + FPI + OFFENSE + Rush.Att + YPC + Opp.Ypc.Allowed +
FALSE     (1 | TEAM)
FALSE    Data: Player_rush1
FALSE
FALSE REML criterion at convergence: 1685.2
FALSE
FALSE Scaled residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -4.0003 -0.4649 -0.0508  0.3516 11.0007
FALSE
FALSE Random effects:
FALSE  Groups   Name        Variance Std.Dev.
```

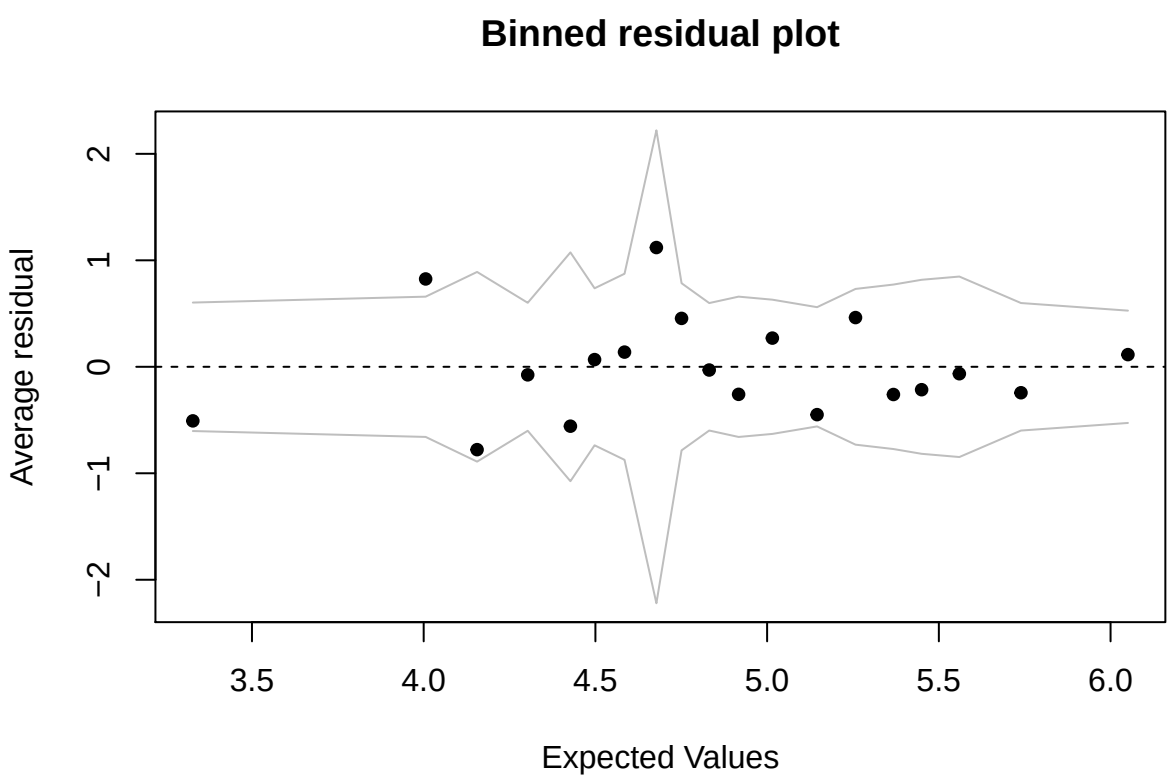
```
FALSE  TEAM     (Intercept) 0.002068 0.04548
FALSE  Residual             4.109455 2.02718
FALSE Number of obs: 388, groups:  TEAM, 86
FALSE
FALSE Fixed effects:
FALSE                    Estimate Std. Error t value
FALSE (Intercept)      2.352e+00  1.942e+00   1.212
FALSE Attempt          2.827e-03  1.318e-03   2.145
FALSE FPI             -2.422e-03  1.472e-02  -0.165
FALSE OFFENSE         -9.643e-03  1.032e-02  -0.934
FALSE Rush.Att        -1.636e-05  1.238e-03  -0.013
FALSE YPC              8.809e-01  2.019e-01   4.362
FALSE Opp.Ypc.Allowed -2.634e-01  4.434e-01  -0.594
FALSE
FALSE Correlation of Fixed Effects:
FALSE             (Intr) Attmpt FPI    OFFENS Rsh.At YPC
FALSE Attempt     -0.050
FALSE FPI         -0.053 -0.056
FALSE OFFENSE     -0.049  0.039 -0.738
FALSE Rush.Att    -0.221  0.001  0.105 -0.025
FALSE YPC          0.009 -0.033  0.030 -0.391 -0.582
FALSE Opp.Ypc.All -0.941  0.005  0.194 -0.025  0.168 -0.175
```

```
FALSE Linear mixed model fit by REML ['lmerMod']
FALSE Formula:
FALSE YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE     (1 | TEAM)
FALSE    Data: Player_rush1
FALSE
FALSE REML criterion at convergence: 1677.8
FALSE
FALSE Scaled residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -4.0447 -0.4525 -0.0696  0.3235 11.0109
FALSE
FALSE Random effects:
FALSE  Groups   Name        Variance Std.Dev.
FALSE  TEAM     (Intercept) 0.000    0.000
FALSE  Residual             4.049    2.012
FALSE Number of obs: 388, groups:  TEAM, 86
FALSE
FALSE Fixed effects:
FALSE                    Estimate Std. Error t value
FALSE (Intercept)      3.796249   2.010516   1.888
FALSE Attempt          0.002849   0.001307   2.180
FALSE FPI              0.085669   0.038998   2.197
FALSE YPC              0.860433   0.185408   4.641
FALSE OVERALL         -0.058564   0.022495  -2.603
FALSE Rush.Att         0.000057   0.001227   0.046
FALSE Opp.Ypc.Allowed -0.042540   0.448294  -0.095
FALSE
FALSE Correlation of Fixed Effects:
FALSE             (Intr) Attmpt FPI    YPC    OVERAL Rsh.At
FALSE Attempt     -0.048
```

```
FALSE FPI          0.251 -0.018
FALSE YPC          0.022 -0.021  0.002
FALSE OVERALL     -0.292  0.007 -0.968 -0.110
FALSE Rush.Att    -0.204  0.002  0.063 -0.635 -0.031
FALSE Opp.Ypc.All -0.826  0.004  0.257 -0.174 -0.198  0.171
```

```
FALSE    df      AIC
FALSE m2  9 1695.785
FALSE m1  9 1703.205
```






```
FALSE Data: Player_rush1
FALSE Models:
FALSE m2: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m2:     (1 | TEAM)
FALSE m1: YPA ~ Attempt + FPI + OFFENSE + Rush.Att + YPC + Opp.Ypc.Allowed +
FALSE m1:     (1 | TEAM)
FALSE    Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
FALSE m2  9 1654.6 1690.3 -818.31   1636.6
FALSE m1  9 1660.6 1696.2 -821.29   1642.6     0      0          1
```

The above models were created by using the fixed effect to YPA (Yard per Attempt), controlling for by-Team variability. Looking at their binnedplot and anova test, I think model m2 is better since all the points in binned residual plot are within the range and it has the lower AIC, BIC and deviance. From the output,

when other variables remain constant, each unit increase of Attempt, the YPA increases by 0.003 on average; with every unit increase of own team FPI, YPA increases 0.085669 ; One unit increase in YPC , the YPA increase by 0.86. For the team overall efficiency factor, with each unit increase , the YPA would decrease by 0.05; one unit increase in opponent yard per carry allowed, YPA will decreases 0.04,which is meaningful in reality becasue the better running defense on opponent teams keep the runner running less.

