

College-to-Pro QB Model

Sam Burch

2024-00-00

Introduction

After defining quarterback (QB) success in my previous article, it is now time to project QB play for the top prospects heading into the 2024 NFL draft. Last year, I developed a linear regression model to predict NFL success (defined as a 50-50 composite of EPA/play and average PFF grade). After optimizing with the greedy algorithm and manually *eliminating* collinearity issues, my model had an adjusted R^2 of 0.55. While this was a step in the right direction, there were several issues, mainly overfitting biases, that I will be correcting in this analysis.

Data Processing

To start, the dataset had a career minimum of 650 plays and 250 for each season. This restricted the sample size to only 29 QBs. Therefore, this dataset (same as for the previous article) uses 400 min. career plays and 150 min. for each season. This allows us to increase the sample size to 47 and include players who didn't have as many plays due to them being bad. I also reduced the number of college single season plays from 200 to 150. (Keep in mind, the sample size is so small because college data is from 2014-2023. Hence, players like Cam Newton, Andrew Luck, or even Tom Brady won't be studied here. This is perhaps the biggest issue that still hasn't been solved.)

After these restrictions were set, I needed college stats that could be used as potential predictors. The PFF datasets I have access to provide information like big-time-throws, pressure-to-sack-rate, and grades. Not many tweaks were made to these individual stats, however one big change was considering different samples. Not only will I include the career play of these QBs, but I will also include first, last, worst, and best college season summaries. On top of this, the cfbfastR data I used last year only considers a handful of stats. This year, I'm expanding this list to include total EPA, sack plays, and even ground plays (sack or rush). I'm also considering career, first, last, worst, and best season splits.

One key predictor to NFL QB play is draft position. Draft pick has a -0.35 correlation with NFL success (defined as value per game metric, min. 10 games) since 1980. Thus, I will add in pick, round, and age from the nflfastR draft picks package.

Other potential predictors can include combine measures, but they were not considered in this analysis. Tej Seth of Sumer Sports wrote about recent correlations with NFL success *here* (<https://sumersports.com/the-zone/nfl-combine-what-influences-production-and-draft-position/>).

Correlations

This dataset has 193 potential predictors that it is considering. So, let us look at those whose absolute correlation is 0.25 or larger.

Right away, we can see ground EPA is very important. Variants of ground EPA make up the top-3 and five of the top-6. In fact, the top-2 have a correlation above 0.40, which is great for football data. Pressure-to-sack-rate (PTSR) is seventh highest at -0.35, meaning a the lower the better. This all points to pocket awareness – maximizing sack avoidance and scrambling – translates to the NFL. (Keep in mind, something like EPA per pressure play is more likely to determine how a QB plays under pressure; this is due to pressures that result in passes. However, I do not have access to pressure play-by-play data.) Average depth of target

(ADOT) in their first year shows up surprisingly well at 0.37. This likely points to the importance of arm strength in the NFL. Total EPA and EPA per play have appearances high, which makes sense as these are based on overall production. Lastly, other total statistics – like TDs, completions, and yards – signal the importance of experience and productivity.

Draft round and pick were relatively low, at -0.27 and -0.25 respectively, when we know these should be high. If we look at the top correlations with pick, this may provide some traits our models may be missing on.

There are a lot of high correlations here, as nearly 40 other predictors have an absolute correlation above 0.40 with draft pick. The highest, above 0.50, include the following: average pass EPA in best year, total EPA in best year, EPA per play in best year, EPA per rush, and EPA per pass. Across the board, these statistics are production based whether that be through film (PFF grading) or analytics (EPA). This of course makes sense and likely will be more helpful statistics than just looking at sack avoidance. While that may ring true, it is possible NFL teams are underrating the importance of pocket awareness, experience, and maybe even arm strength as these have higher importance to NFL success in this dataset.

Data Mining

Let's run through some interesting graphs before we begin modeling. For these, I will look at the 47 NFL QBs in the dataset along with the top-7 2024 draft prospects. In my last article, the 67th percentile was defined as the average starter, so we will continue to use this definition here.

Looking at career production, Williams is atop, with Maye being second in grading and McCarthy being second in efficiency. While good production in college doesn't necessarily translate to NFL production – look at the mix of blue and red dots in the top right and bottom left – we know that production is important for draft position. Thus, with draft position being important historically, we know college production should lead to better prospects in general.

