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. ##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.

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

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:
12) Rush.Yard
13) YPC 14) Opp.Att.Allowed 15) Opp.Yds.Allowed 16) Opp.Ypc.Allowed 17) Opp.Ypg.Allowed 18) Opp.FPI 19) Opp.Def.Eff 20) Opp.Overall.Eff

Model Used

EDA and Result

EDA

Model Choice

Interpretation

Model Checking

Discussion

Limitation

Future Direction

```
## -- Attaching packages ----------------------- tidyverse 1.2.1 --

## 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 -------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Warning: Column `TEAM` joining factors with different levels, coercing to
## character vector

## Warning: Column `TEAM` joining character vector and factor, coercing into
## character vector

## Warning: Column `TEAM` joining character vector and factor, coercing into
## character vector

## Warning: Column `TEAM` joining character vector and factor, coercing into
```

```
## character vector
```

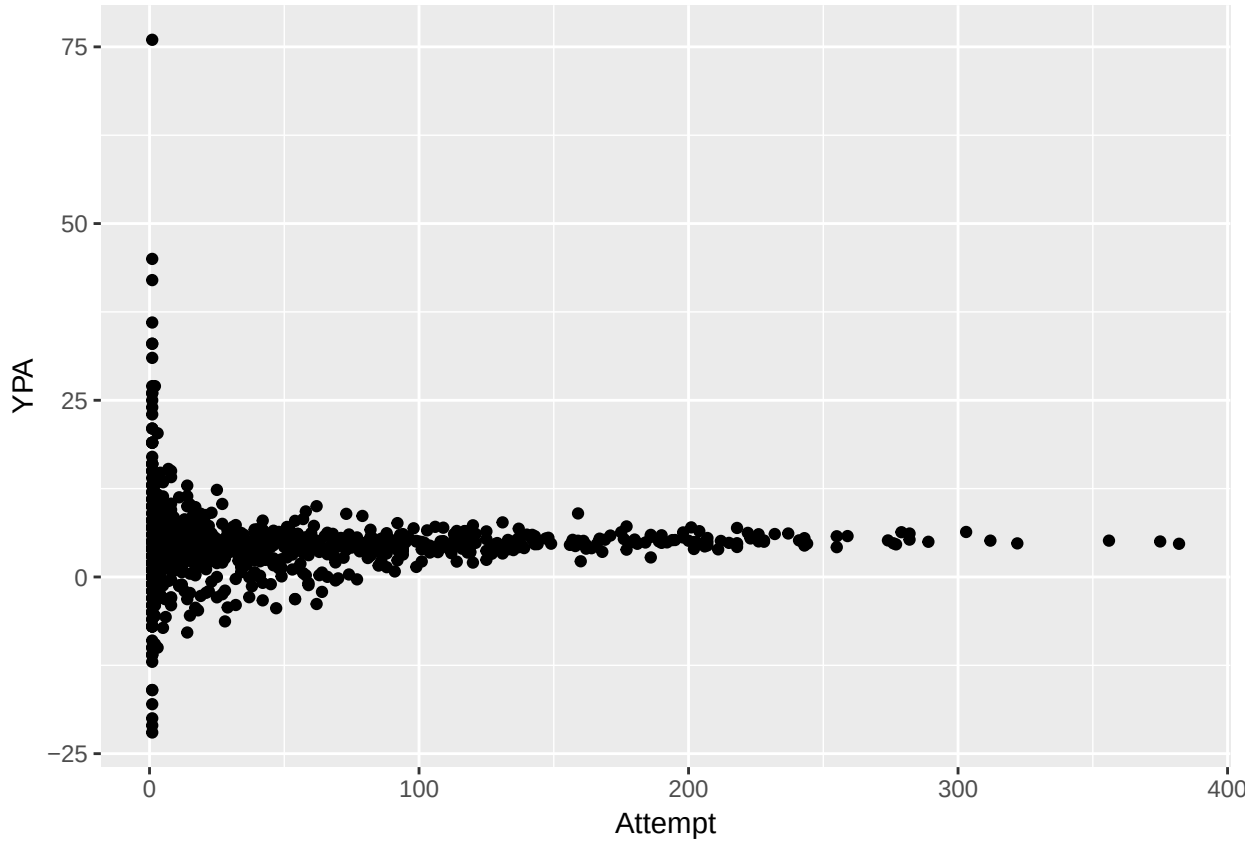
```
## Warning: Column `TEAM` joining factors with different levels, coercing to
## character vector
```