```
FALSE Linear mixed model fit by REML ['lmerMod']
FALSE Formula:
```

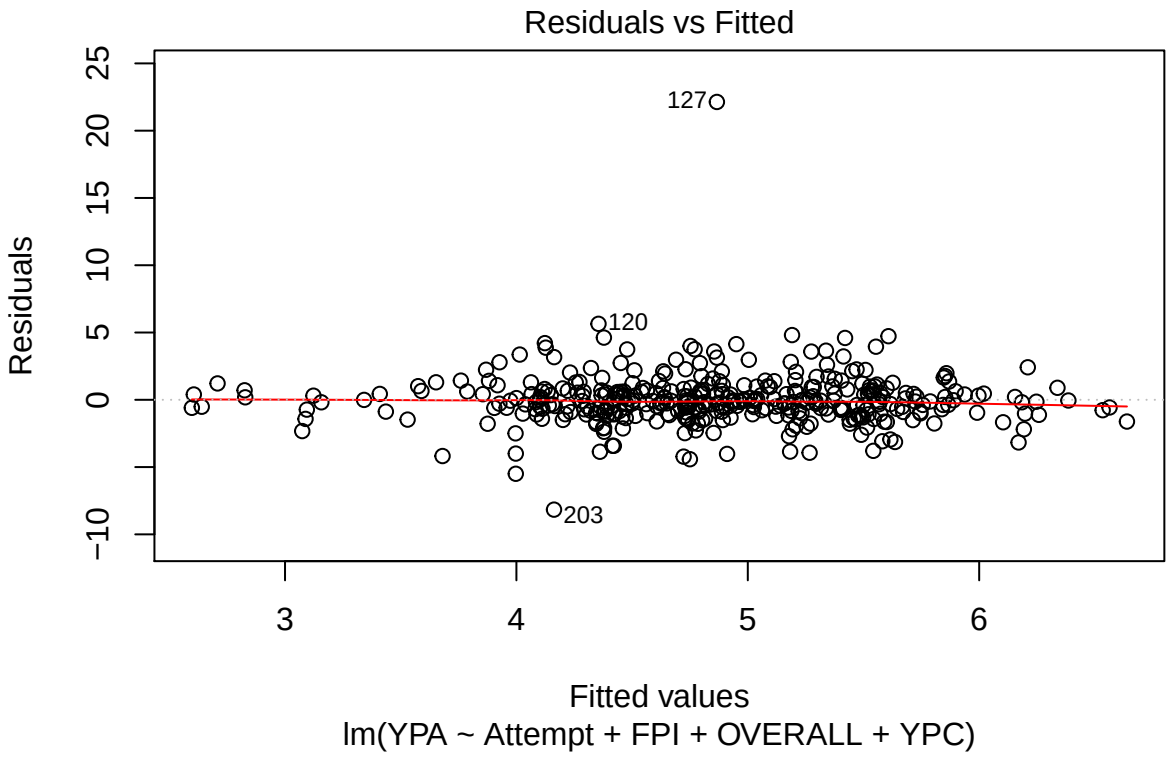
```
FALSE YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE     (1 + Attempt | TEAM)
FALSE    Data: Player_rush1
FALSE
FALSE REML criterion at convergence: 1654.3
FALSE
FALSE Scaled residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -3.5233 -0.4273 -0.0085  0.3485  8.2916
FALSE
FALSE Random effects:
FALSE  Groups   Name        Variance  Std.Dev. Corr
FALSE  TEAM     (Intercept) 2.2378146 1.49593
FALSE           Attempt     0.0001142 0.01068  -1.00
FALSE  Residual             3.1119128 1.76406
FALSE Number of obs: 388, groups:  TEAM, 86
FALSE
FALSE Fixed effects:
FALSE                    Estimate Std. Error t value
FALSE (Intercept)      3.4493113  2.0818648   1.657
FALSE Attempt          0.0021924  0.0017279   1.269
FALSE FPI              0.0585669  0.0412188   1.421
FALSE YPC              0.7680344  0.2065659   3.718
FALSE OVERALL         -0.0398169  0.0240158  -1.658
FALSE Rush.Att        -0.0001711  0.0014121  -0.121
FALSE Opp.Ypc.Allowed -0.0344910  0.4707240  -0.073
FALSE
FALSE Correlation of Fixed Effects:
FALSE             (Intr) Attmpt FPI    YPC    OVERAL Rsh.At
FALSE Attempt     -0.009
FALSE FPI          0.249  0.028
FALSE YPC         -0.001 -0.027  0.024
FALSE OVERALL     -0.282 -0.044 -0.967 -0.132
FALSE Rush.Att    -0.162 -0.047  0.047 -0.617 -0.021
FALSE Opp.Ypc.All -0.815 -0.027  0.262 -0.149 -0.209  0.102
FALSE convergence code: 0
FALSE Model is nearly unidentifiable: very large eigenvalue
FALSE  - Rescale variables?
```

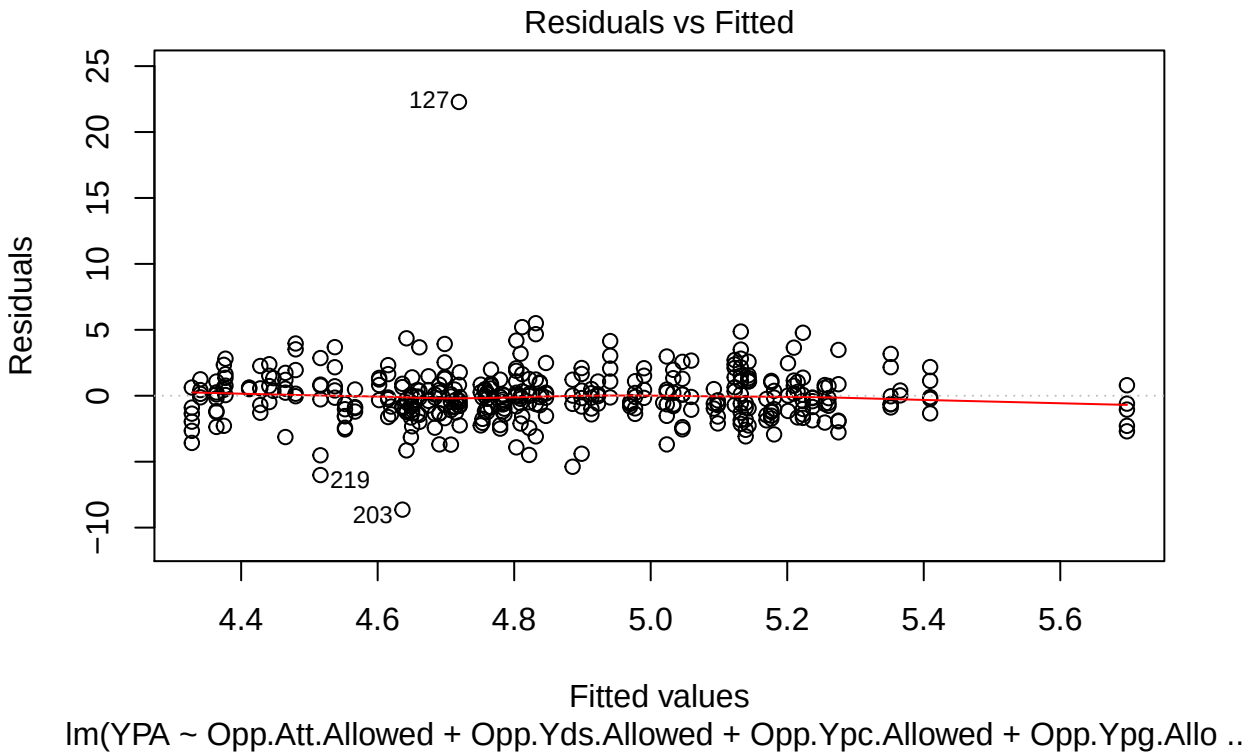