Looking at best season production for efficiency and grading, Daniels is by far the best of the 24' prospects. Nix is second, followed by a cluster including Williams, Penix, and McCarthy. Maye & Rattler had elite grading in their best seasons, but efficiency was below average here. Regardless, all of these prospects have had good best seasons, which should be a good sign.

Here, we look at where on the field a QBs efficiency is coming from. The top right (good passing, rushing, and sack avoidance) is where one wants to be. Of the NFL QBs, only three are below average and six are above average. All of the 24' prospects, except Rattler, are in this quadrant. Also, having good ground efficiency gives QBs a solid floor, since very few above average starters had below average ground efficiency and 8 of the top 9 were above average starters. Nix and Daniels show up the best in ground efficiency giving them a solid floor. Keep in mind though, some of the *successful* QBs with great ground EPA include Hill and Mariota.

The top couple predictors are all dealing with ground efficiency. However, ADOT in a QBs first year is very high as well. As we compare the two, we can see there have only been three failures in the top right quadrant, compared to nine successes. Williams, McCarthy, Maye, and Penix are all in this area. Nix and Daniels don't have a high ADOT in their first year, however they have the second and third best max total ground EPA respectively. Again, this should lead to a high floor and is a good sign regardless of their ADOT.

Just looking at how QBs deal with pressure in college can tell us a lot. Penix has the best PTSR in this entire dataset. Nix also doesn't take many sacks and has the added benefit of his rushing. While Daniels has elite rushing too, he takes a lot of sacks. McCarthy has good (not great) sack avoidance, while Williams and Maye are on the border between good and bad. Rattler's below average in both, which is concerning.

Before moving onto modeling, let's see how draft position tends to NFL success. 15 of the 25 QBs drafted in round 1 become above average NFL starters. Meanwhile, only 4 of the 22 non-first rounders become above average starters. Hurts and Prescott were genuinely non-first round hits. Purdy still has something to prove after only playing with an elite offense around him. Hill is just efficient on a small sample. These players skew the correlations in recent history, along with first-round busts. However, they will be kept in the sample as we move onto modeling because of the already small sample of QBs in the dataset. Regardless,

if NFL teams don't value a QB as a first rounder, this is a bad sign. This will be important when it comes to players like Nix and Penix, as they are fringe first rounders.

Modeling

For our train-test split – not considered last year due to a small sample size – we will use 85-15. The training set needs to have enough data points, otherwise the results will not be trustworthy. However, our testing results cannot be fully trusted here. While they can provide some small signal, it will likely be very noisy due to a small sample. Thus, we must keep this in mind when doing model selection.

To get an idea of what a good rmse and r^2 might look like, when we use the average (0.547) to predict QB success. We get an rmse of 0.348. This means, on average, we are off by ~35 on a scale from 0-100 (aka percentile).

Linear Regression

For the linear regression we can't fit a model with all the potential predictors, because of dimensional issues. Therefore, we start by fitting one with predictors where their absolute correlations are 0.275 or more. Then, by optimizing using the greedy algorithm this gives us the following: an adjusted r^2 of 0.466 and a RMSE of 0.232. The adjusted r^2 is very good and the error is not bad, but collinearity is a problem. Hence, it is not smart to consider this model in the model selection later now.

This model considers predictors with an absolute correlation above 0.30. When optimizing, again with the greedy algorithm, the adjusted r^2 is 0.343 and the error is 0.260. These numbers are both good. The diagnostics pass here, so this is an improvement over the last model. The predictors selected here are average ground EPA, total EPA, and PTSR.

Using the maximum total ground EPA as the only predictor, we get an RMSE of 0.34. The adjusted r^2 is 0.19, which isn't awful. If we instead use PTSR as the only predictor, the r^2 drops to 0.05, but the error improves to 0.31. When using draft pick, both r^2 and RMSE are bad – 0.01 and 0.33 respectively. Considering these factors, I will try to construct a model which has a high adjusted r^2 , low RMSE, and low collinearity error. This will not be perfect, and is not recommended as it will introduce bias, although it will be an interesting exercise.