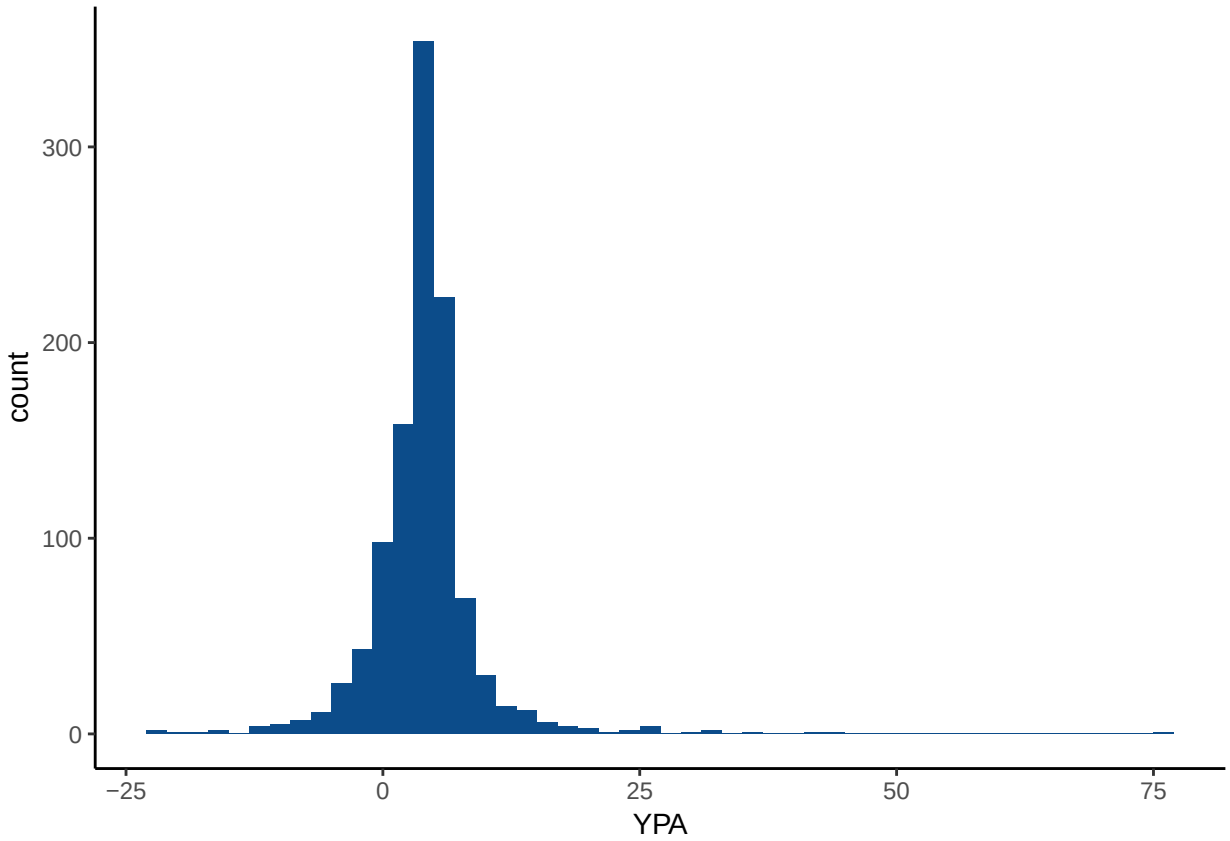
```
## Warning: Column `TEAM` joining character vector and factor, coercing into
## character vector
```

```
## Warning: Column `TEAM` joining character vector and factor, coercing into
## character vector
```

