NFL Draft

Load Libraries

```
options(warn=-1)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(repr)
```

Read the Data

```
draft <- read.csv('./NFL Draft.csv')
head(draft)
```

```
##   Player_Id Year Rnd Pick  Tm           Player Pos Position.Standard
## 1  WinsJa00 2015   1    1 TAM   Jameis Winston  QB                QB
## 2  MariMa01 2015   1    2 TEN   Marcus Mariota  QB                QB
## 3  FowlDa00 2015   1    3 JAX     Dante Fowler OLB                LB
## 4  CoopAm00 2015   1    4 OAK     Amari Cooper  WR                WR
## 5  ScheBr00 2015   1    5 WAS  Brandon Scherff   T                 T
## 6  WillLe02 2015   1    6 NYJ Leonard Williams  DE                DE
##   First4AV Age   To AP1 PB St CarAV DrAV  G Cmp Pass_Att Pass_Yds
Pass_TD
## 1       13  21 2016   0  1  2    13   13 26 540      913     6722
42
## 2        9  21 2016   0  0  2     9    9 23 458      725     5590
42
## 3        0  21 2016   0  0  0     0    0 10  NA       NA       NA
NA
## 4        9  21 2016   0  1  1     9    9 26  NA       NA       NA
NA
## 5        7  23 2016   0  0  2     7    7 27  NA       NA       NA
NA
## 6        9  21 2016   0  0  2     9    9 26  NA       NA       NA
NA
##   Pass_Int Rush_Att Rush_Yds Rush_TDs Rec Rec_Yds Rec_Tds Tkl
Def_Int  Sk
```

```
## 1       25       86      311        7  NA      NA      NA  NA
NA
## 2       18       79      516        4   1      41       1  NA
NA
## 3       NA       NA       NA       NA  NA      NA      NA  16
NA 2.5
## 4       NA        3       -3        0 134    1970       9  NA
NA
## 5       NA       NA       NA       NA  NA      NA      NA  NA
NA
## 6       NA       NA       NA       NA  NA      NA      NA  56
NA   9
##   College.Univ X
## 1  Florida St.
## 2       Oregon
## 3      Florida
## 4      Alabama
## 5         Iowa
## 6          USC
```

```
str(draft)
```

```
## 'data.frame':    8435 obs. of  33 variables:
##  $ Player_Id        : Factor w/ 7215 levels "","  Thigpen",..: 7061
3955 2030 1315 5547 6948 6821 399 1978 2467 ...
##  $ Year             : int  2015 2015 2015 2015 2015 2015 2015 2015
2015 2015 ...
##  $ Rnd              : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Pick             : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Tm               : Factor w/ 35 levels "ARI","ATL","BAL",..: 33
34 15 23 35 22 6 2 21 32 ...
##  $ Player           : Factor w/ 8230 levels "'Omar Ellison",..: 3391
5159 1855 230 814 4977 4664 7985 2655 7558 ...
##  $ Pos              : Factor w/ 25 levels "C","CB","DB",..: 19 19 17
25 23 4 25 17 23 20 ...
##  $ Position.Standard: Factor w/ 15 levels "C","DB","DE",..: 11 11 8
15 13 3 15 8 13 12 ...
##  $ First4AV         : int  13 9 0 9 7 9 0 7 9 8 ...
##  $ Age              : int  21 21 21 21 23 21 23 23 21 21 ...
##  $ To               : int  2016 2016 2016 2016 2016 2016 2016 2016
2016 2016 ...
##  $ AP1              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PB               : int  1 0 0 1 0 0 0 0 0 1 ...
##  $ St               : int  2 2 0 1 2 2 0 1 2 2 ...
##  $ CarAV            : int  13 9 0 9 7 9 0 7 9 8 ...
##  $ DrAV             : int  13 9 0 9 7 9 0 7 9 8 ...
##  $ G                : int  26 23 10 26 27 26 4 26 25 23 ...
##  $ Cmp              : int  540 458 NA NA NA NA NA NA NA NA ...
##  $ Pass_Att         : int  913 725 NA NA NA NA NA NA NA NA ...
##  $ Pass_Yds         : int  6722 5590 NA NA NA NA NA NA NA NA ...
```

```
##  $ Pass_TD          : int  42 42 NA NA NA NA NA NA NA NA ...
##  $ Pass_Int         : int  25 18 NA NA NA NA NA NA NA NA ...
##  $ Rush_Att         : int  86 79 NA 3 NA NA 1 NA NA 416 ...
##  $ Rush_Yds         : int  311 516 NA -3 NA NA 9 NA NA 1697 ...
##  $ Rush_TDs         : int  7 4 NA 0 NA NA 0 NA NA 14 ...
##  $ Rec              : int  NA 1 NA 134 NA NA 19 NA 1 47 ...
##  $ Rec_Yds          : int  NA 41 NA 1970 NA NA 187 NA -11 390 ...
##  $ Rec_Tds          : int  NA 1 NA 9 NA NA 0 NA 0 0 ...
##  $ Tkl              : int  NA NA 16 NA NA 56 NA 44 NA NA ...
##  $ Def_Int          : num  NA NA NA NA NA NA NA 1 NA NA ...
##  $ Sk               : Factor w/ 443 levels "","0.5","1","1.5",..: 1
1 34 1 1 142 1 15 1 1 ...
##  $ College.Univ     : Factor w/ 318 levels "","Abilene
Christian",..: 85 207 81 6 114 278 296 57 151 91 ...
##  $ X                : Factor w/ 2 levels "","College Stats": 1 1 1 1
1 1 1 1 1 1 ...
```