One unit increase in Attempt, the YPA (Yard per attempt) increase by 0.002. One unit increase in FPI, the YPA increase by 0.06. With each unit increase in team YPC (Yard per carry), the YPA for individual player would increase by 0.77. Compared to own team, when the opponent yard per carry allowed changed one unit, the YPA will decrease by 0.034.

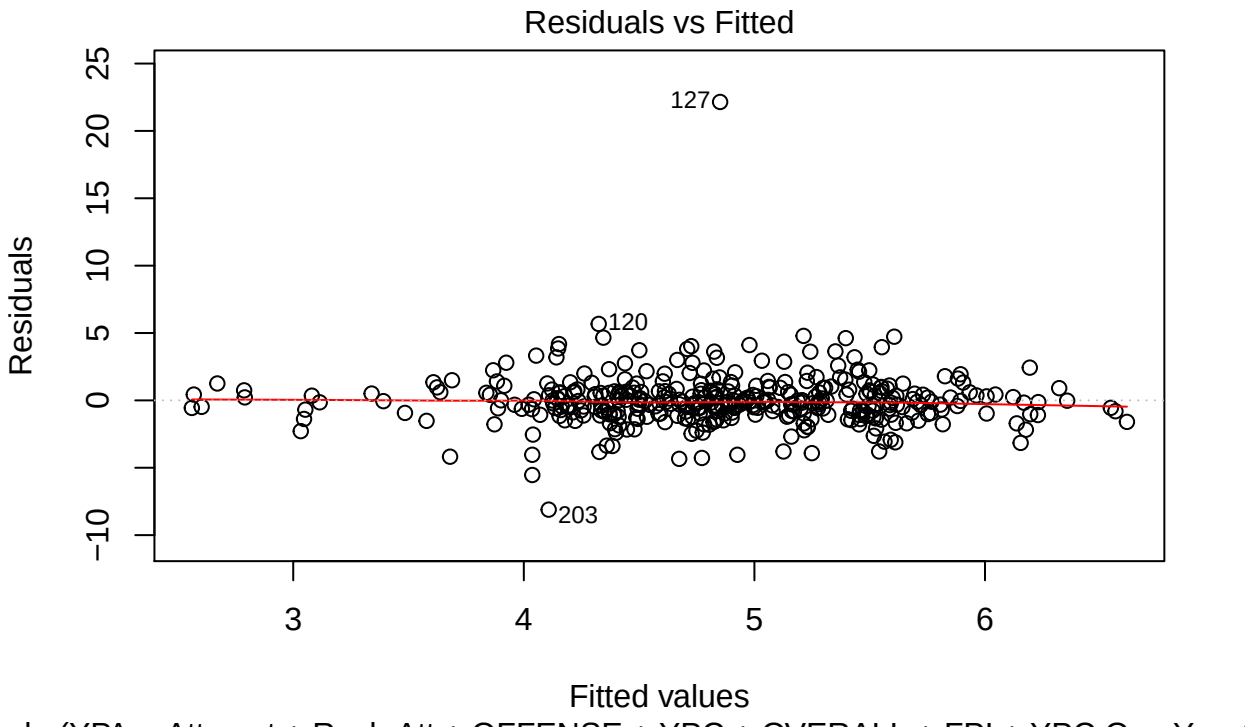
```
FALSE Linear mixed model fit by REML ['lmerMod']
FALSE Formula:
FALSE YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE     (1 | TEAM) + (1 + Attempt | TEAM)
FALSE    Data: Player_rush1
FALSE
FALSE REML criterion at convergence: 1654.3
```

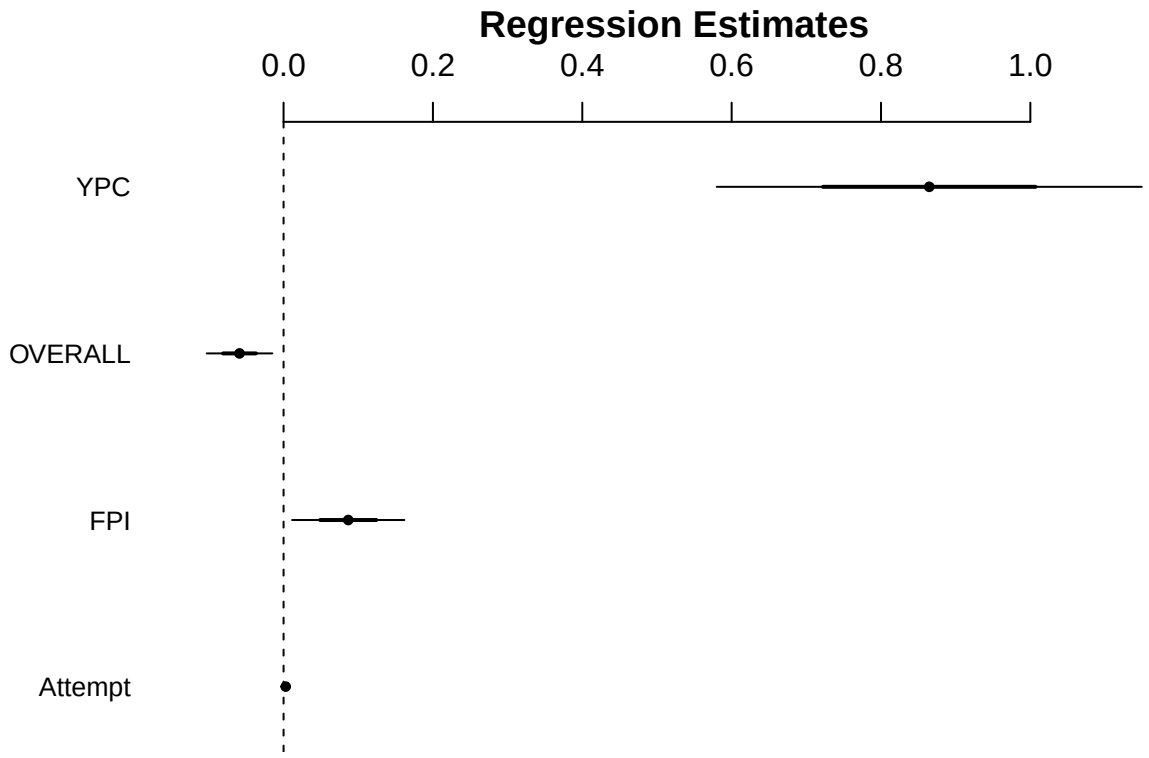
```
FALSE
FALSE Scaled residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -3.5233 -0.4273 -0.0085  0.3485  8.2916
FALSE
FALSE Random effects:
FALSE  Groups   Name        Variance  Std.Dev. Corr
FALSE  TEAM     (Intercept) 0.0000000 0.00000
FALSE  TEAM.1   (Intercept) 2.2377728 1.49592
FALSE           Attempt     0.0001142 0.01068  -1.00
FALSE  Residual             3.1119215 1.76406
FALSE Number of obs: 388, groups:  TEAM, 86
FALSE
FALSE Fixed effects:
FALSE                    Estimate Std. Error t value
FALSE (Intercept)      3.4493080  2.0818664   1.657
FALSE Attempt          0.0021924  0.0017279   1.269
FALSE FPI              0.0585671  0.0412188   1.421
FALSE YPC              0.7680352  0.2065659   3.718
FALSE OVERALL         -0.0398171  0.0240159  -1.658
FALSE Rush.Att        -0.0001711  0.0014121  -0.121
FALSE Opp.Ypc.Allowed -0.0344892  0.4707244  -0.073
FALSE
FALSE Correlation of Fixed Effects:
FALSE             (Intr) Attmpt FPI    YPC    OVERAL Rsh.At
FALSE Attempt     -0.009
FALSE FPI          0.249  0.028
FALSE YPC         -0.001 -0.027  0.024
FALSE OVERALL     -0.282 -0.044 -0.967 -0.132
FALSE Rush.Att    -0.162 -0.047  0.047 -0.617 -0.021
FALSE Opp.Ypc.All -0.815 -0.027  0.262 -0.149 -0.209  0.102
FALSE convergence code: 0