Again, this process is not ideal, but I looked at the top correlations and did trial and error to optimize r^2 and the prediction error. I ended up with PTSR, average ground EPA, ADOT in first year, and total EPA in best year as the predictors. (Total EPA in best year had a high correlation with round and a better trade off with the comparison metrics.) This model has a adjusted r^2 of 0.321, rmse of 0.240, and low collinearity. If I run the greedy algorithm on this model, it suggests to drop total EPA in best year. This would improve the r^2 to 0.326, but the rmse would worsen to 0.251. Thus, we will keep this predictor considering the better rmse and the tell of overall play, on top of pocket awareness and arm strength.

Lasso

We will next consider a lasso regression, as not all of these predictors should be included due to their massive collinearity.

After completing Lasso, this tells us that no predictors are worth keeping. The intercept (0.53) is what the model will predict each time. By doing so, the error is 0.353; this is worse than our average model.

KNN

The prediction error for KNN was optimized at $K = 22$. This led to an error of 0.343, again not great.

Random Forest

After tuning on mtry and nodesize, the selected model has an r^2 of 0.031, with an OOB rate of 0.082. The rmse is bad too at 0.386. So, while this model won't be selected, the top variables for this model include EPA totals, ADOT, and completion percentage (CP).

XGBoost

The best prediction error for XGBoost also does not perform with an error of 0.366. The optimal tuning here was 0.1 ETA, 5 max depth, and 2 ntrees.

Model Selection

Across the board, most models perform very similarly. Looking at adjusted r^2 , the first linear regression (which included all predictors with an abs. correlation of 0.275 or higher) is by far the best at 0.466. However, there are certainly overfitting issues as the model does not pass the collinearity diagnostics. Next tier here is the second and third linear regression respectively. These have good r^2 s, but not great at around a third.

For prediction error, the baseline is 0.348 for the average. Five models did perform better than this. The best here was the first linear regression, however note the collinearity issues. Next comes the other linear regression models, with this time the third performing slightly better. With the testing sample size being so small, this rmse statistic should be taken with a grain of salt though.

Therefore, we will select the second linear regression model, with hesitancy. This model has the second highest adjusted r^2 and the third lowest prediction error. The first LR model would have been chosen, but it did not pass the collinearity diagnostics. Meanwhile, the third linear regression model does have a better rmse, however the more stable stat (r^2) is worse than the second LR. On top of this, bias was introduced when modeling the third linear regression. The runners up are the average, the best predictor, and the KNN model. The best predictor and KNN model each had a slightly better prediction error than the average. Although that is the case, the errors are all close enough that the average is probably the best of the runner ups. The advanced models all didn't perform well. This can possibly be improved by transforming the predictors into principal components. However, it is not necessary as the advanced models all pass the diagnostics by tuning them.

Projections

Let's see how the selected model projects the 2024 NFL draft prospects!

Nix and Penix are in their own tier due to their elite pocket awareness and experience. The issue with both is their consensus rankings have them as fringe first rounders. Considering this, if they fall out of the first round, they shouldn't necessarily be looked at as values. The top-4 on the consensus board are next at slightly above average. McCarthy is hurt by his limited number of plays, same with Williams and Maye, but Daniels is helped by it. Take on the fact that all have had elite seasons makes them good prospects. The biggest issue is Daniels needs a good surrounding as his PTSR is concerning low. Daniels' experience and rushing should help him though. Williams and Maye weren't elite in pocket awareness in college, but their elite production make them elite prospects regardless. McCarthy was great across the board, but it was on a small sample, as stated earlier. On the other hand, Rattler projects poorly being worst in every category but PTSR.