Checking for the NA Values

```
any(is.na(draft))
```

```
## [1] TRUE
```

Creating the Linear Regression

```
draft <- draft %>% select(Pick, DrAV) %>% filter(Pick<257) %>%
na.omit()
fit.1 <- lm(DrAV ~ Pick, data=draft)
fit.2 <- lm(DrAV ~ poly(Pick,2), data=draft)
fit.3 <- lm(DrAV ~ poly(Pick,3), data=draft)
fit.4 <- lm(DrAV ~ poly(Pick,4), data=draft)
fit.5 <- lm(DrAV ~ poly(Pick,5), data=draft)
fit.6 <- lm(DrAV ~ poly(Pick,6), data=draft)
fit.7 <- lm(DrAV ~ poly(Pick,7), data=draft)

anova(fit.1, fit.2, fit.3, fit.4, fit.5, fit.6, fit.7)
```

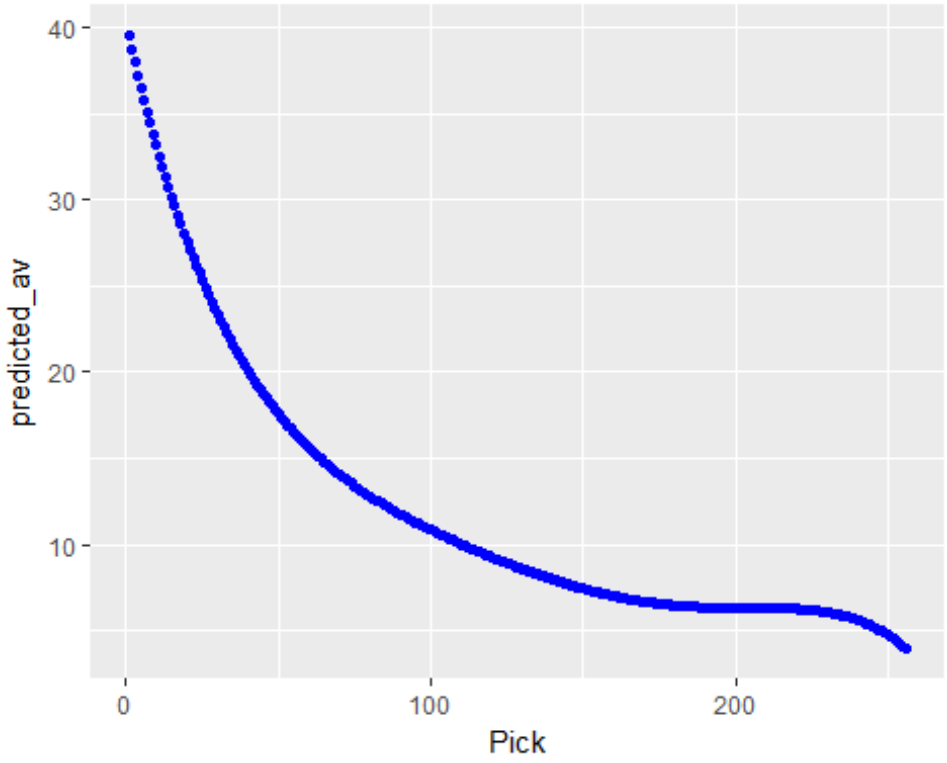
```
## Analysis of Variance Table
##
## Model 1: DrAV ~ Pick
## Model 2: DrAV ~ poly(Pick, 2)
## Model 3: DrAV ~ poly(Pick, 3)
## Model 4: DrAV ~ poly(Pick, 4)
## Model 5: DrAV ~ poly(Pick, 5)
## Model 6: DrAV ~ poly(Pick, 6)
## Model 7: DrAV ~ poly(Pick, 7)
##   Res.Df     RSS Df Sum of Sq        F    Pr(>F)
## 1   6066 1792400
## 2   6065 1703228  1     89172 321.3354 < 2.2e-16 ***
```