FALSE Model is nearly unidentifiable: very large eigenvalue
FALSE  - Rescale variables?
```

In this model. For the random effect output, it represents the estimated variability in the intercept. For the fixed effect output, the coefficient of Attempt, FPI and YPC (Team Yard per Carry) are positive, which indicates a change in unit would lead to an increase in the YPA.The other variables' coefficients are negative change indicate that in one unit change would lead to a decrease in YPA. ## Model Check

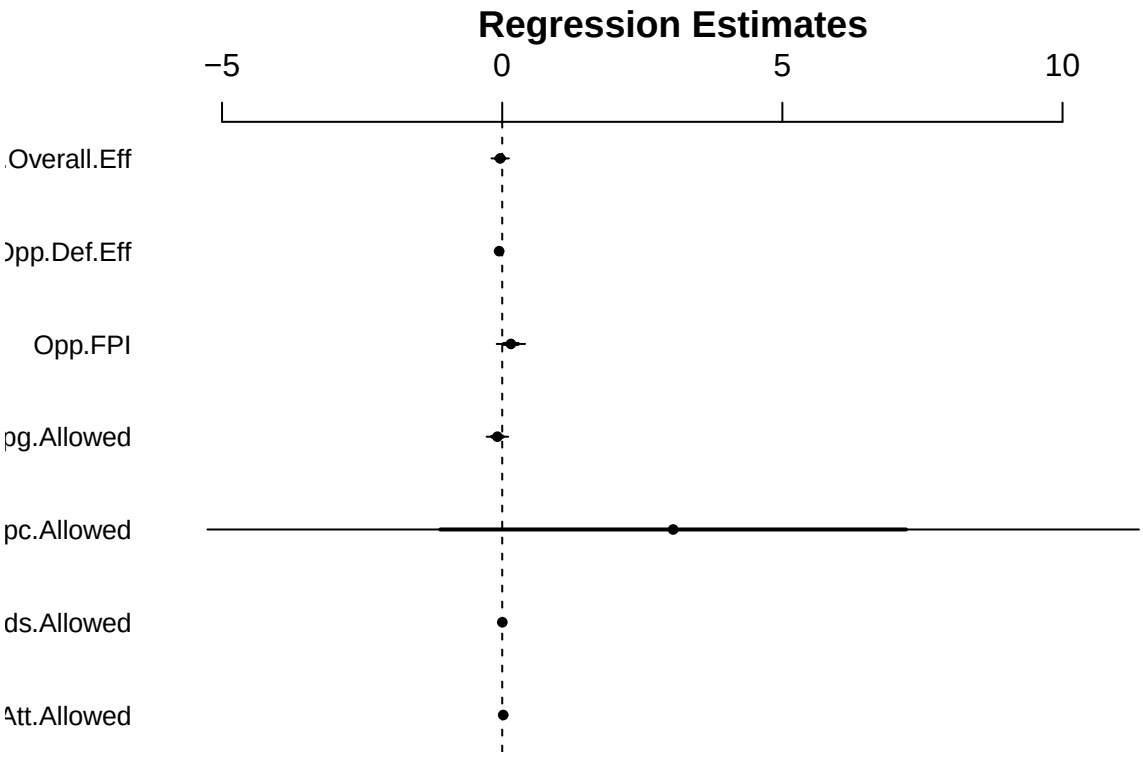

Residuals vs Fitted
Residuals
25
20
15
10
5
0
−10
127
120
203
3
4
5
6
Fitted values
lm(YPA ~ Attempt + FPI + OVERALL + YPC)


Residuals vs Fitted
Residuals
25
20
15
10
5
0
−10
127
219
203
4.4
4.6
4.8
5.0
5.2
5.4
5.6
Fitted values
lm(YPA ~ Opp.Att.Allowed + Opp.Yds.Allowed + Opp.Ypc.Allowed + Opp.Ypg.Allo ..

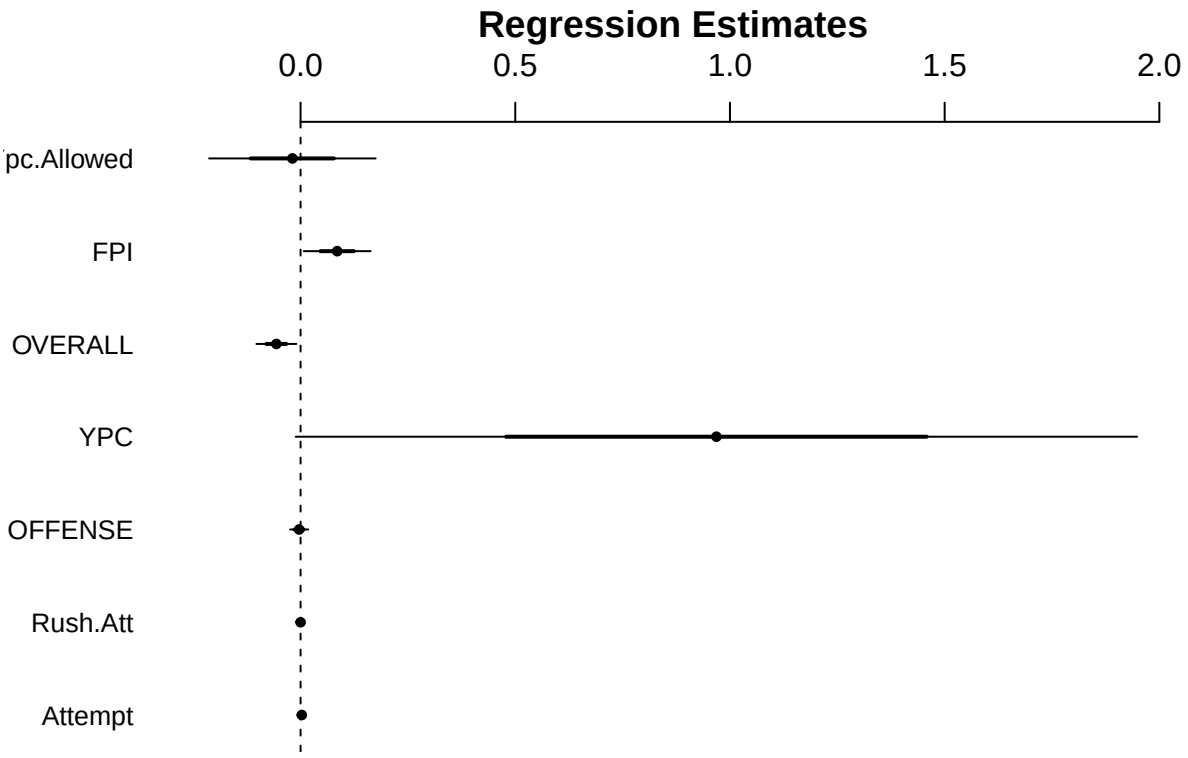

Residuals vs Fitted
Residuals
25
20
15
10
5
0
−10
127
120
203
3
4
5
6
Fitted values
lm(YPA ~ Attempt + Rush.Att + OFFENSE + YPC + OVERALL + FPI + YPC:Opp.Ypc.A



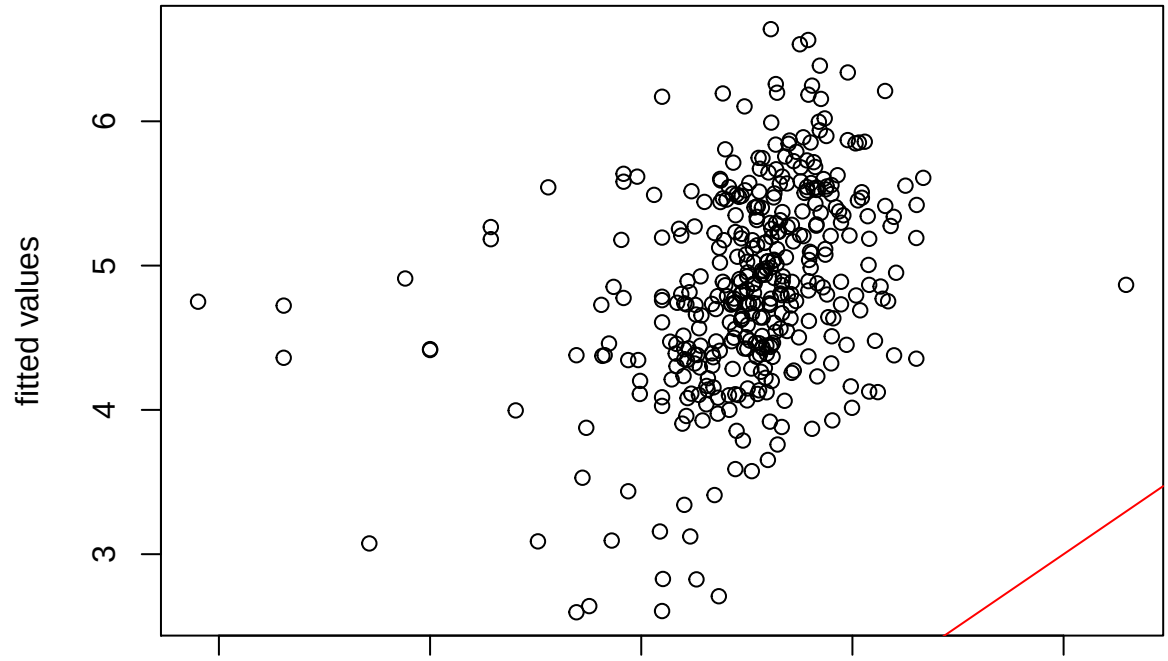
	2.5 %	97.5 %
(Intercept)	1.4396607	5.8538719
Attempt	0.0002865	0.0054123
FPI	0.0126710	0.1604703
OVERALL	-0.1022268	-0.0157549
YPC	0.5848385	1.1444156



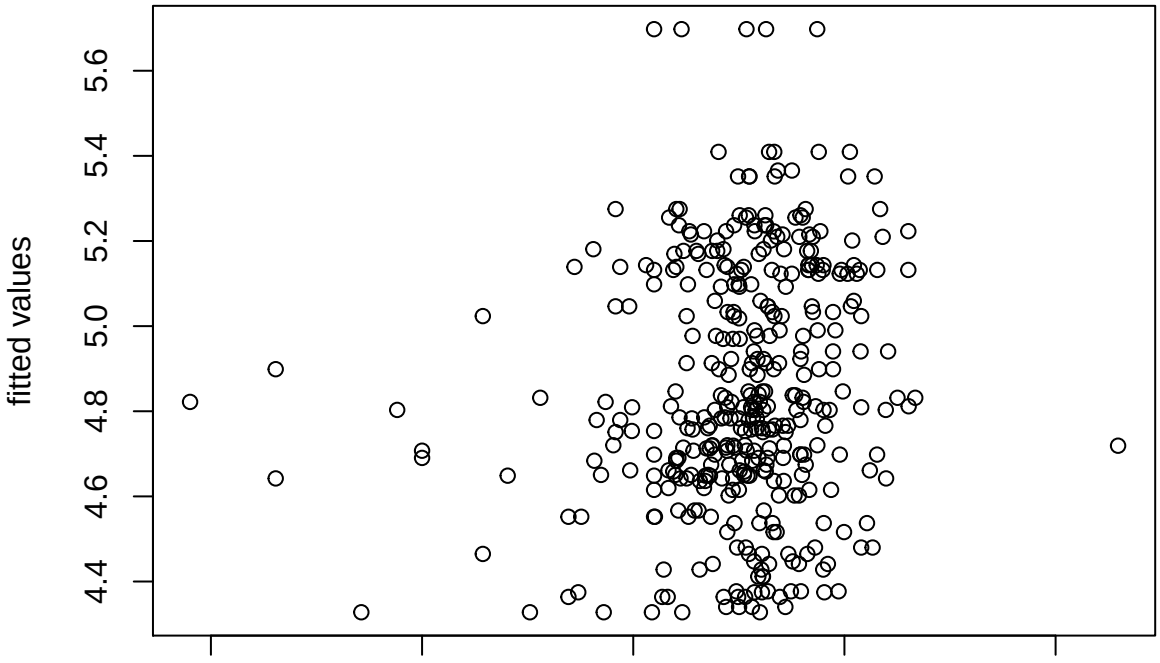
	2.5 %	97.5 %
(Intercept)	-38.2746317	33.4928921
Opp.Att.Allowed	-0.0550274	0.0917774
Opp.Yds.Allowed	-0.0209907	0.0252127
Opp.Ypc.Allowed	-5.1213327	11.2228052
Opp.Ypg.Allowed	-0.2756737	0.1029725
Opp.FPI	-0.0940511	0.4037736



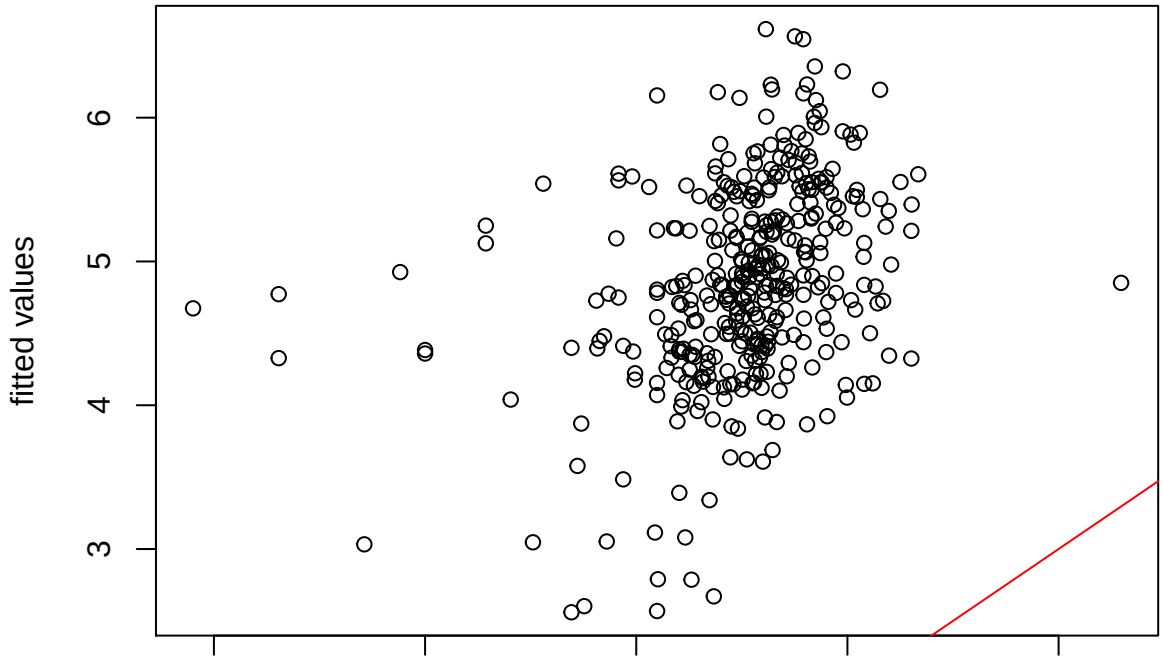
	2.5 %	97.5 %
(Intercept)	1.2792758	5.8475962
Attempt	0.0002599	0.0054088
Rush.Att	-0.0023699	0.0024593
OFFENSE	-0.0242067	0.0174705
YPC	0.0051929	1.9316065
OVERALL	-0.1020645	-0.0106643
FPI	0.0087331	0.1618416
YPC:Opp.Ypc.Allowed	-0.2098138	0.1716156