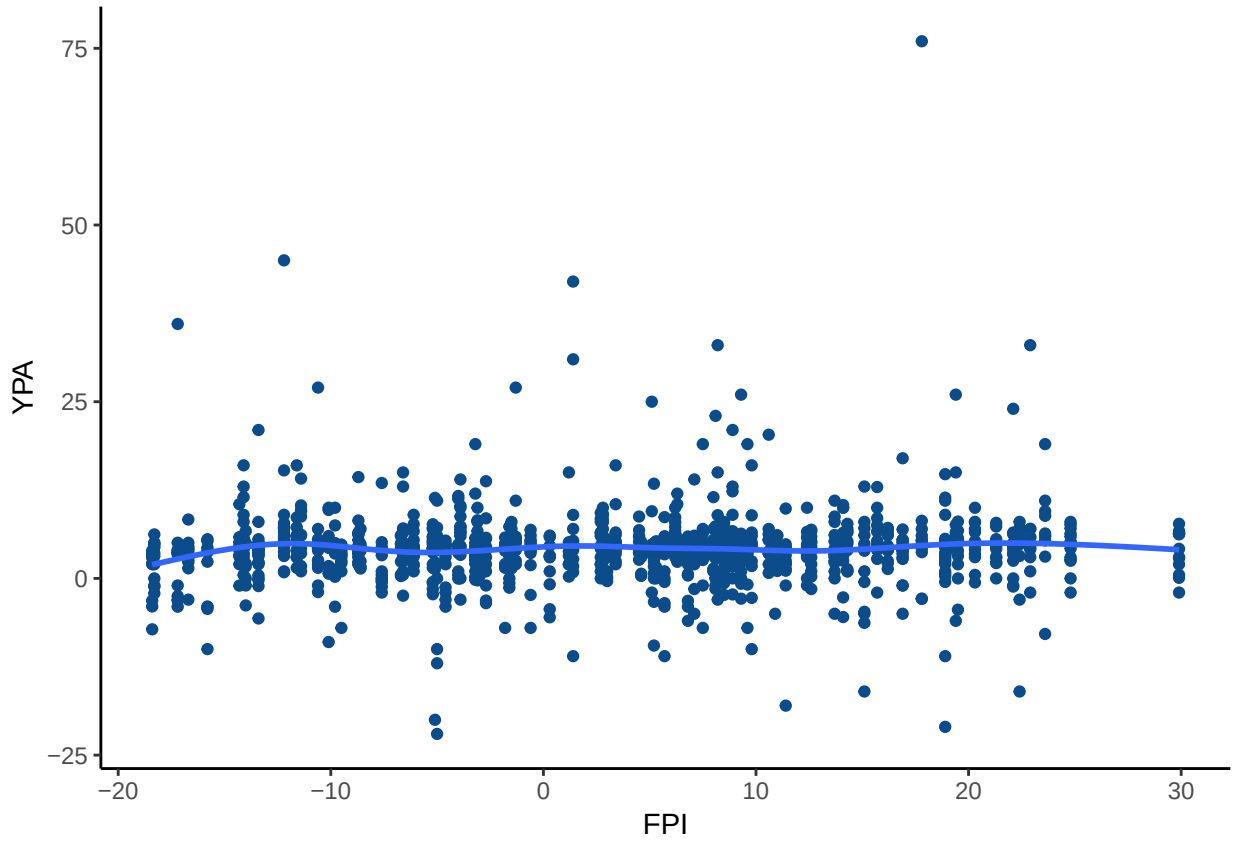
```
## Warning: Column `TEAM` joining character vector and factor, coercing into
## character vector
```

```
## Warning: Column `TEAM` joining factor and character vector, coercing into
## character vector
```