```
## 3   6064 1686026  1     17202  61.9862 4.069e-15 ***
## 4   6063 1683502  1      2525   9.0975   0.00257 **
## 5   6062 1682316  1      1186   4.2737   0.03875 *
## 6   6061 1681686  1       630   2.2691   0.13203
## 7   6060 1681685  1         1   0.0046   0.94571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Created a linear model to fit all seven rounds of the NFL Draft for predictions with the modern draft.

Draft Pick Exploratory Analysis

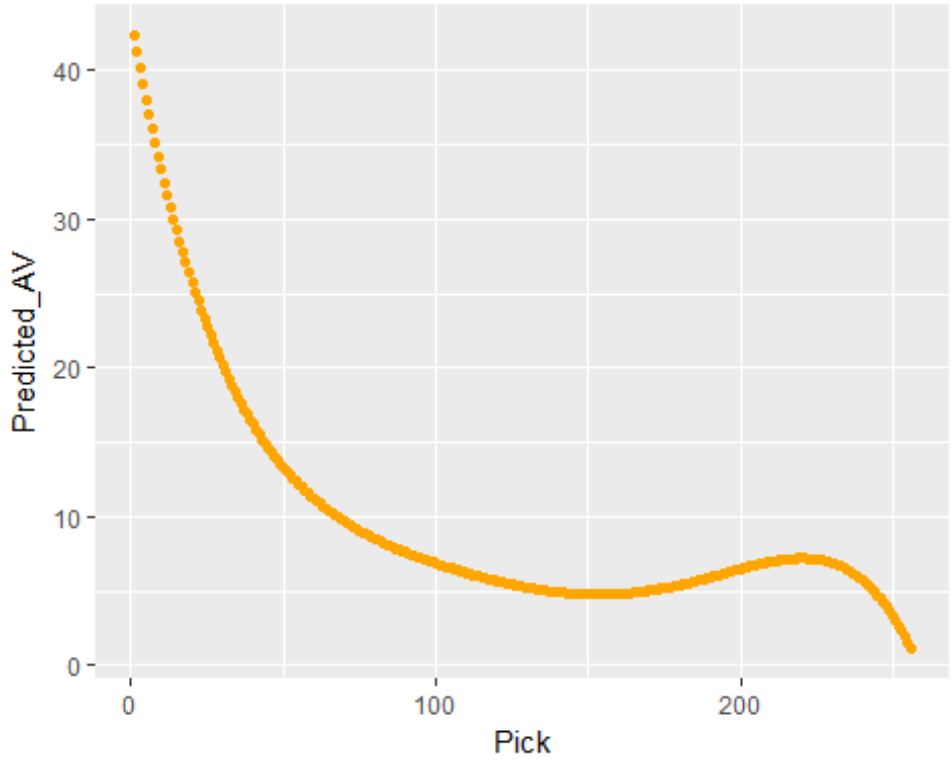
```
draft$y_hat <- predict(fit.5)
group_by_pick <- draft %>% group_by(Pick) %>% summarise(predicted_av =
mean(y_hat)) %>% data.frame()
options(repr.plot.width=4, repr.plot.height=3)
ggplot(group_by_pick, aes(Pick, predicted_av)) +
geom_point(color='blue')
```



Quarterback Draft Analysis