```
FALSE integer(0)
```

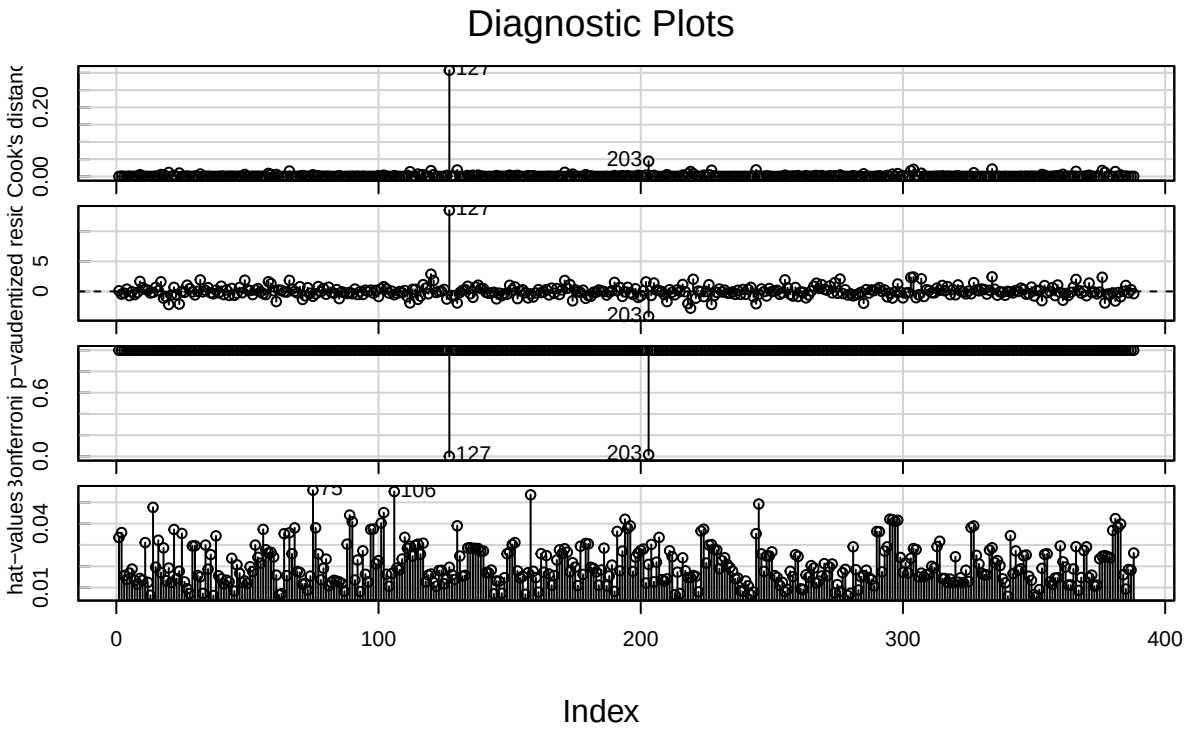


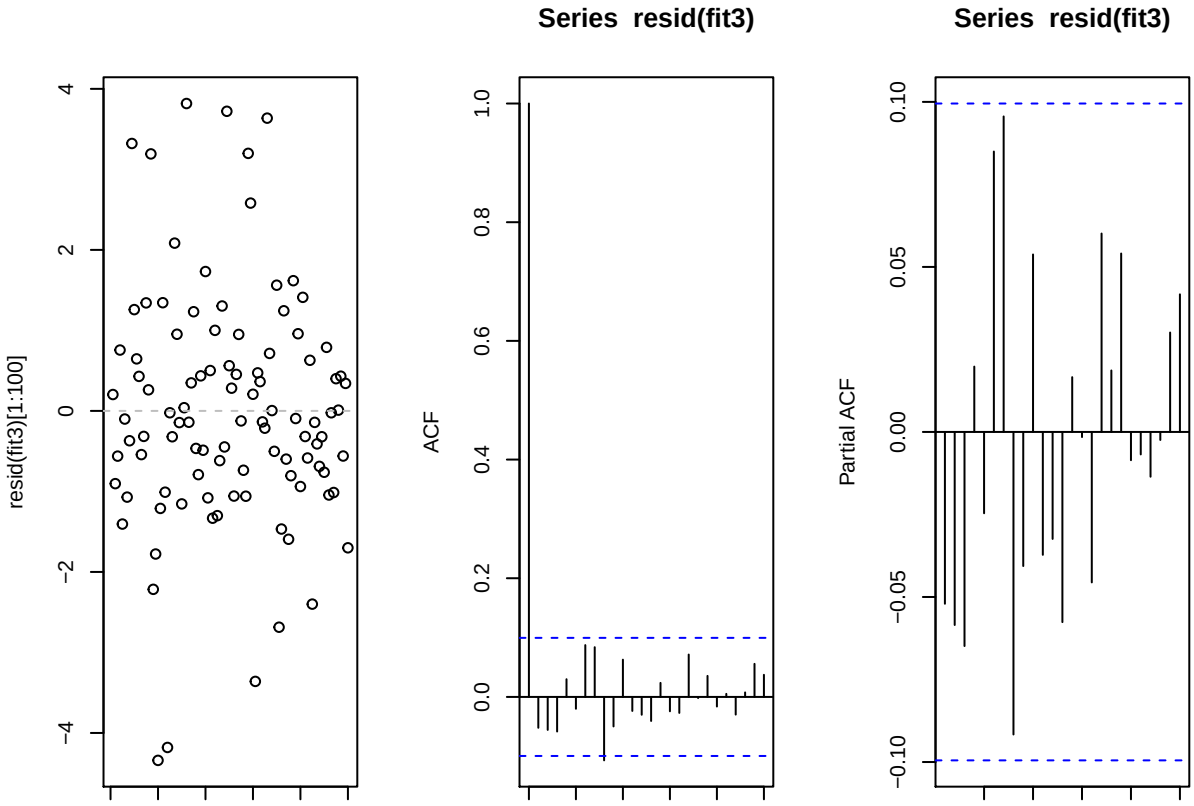