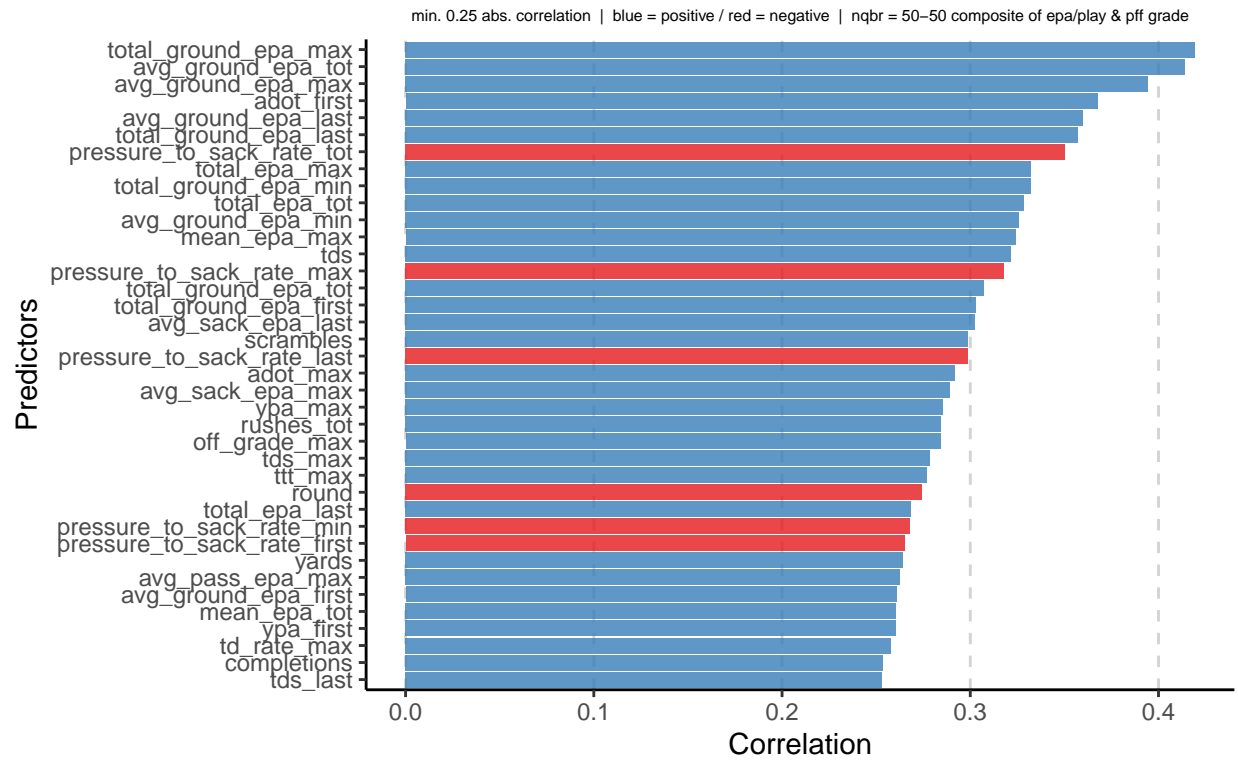
Some late round QBs are interesting. Jordan Travis has the third highest projection (77th percentile) out of the 11 prospects considered, due to his experience and good (not great) pocket awareness. The other three (Hartman, Pratt, and Milton) all project as below average starter or backup. Hartman and Pratt are helped by their production & experience, but hurt by their pocket awareness, which is concerning. Milton might be worth a flyer with his strong arm and decent pocket awareness.

Conclusion

Next steps include using a larger sample, implementing principal components into the models, and adding even more predictors. The first is by far the most important, as having a larger sample size will almost certainly create a better model than using the average. Implementing PCs will help the advanced models perform better. More predictors, like performance in a clean pocket, on play action plays, or percentage of pressures assigned to the QB will likely have some significance. Regardless, this analysis is a big step forward from last year's and provides us with a more concrete projection of the QB prospects. One continued theme is the stability of pocket awareness. Metrics like PTSR and ground EPA need to be considered more when ranking prospects.