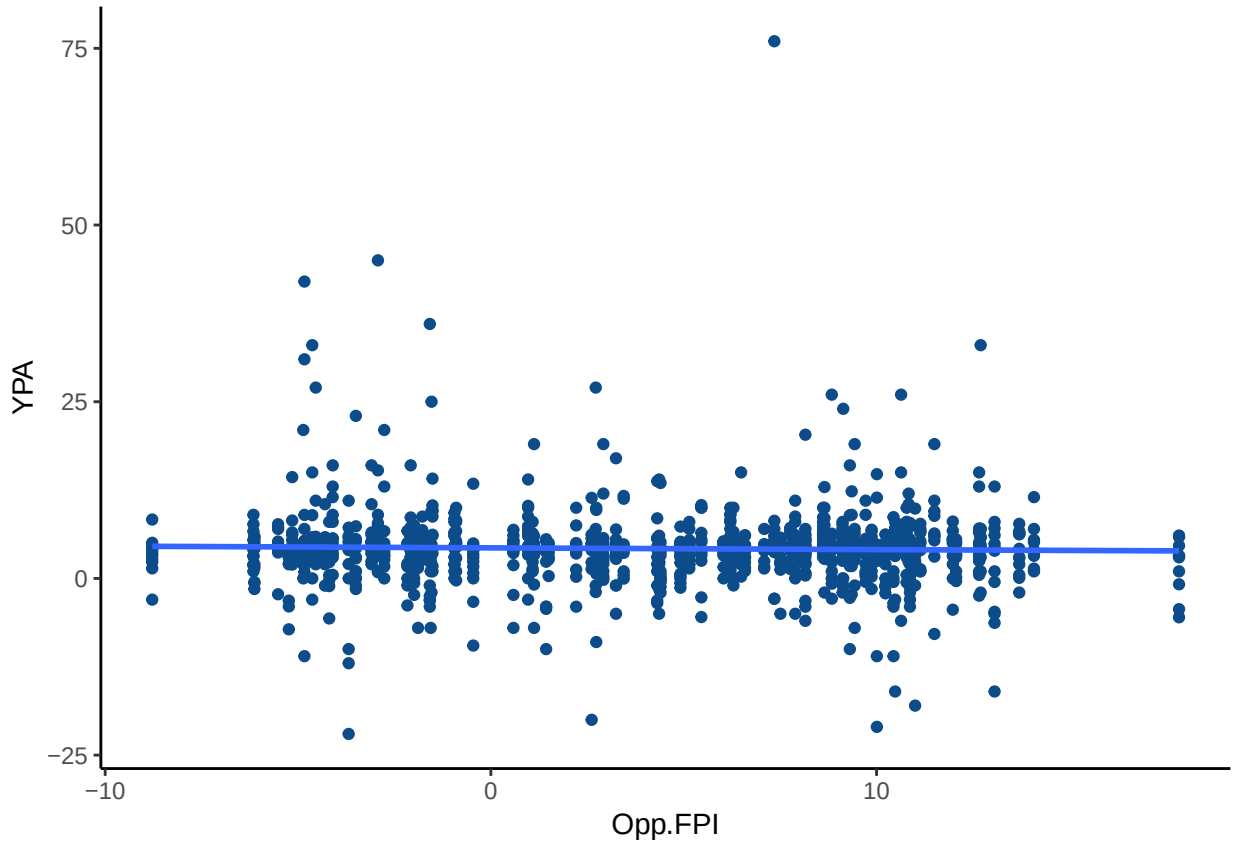




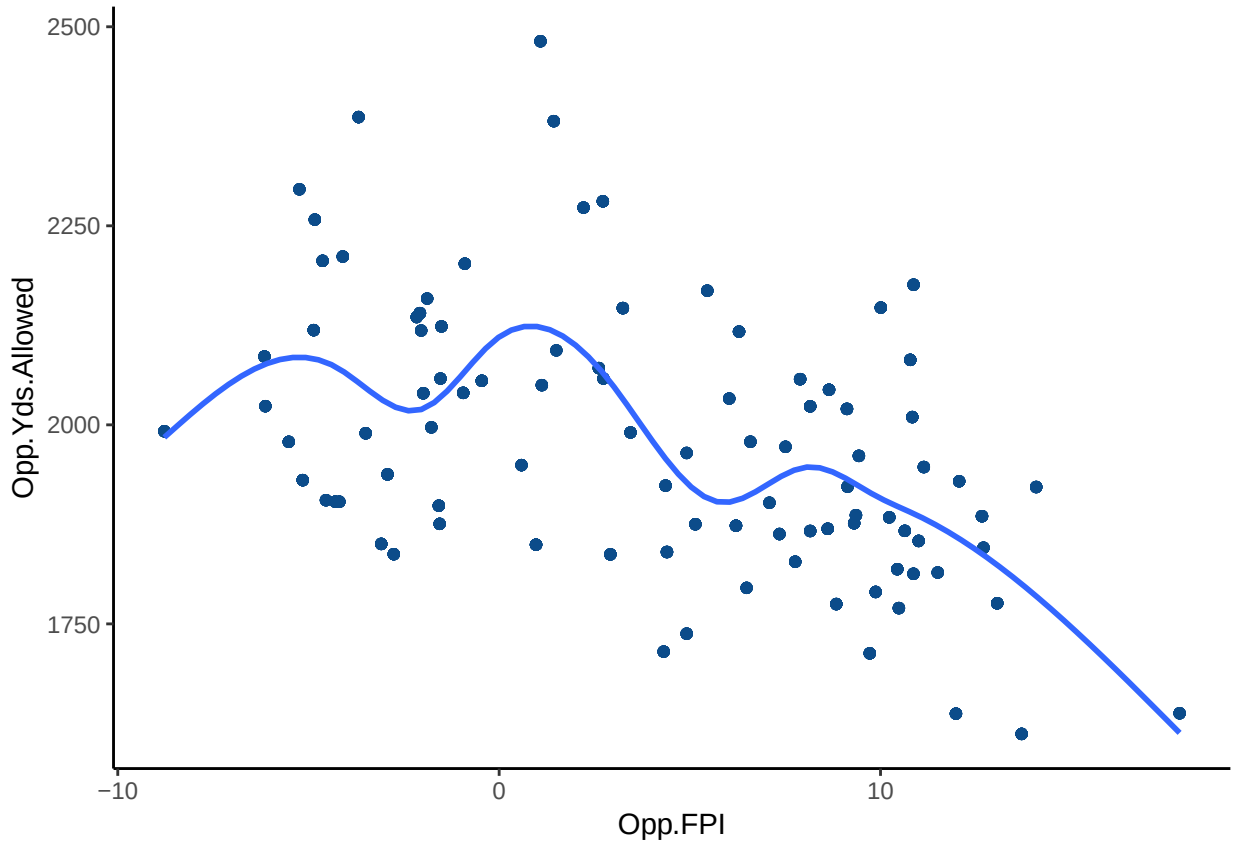

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
##
## Attaching package: 'nlme'
## The following object is masked from 'package:dplyr':
##
##     collapse
## Linear mixed-effects model fit by REML
##  Data: Player_rush
##        AIC      BIC    logLik
##   6852.178 6892.063 -3418.089
##
## Random effects:
##  Formula: ~1 | TEAM
##          (Intercept) Residual
## StdDev: 0.0004811699 5.568419
##
## Fixed effects: YPA ~ OFFENSE + Rush.Att + YPC + Opp.Ypc.Allowed + Opp.FPI
##                       Value Std.Error  DF   t-value p-value
## (Intercept)      2.9057005  3.379082 994  0.859908  0.3900
## OFFENSE         -0.0097366  0.012393  87 -0.785638  0.4342
## Rush.Att        -0.0008108  0.002120  87 -0.382414  0.7031
## YPC              1.4819989  0.340818  87  4.348352  0.0000
## Opp.Ypc.Allowed -0.9640452  0.774982  87 -1.243959  0.2169
## Opp.FPI         -0.0534901  0.032019  87 -1.670575  0.0984
##  Correlation:
##                 (Intr) OFFENS Rsh.At YPC    Op.Y.A
```

```
## OFFENSE          0.026
## Rush.Att        -0.215  0.029
## YPC             -0.002 -0.544 -0.555
## Opp.Ypc.Allowed -0.952  0.035  0.134 -0.170
## Opp.FPI         -0.317 -0.415  0.129  0.099  0.281
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -4.7400945 -0.3554139  0.0107298  0.2631192 12.7303688
##
## Number of Observations: 1087
## Number of Groups: 93
```

```
## 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
```