```
FALSE integer(0)
```



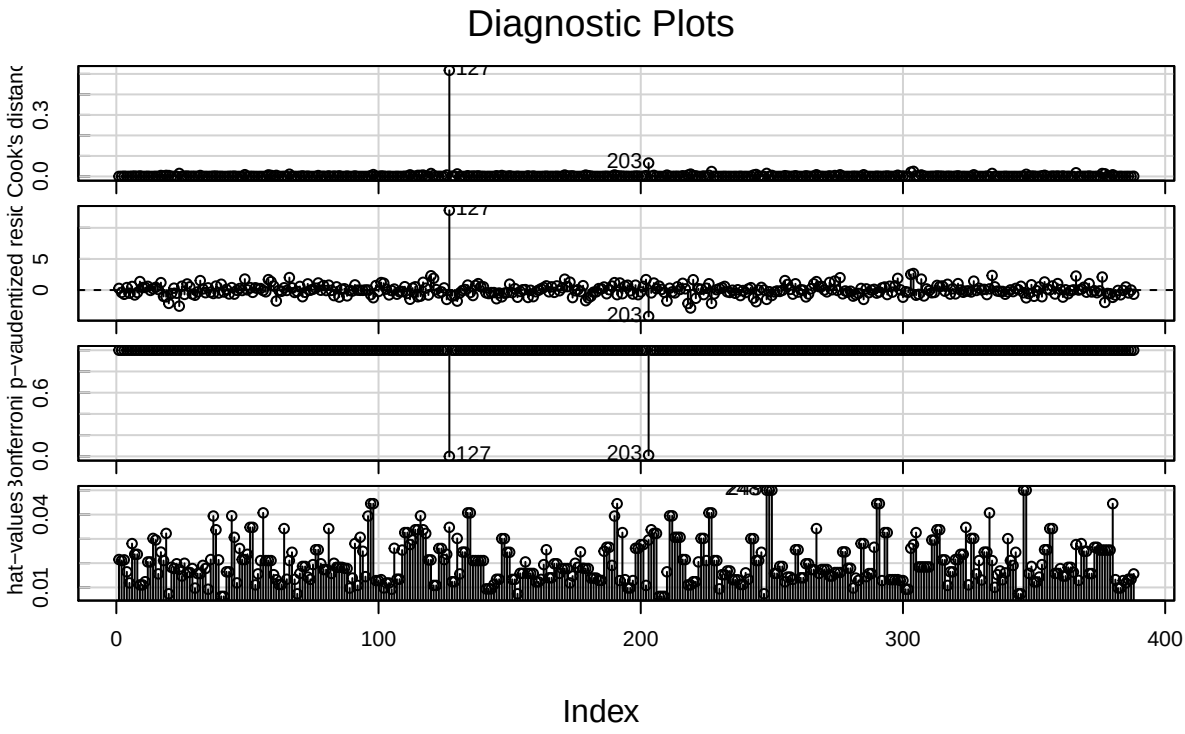

```
FALSE integer(0)
```

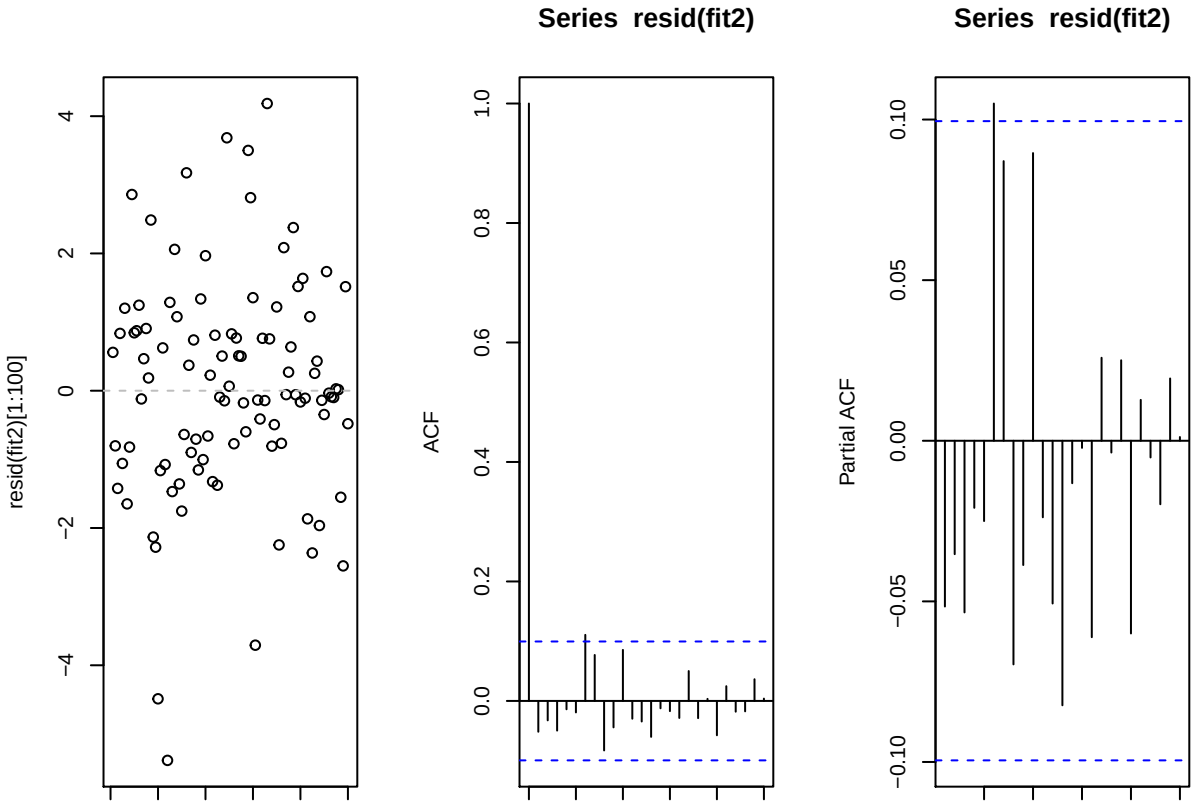
```
FALSE      rstudent unadjusted p-value Bonferonni p
FALSE 127 13.492017         3.7412e-34   1.4516e-31
FALSE 203 -4.153488         4.0492e-05   1.5711e-02
```


Diagnostic Plots
Cook's distanc
0.00
0.20
127
203
standardized resic
0
5
127
203
p−vaudenitized
onferroni
0.0
0.6
127
203
hat−values
3
0.01
0.04
75
106
0
100
200
300
400
Index

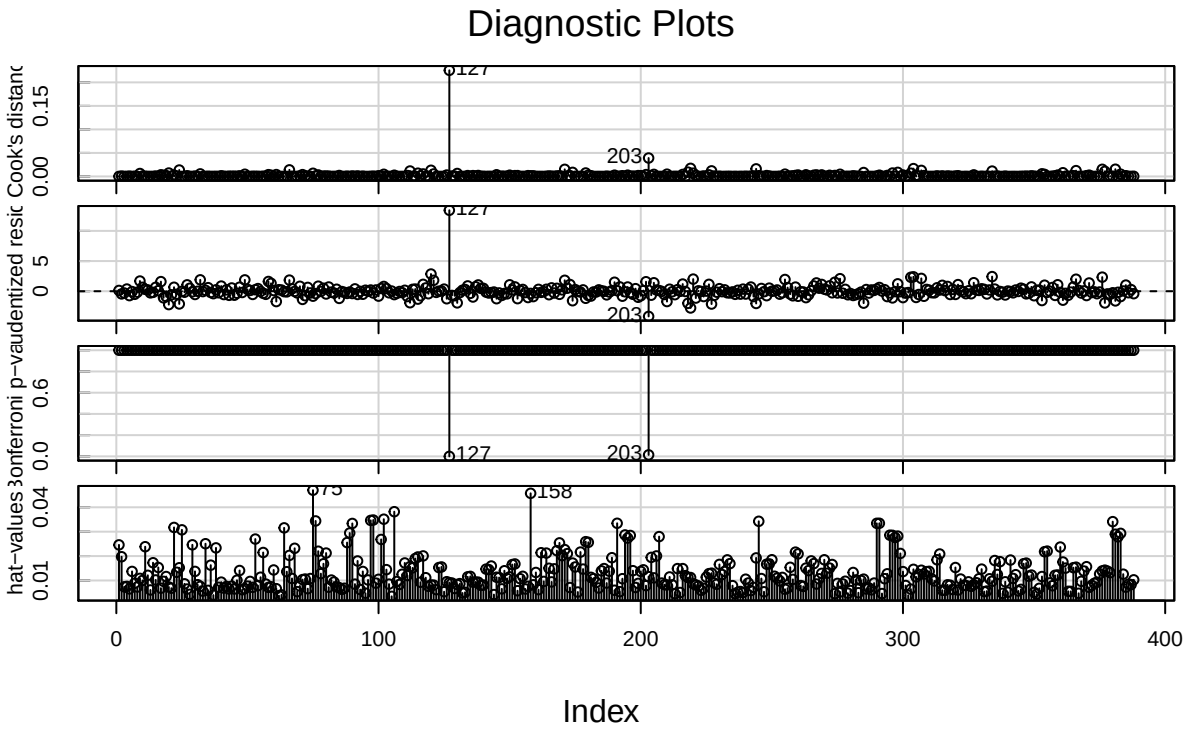


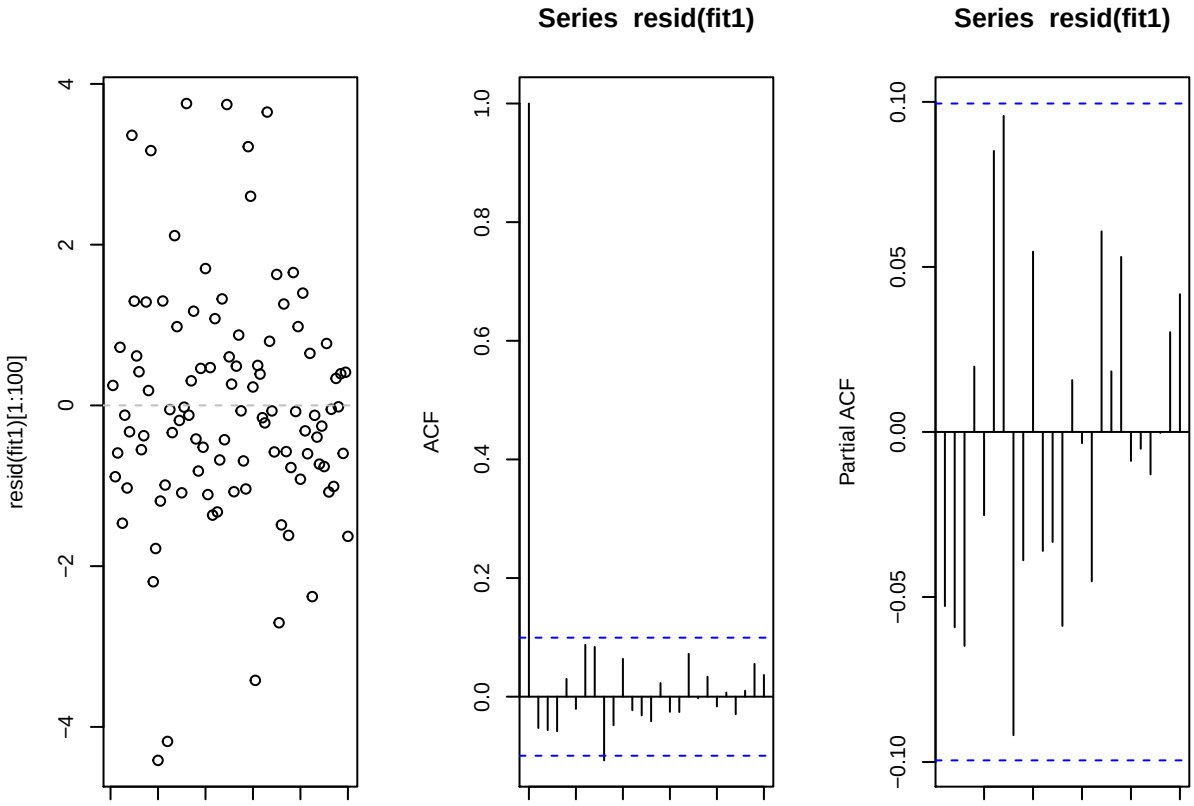