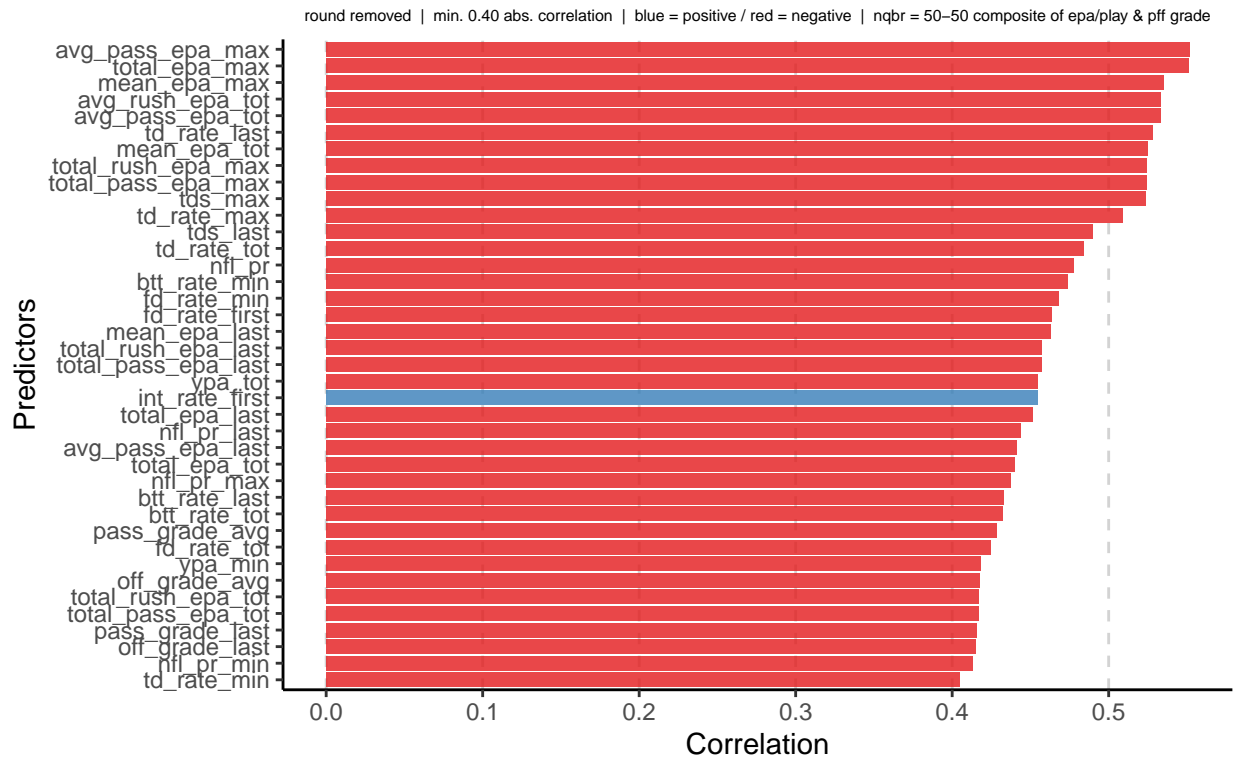
Thanks for reading!

Absolute Correlation with NFL nQBR



By: Sam Burch | Data: nflfastR, cfbfastR, & pff (2014–2023)

Absolute Correlation with Draft Pick



College Production Career

```
df_final |> filter(
  # player == "Anthony Richardson" |
  # player == "Will Levis" |

  player == "Caleb Williams" |
  player == "Drake Maye" |
  player == "Jayden Daniels" |
  player == "Bo Nix" |
  player == "Michael Penix Jr." |
  player == "J.J. McCarthy" |
  player == "Spencer Rattler" |
  !is.na(nfl_pct),
  off_grade_avg >= 70,
  mean_epa_tot >= 0
) |>
ggplot(aes(y = mean_epa_tot, x = off_grade_avg)) +
labs(
  title = "QB College Production (Career)",
  subtitle = "outliers (Kyle Allen & Trevor Siemian) removed | blue = above average starter | red = below average starter",
  x = "Avg. PFF Grade",
  y = "EPA / Play",
  caption = "By: Sam Burch | Data: nflfastR, pff, and cfbfastR"
) +
geom_point(aes(color = nfl_pct >= .67)) +
```

```

geom_text_repel(size = 2, aes(label = player)) +
stat_smooth(formula = y ~ x, method = 'lm', geom = 'line', se=FALSE, color='grey30', linetype = 2, alpha = .3) +
nflplotR::geom_mean_lines(aes(y0 = mean_epa_tot, x0 = off_grade_avg), alpha = .3) +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5, size = 7),
  axis.line = element_line(color = "black", size = 0.5),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  legend.position = "none"
)

# ggsave("college-qb-career-production.png", width = 16, height = 9, units = "cm")