```
qb <- read.csv("./Nfl Draft.csv")
qb <- filter(qb, Position.Standard=="QB")
qb <- qb %>%
```

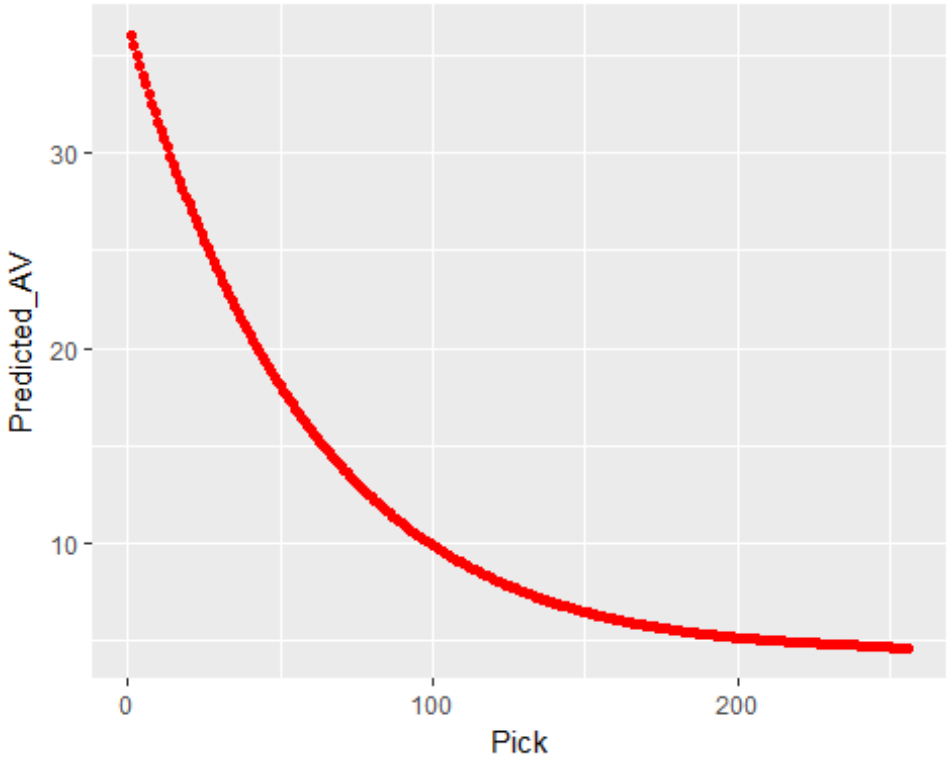
```
  select(Pick, DrAV) %>%
  filter(Pick<257) %>%
  na.omit()
fit.5 <- lm(DrAV ~ poly(Pick,5), data=qb)
new <- data.frame(Pick = seq_len(256))
y_hat <- predict(fit.5, new, se.fit = TRUE)
df <- data.frame(y_hat = matrix(unlist(y_hat)))
Predicted_AV <- df[seq(1,256),]
df <- data.frame(Pick = new$Pick, Predicted_AV)
ggplot(df, aes(Pick, Predicted_AV)) + geom_point(color='orange')
```



Running Backs in Draft

```
rb <- read.csv("./Nfl Draft.csv")
rb <- filter(rb, Position.Standard=="RB")
rb <- rb %>%
  select(Pick, DrAV) %>%
  filter(Pick<257) %>%
  na.omit()
fit.5 <- lm(DrAV ~ poly(Pick,5), data=rb)
new <- data.frame(Pick = seq_len(256))
y_hat <- predict(fit.5, new, se.fit = TRUE)
df <- data.frame(y_hat = matrix(unlist(y_hat)))
Predicted_AV <- df[seq(1,256),]
```

```
df <- data.frame(Pick = new$Pick, Predicted_AV)
ggplot(df, aes(Pick, Predicted_AV)) + geom_point(color='red')
```




```
head(df)
```

```
##   Pick Predicted_AV
## 1    1     36.06027
## 2    2     35.53884
## 3    3     35.02523
## 4    4     34.51934
## 5    5     34.02107
## 6    6     33.53032
```