```
FALSE     rstudent unadjusted p-value Bonferonni p
FALSE 127 12.82492         1.6009e-31   6.2113e-29
FALSE 203 -4.23694         2.8483e-05   1.1052e-02
```

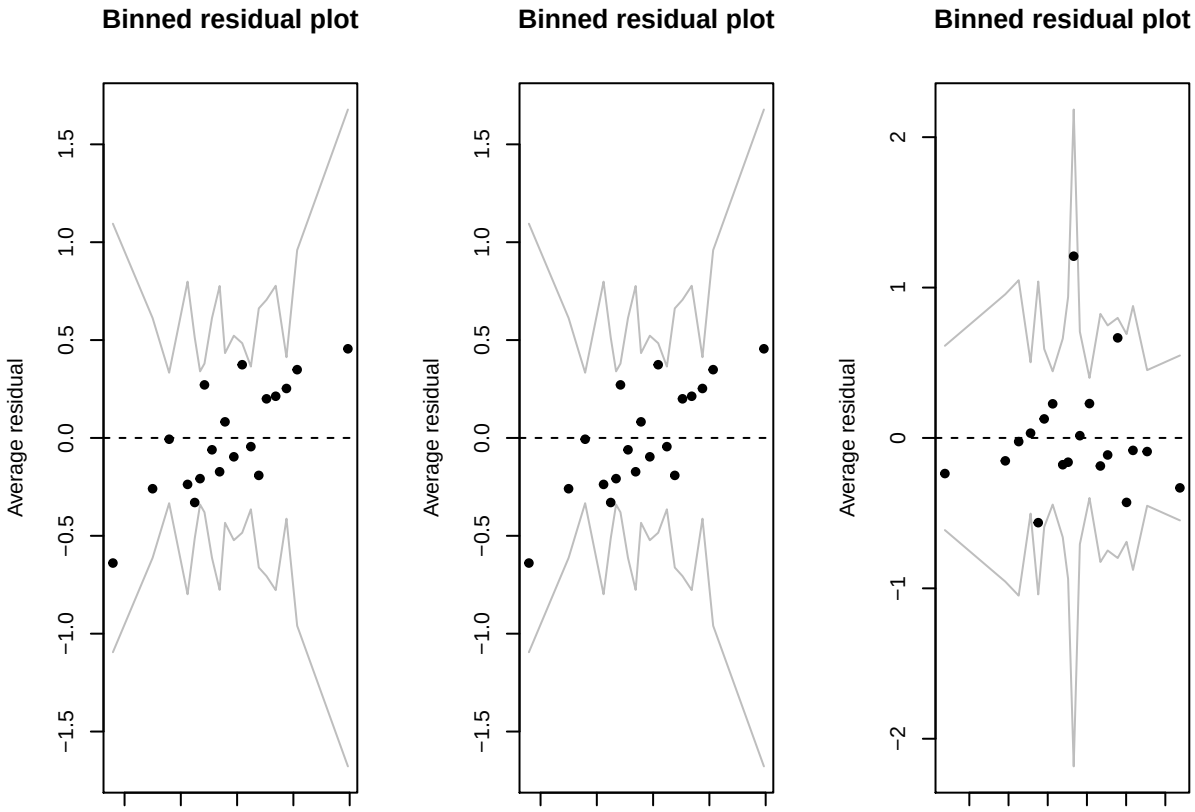

Diagnostic Plots
Cook's distanc
0.0
0.3
127
203
standardized resic
0
5
127
203
Bonferroni p−value
0.0
0.6
127
203
hat−values
0.01
0.04
0
100
200
300
400
Index




```
FALSE      rstudent unadjusted p-value Bonferonni p
FALSE 127 13.422602         6.3868e-34   2.4781e-31
FALSE 203 -4.177857         3.6500e-05   1.4162e-02
```


Diagnostic Plots
Cook's distanc
0.00
0.15
127
203
rstudent
0
5
127
203
p−vaudenitized resic
0.0
0.6
127
203
hat−values
0.01
0.04
75
158
0
100
200
300
400
Index


resid(fit1)[1:100]
4
2
0
−2
−4
Series resid(fit1)
ACF
1.0
0.8
0.6
0.4
0.2
0.0
Series resid(fit1)
Partial ACF
0.10
0.05
0.00
−0.05
−0.10

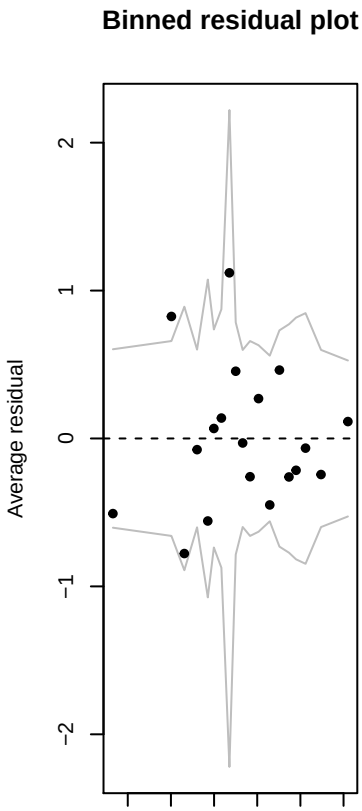



```
FALSE Data: Player_rush1
FALSE Models:
FALSE m2: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m2:     (1 | TEAM)
FALSE m1: YPA ~ Attempt + FPI + OFFENSE + Rush.Att + YPC + Opp.Ypc.Allowed +
FALSE m1:     (1 | TEAM)
FALSE m3: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m3:     (1 + Attempt | TEAM)
FALSE m4: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m4:     (1 | TEAM) + (1 + Attempt | TEAM)
FALSE    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
FALSE m2  9 1695.8 1731.4 -838.89   1677.8
FALSE m1  9 1703.2 1738.8 -842.60   1685.2  0.000      0          1
FALSE m3 11 1676.2 1719.8 -827.13   1654.2 30.955      2  1.898e-07 ***
FALSE m4 12 1678.2 1725.8 -827.13   1654.2  0.000      1          1
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
FALSE Data: Player_rush1
FALSE Models:
FALSE fit1: YPA ~ Attempt + FPI + OVERALL + YPC
FALSE m2: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m2:     (1 | TEAM)
FALSE m1: YPA ~ Attempt + FPI + OFFENSE + Rush.Att + YPC + Opp.Ypc.Allowed +
FALSE m1:     (1 | TEAM)
FALSE fit2: YPA ~ Opp.Att.Allowed + Opp.Yds.Allowed + Opp.Ypc.Allowed + Opp.Ypg.Allowed +
FALSE fit2:     Opp.FPI + Opp.Def.Eff + Opp.Overall.Eff
```