```

Looking at career production, Williams is atop, with Maye being second in grading and McCarthy being second in efficiency. While good production in college doesn't necessarily translate to NFL production – look at the mix of blue and red dots in the top right and bottom left – we know that production is important for draft position. Thus, with draft position being important historically, we know college production should lead to better prospects in general.

```

# Best College Production

df_final |> filter(
  # player == "Anthony Richardson" |
  # player == "Will Levis" |

  player == "Caleb Williams" |
  player == "Drake Maye" |
  player == "Jayden Daniels" |
  player == "Bo Nix" |
  player == "Michael Penix Jr." |
  player == "J.J. McCarthy" |
  player == "Spencer Rattler" |
  !is.na(nfl_pct),
  off_grade_avg >= 70,
  mean_epa_tot >= 0
) |>
ggplot(aes(y = mean_epa_max, x = off_grade_max)) +
labs(
  title = "QB College Production (Best Season)",
  subtitle = "outliers (Kyle Allen & Trevor Siemian) removed | blue = above average starter | red = below average starter",
  x = "PFF Grade",
  y = "EPA / Play",
  caption = "By: Sam Burch | Data: nflfastR, pff, and cfbfastR"
) +
geom_point(aes(color = nfl_pct >= .67)) +
geom_text_repel(size = 1.5, aes(label = player)) +
stat_smooth(formula = y ~ x, method = 'lm', geom = 'line', se=FALSE, color='grey30', linetype = 2, alpha = .3) +
nflplotR::geom_mean_lines(aes(y0 = mean_epa_max, x0 = off_grade_max), alpha = .3) +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5, size = 7),
  axis.line = element_line(color = "black", size = 0.5),

```

```

panel.grid = element_blank(),
panel.background = element_blank(),
legend.position = "none"
)

```

```
# ggsave("college-qb-max-production.png", width = 16, height = 9, units = "cm")
```

Looking at best season production for efficiency and grading, Daniels is by far the best of the 24' prospects. Nix is second, followed by a cluster including Williams, Penix, and McCarthy. Maye & Rattler had elite grading in their best seasons, but efficiency was below average here. Regardless, all of these prospects have had good best seasons, which should be a good sign.

```
# Ground vs. Air EPA
```

```

df_final |> filter(
  # player == "Anthony Richardson" |
  # player == "Will Levis" |

  player == "Caleb Williams" |
  player == "Drake Maye" |
  player == "Jayden Daniels" |
  player == "Bo Nix" |
  player == "Michael Penix Jr." |
  player == "J.J. McCarthy" |
  player == "Spencer Rattler" |
  !is.na(nfl_pct)
  # off_grade_avg >= 70
) |>
ggplot(aes(avg_ground_epa_tot, avg_pass_epa_tot)) +
labs(
  title = "Ground vs. Air EPA in College",
  subtitle = "blue = above average starter | red = below average | gray = prospect",
  y = "EPA / Pass",
  x = "EPA / (Rush + Sack)",
  caption = "By: Sam Burch | Data: nflfastR, pff, and cfbfastR"
) +
# scale_y_reverse() +
geom_point(aes(color = nfl_pct >= .67)) +
geom_text_repel(size = 2, aes(label = player)) +
stat_smooth(formula = y ~ x, method = 'lm', geom = 'line', se=FALSE, color='grey30', linetype = 2, alpha = .3) +
nflplotR::geom_mean_lines(aes(x0 = avg_ground_epa_tot, y0 = avg_pass_epa_tot), alpha = .3) +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5, size = 8),
  axis.line = element_line(color = "black", size = 0.5),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  legend.position = "none"
)

# ggsave("college-qb-ground-vs-air.png", width = 16, height = 9, units = "cm")

```

Here, we look at where on the field a QBs efficiency is coming from. The top right (good passing, rushing, and sack avoidance) is where one wants to be. Of the NFL QBs, only three are below average and six are