```
FALSE fit3: YPA ~ Attempt + Rush.Att + OFFENSE + YPC + OVERALL + FPI + YPC:Opp.Ypc.Allowed
FALSE m3: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m3:     (1 + Attempt | TEAM)
FALSE m4: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m4:     (1 | TEAM) + (1 + Attempt | TEAM)
FALSE      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
FALSE fit1  6 1648.6 1672.4 -818.32   1636.6
FALSE m2    9 1695.8 1731.4 -838.89   1677.8  0.000      3          1
FALSE m1    9 1703.2 1738.8 -842.60   1685.2  0.000      0          1
FALSE fit2  9 1692.2 1727.8 -837.09   1674.2 11.018      0     <2e-16 ***
FALSE fit3  9 1654.5 1690.1 -818.25   1636.5 37.693      0     <2e-16 ***
FALSE m3   11 1676.2 1719.8 -827.13   1654.2  0.000      2          1
FALSE m4   12 1678.2 1725.8 -827.13   1654.2  0.000      1          1
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

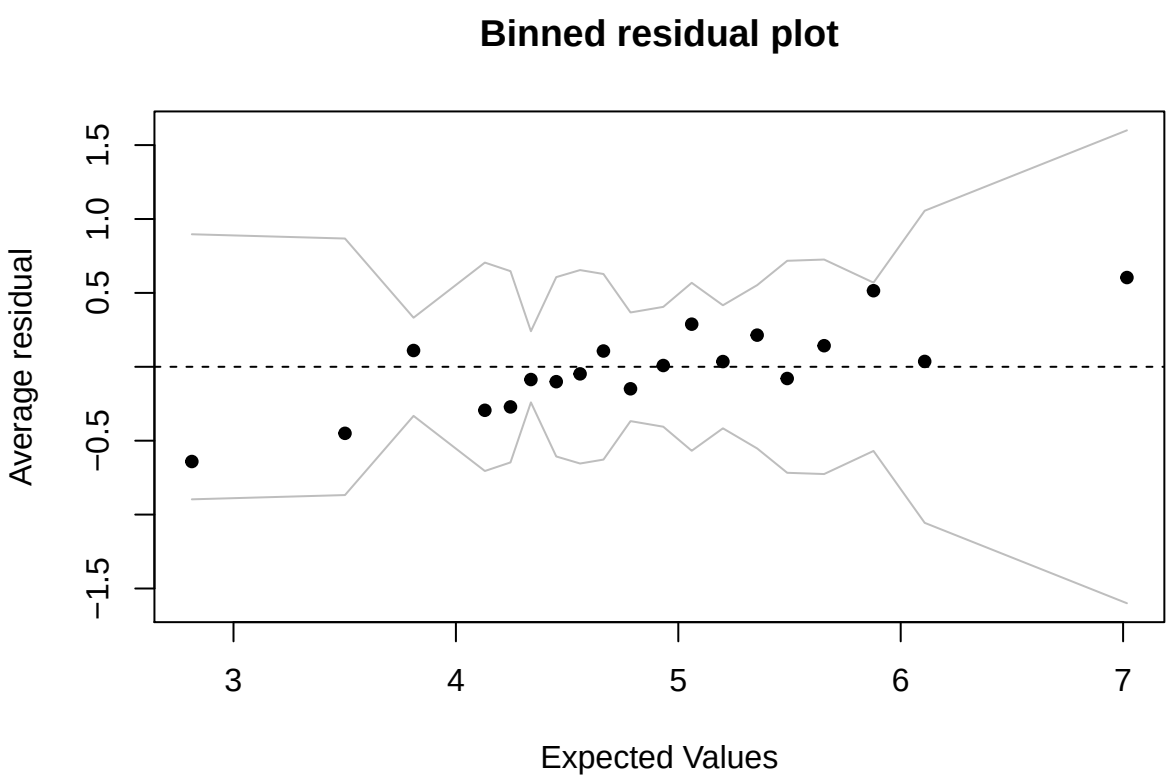
```
FALSE [1] 3.725749
```



Model Improvement

```
FALSE Linear mixed model fit by REML ['lmerMod']
FALSE Formula: YPA ~ Attempt + FPI + Rush.Att + YPC * Opp.Ypc.Allowed + (1 +
FALSE     Attempt | TEAM)
FALSE    Data: Player_rush1
FALSE
FALSE REML criterion at convergence: 1650.5
FALSE
```

```
FALSE Scaled residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -3.5077 -0.4350 -0.0212  0.3719  8.1835
FALSE
FALSE Random effects:
FALSE  Groups   Name        Variance  Std.Dev. Corr
FALSE  TEAM     (Intercept) 2.4205857 1.55582
FALSE           Attempt     0.0001185 0.01088  -1.00
FALSE  Residual             3.1026763 1.76144
FALSE Number of obs: 388, groups:  TEAM, 86
FALSE
FALSE Fixed effects:
FALSE                        Estimate Std. Error t value
FALSE (Intercept)          1.8407183 11.3969661   0.162
FALSE Attempt              0.0020392  0.0017410   1.171
FALSE FPI                 -0.0072119  0.0106598  -0.677
FALSE Rush.Att            -0.0002056  0.0014348  -0.143
FALSE YPC                  0.8620016  2.4481363   0.352
FALSE Opp.Ypc.Allowed     -0.0433409  2.6989256  -0.016
FALSE YPC:Opp.Ypc.Allowed -0.0341866  0.5784207  -0.059
FALSE
FALSE Correlation of Fixed Effects:
FALSE             (Intr) Attmpt FPI    Rsh.At YPC    Op.Y.A
FALSE Attempt     -0.010
FALSE FPI          0.005 -0.059
FALSE Rush.Att    -0.045 -0.050  0.104
FALSE YPC         -0.981  0.003 -0.057 -0.039
FALSE Opp.Ypc.All -0.998 -0.001  0.019  0.031  0.979
FALSE YPC:Opp.Y.A  0.984 -0.006  0.022 -0.014 -0.996 -0.985
FALSE convergence code: 0
FALSE Model failed to converge with max|grad| = 0.00224554 (tol = 0.002, component 1)
FALSE Model is nearly unidentifiable: very large eigenvalue
FALSE  - Rescale variables?

FALSE Data: Player_rush1
FALSE Models:
FALSE m5: YPA ~ Attempt + FPI + Rush.Att + YPC * Opp.Ypc.Allowed + (1 +
FALSE m5:     Attempt | TEAM)
FALSE m3: YPA ~ Attempt + FPI + YPC + OVERALL + Rush.Att + Opp.Ypc.Allowed +
FALSE m3:     (1 + Attempt | TEAM)
FALSE    Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
FALSE m5 11 1639.4 1683.0 -808.7   1617.4
FALSE m3 11 1636.6 1680.2 -807.3   1614.6 2.7959      0  < 2.2e-16 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




```
FALSE [1] 3.725311
```

I try to improve the model by considering interaction to the Model. However, the output shows that it doesn't fit the data better with interaction. Therefore, I think model 3 (m3) still should be chose in to compare the player's performance.

Discussion

The goal of this project is trying to determine factors that may have impact on player's yards gained per rush and who could be the better runner. After EDA and multilevel model building process, I found out that number of rush ofr individual(Attempt), Team average yards gain per carry, Team Power index(FPI), Opponent yards per carry allowed and Team average rushing attempts are factors that may affact the YPA for individual player. Attempts, FPI and YPC have positive effect on the YPA, other variables have negative effect. Although, the anova test shows that multilevel model doesn't fit my data well, I still think multilevel model should be use in my model choice. The reason is that different team will have thier own factors that may influence the player's performance as well as their opponent average power. Also, the AIC, BIC and deviance has no big difference with classic linear modelling.

Implitcation

From all the results and the interpretations of the models fitted above, it's not reasonable to use these models to predict player's performance; while the models do explain relationship between the YPA and other variables.

Limitation

1. The data are limited size. Although there are lots of variables and observations in college football, but my project are focusing on a small scope that constrain the data size.
2. Some variables are not real stats, such as FPI and Team Efficiency. All those data are from ESPN and based on their own math formula.

Future Direction

I nned do further research and study for the multilevel model building to improve the model from this project. It is possible to use multilevel model to predict player's performance for next season. However, it requires more knowlege that I'm not understdand well. I think this model could used to calculated "correct" performace on all players even though they are in different team and facing different opponent.