above average. All of the 24' prospects, expect Rattler, are in this quadrant. Also, having good ground efficiency gives QBs a solid floor, since very few above average starters had below average ground efficiency and 8 of the top 9 were above average starters. Nix and Daniels show up the best in ground efficiency giving them a solid floor. Keep in mind though, some of the *successful* QBs with great ground EPA include Hill and Mariota.

```
# Best Predictors Comp

df_final |> filter(
  # player == "Anthony Richardson" |
  # player == "Will Levis" |

  player == "Caleb Williams" |
  player == "Drake Maye" |
  player == "Jayden Daniels" |
  player == "Bo Nix" |
  player == "Michael Penix Jr." |
  player == "J.J. McCarthy" |
  player == "Spencer Rattler" |
  !is.na(nfl_pct)
) |>
ggplot(aes(y = adot_first, x = total_ground_epa_max)) +
labs(
  title = "Top 2 Predictors from College to Pro",
  subtitle = "correlation of 0.42 & 0.37 for ground epa and adot respectively | blue = above average",
  y = "ADOT in First Year",
  x = "Max Total Ground (Rush + Sack) EPA",
  caption = "By: Sam Burch | Data: nflfastR, pff, and cfbfastR"
) +
geom_point(aes(color = nfl_pct >= .67)) +
geom_text_repel(size = 1.8, aes(label = player)) +
stat_smooth(formula = y ~ x, method = 'lm', geom = 'line', se=FALSE, color='grey30', linetype = 2, alpha = .3) +
nflplotR::geom_mean_lines(aes(y0 = adot_first, x0 = total_ground_epa_max), alpha = .3) +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5, size = 6.5),
  axis.line = element_line(color = "black", size = 0.5),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  legend.position = "none"
)

# ggsave("college-qb-best-predictors-comp.png", width = 16, height = 9, units = "cm")
```

The top couple predictors are all dealing with ground efficiency. However, ADOT in a QBs first year is very high as well. As we compare the two, we can see there have only been three failures in the top right quadrant, compared to nine successes. Williams, McCarthy, Maye, and Penix are all in this area. Nix and Daniels don't have a high ADOT in their first year, however they have the second and third best max total ground EPA respectively. Again, this should lead to a high floor and is a good sign regardless of their ADOT.

```
# How QBs Deal with Pressure
```

```
df_final |> filter(
```

```

# player == "Anthony Richardson" |
# player == "Will Levis" |

player == "Caleb Williams" |
player == "Drake Maye" |
player == "Jayden Daniels" |
player == "Bo Nix" |
player == "Michael Penix Jr." |
player == "J.J. McCarthy" |
player == "Spencer Rattler" |
!is.na(nfl_pct)
) |>
ggplot(aes(avg_ground_epa_tot, pressure_to_sack_rate_tot)) +
labs(
  title = "How QBs Deal with Pressure in College",
  subtitle = "blue = above average starter | red = below average | gray = prospect",
  y = "PTSR",
  x = "EPA / (Rush + Sack)",
  caption = "By: Sam Burch | Data: nflfastR, pff, and cfbfastR"
) +
scale_y_reverse() +
geom_point(aes(color = nfl_pct >= .67)) +
geom_text_repel(size = 2, aes(label = player)) +
stat_smooth(formula = y ~ x, method = 'lm', geom = 'line', se=FALSE, color='grey30', linetype = 2, alpha = .3) +
nflplotR::geom_mean_lines(aes(x0 = avg_ground_epa_tot, y0 = pressure_to_sack_rate_tot), alpha = .3) +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5, size = 8),
  axis.line = element_line(color = "black", size = 0.5),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  legend.position = "none"
)

# ggsave("college-qb-best-pressure-performance.png", width = 16, height = 9, units = "cm")

```

Just looking at how QBs deal with pressure in college can tell us a lot. Penix has the best PTSR in this entire dataset. Nix also doesn't take many sacks and has the added benefit of his rushing. While Daniels has elite rushing too, he takes a lot of sacks. McCarthy has good (not great) sack avoidance, while Williams and Maye are on the border between good and bad. Rattler's below average in both, which is concerning.

Before moving onto modeling, let's see how draft position tends to NFL success. 15 of the 25 QBs drafted in round 1 become above average NFL starters. Meanwhile, only 4 of the 22 non-first rounders become above average starters. Hurts and Prescott were genuinely non-first round hits. Purdy still has something to prove after only playing with an elite offense around him. Hill is just efficient on a small sample. These players skew the correlations in recent history, along with first-round busts. However, they will be kept in the sample as we move onto modeling because of the already small sample of QBs in the dataset. Regardless, if NFL teams don't value a QB as a first rounder, this is a bad sign. This will be important when it comes to players like Nix and Penix, as they are fringe first rounders.

Modeling

Linear Regression

To get an idea of what a good RMSE and r^2 might look like, when we use the average (0.547) to predict QB success. We get an rmse of 0.348. This means, on average, we are off by ~35 on a scale from 0-100 (aka percentile).

```
##
## Call:
## lm(formula = nfl_pct ~ ., data = train_data[, c(top_predictors,
##       "nfl_pct")])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35996 -0.12196  0.02203  0.11325  0.33037
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.7525983   2.9271214  -0.599   0.560
## total_ground_epa_max -0.0142845   0.0138227  -1.033   0.322
## avg_ground_epa_tot    0.7379387   1.5374389   0.480   0.640
## avg_ground_epa_max    0.1156101   1.0043049   0.115   0.910
## adot_first         0.1787049   0.1264392   1.413   0.183
## avg_ground_epa_last  -0.2524811   0.9530846  -0.265   0.796
## total_ground_epa_last -0.0017298   0.0130998  -0.132   0.897
## total_epa_max        0.0022330   0.0035179   0.635   0.538
## total_ground_epa_min  0.0128665   0.0172220   0.747   0.469
## total_epa_tot       -0.0001300   0.0016087  -0.081   0.937
## avg_ground_epa_min  -0.5830096   0.8574345  -0.680   0.509
## mean_epa_max        0.1205732   1.6807682   0.072   0.944
## tds                 0.0094864   0.0091800   1.033   0.322
## total_ground_epa_tot  0.0013628   0.0039996   0.341   0.739
## total_ground_epa_first 0.0012807   0.0084897   0.151   0.883
## avg_sack_epa_last    0.2010000   0.4417726   0.455   0.657
## scrambles          0.0028963   0.0038404   0.754   0.465
## adot_max           -0.1040460   0.0887798  -1.172   0.264
## avg_sack_epa_max    -0.3703093   0.5515604  -0.671   0.515
## ypa_max            0.0006736   0.1639914   0.004   0.997
## rushes_tot         0.0010019   0.0014827   0.676   0.512
## off_grade_max       0.0079759   0.0247714   0.322   0.753
## tds_max            -0.0224596   0.0201321  -1.116   0.286
## ttt_max             0.4225154   0.5857394   0.721   0.485
## pressure_to_sack_rate_last -0.0281378   0.0396703  -0.709   0.492
## pressure_to_sack_rate_max -0.0501019   0.0452466  -1.107   0.290
## pressure_to_sack_rate_tot  4.3444263   6.7516534   0.643   0.532
##
## Residual standard error: 0.2844 on 12 degrees of freedom
## Multiple R-squared:  0.698, Adjusted R-squared:  0.04368
## F-statistic: 1.067 on 26 and 12 DF,  p-value: 0.4731

## Start:  AIC=-90.04
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
```

```

##      adot_first + avg_ground_epa_last + total_ground_epa_last +
##      total_epa_max + total_ground_epa_min + total_epa_tot + avg_ground_epa_min +
##      mean_epa_max + tds + total_ground_epa_tot + total_ground_epa_first +
##      avg_sack_epa_last + scrambles + adot_max + avg_sack_epa_max +
##      ypa_max + rushes_tot + off_grade_max + tds_max + ttt_max +
##      pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##      pressure_to_sack_rate_tot
##
##              Df Sum of Sq      RSS      AIC
## - ypa_max          1  0.000001  0.97070 -92.039
## - mean_epa_max      1  0.000416  0.97111 -92.022
## - total_epa_tot      1  0.000528  0.97122 -92.018
## - avg_ground_epa_max  1  0.001072  0.97177 -91.996
## - total_ground_epa_last  1  0.001411  0.97210 -91.982
## - total_ground_epa_first  1  0.001841  0.97253 -91.965
## - avg_ground_epa_last  1  0.005677  0.97637 -91.812
## - off_grade_max      1  0.008386  0.97908 -91.703
## - total_ground_epa_tot  1  0.009391  0.98009 -91.663
## - avg_sack_epa_last   1  0.016745  0.98744 -91.372
## - avg_ground_epa_tot   1  0.018636  0.98933 -91.297
## - total_epa_max       1  0.032592  1.00329 -90.751
## - pressure_to_sack_rate_tot  1  0.033492  1.00419 -90.716
## - avg_sack_epa_max     1  0.036462  1.00716 -90.601
## - rushes_tot          1  0.036936  1.00763 -90.582
## - avg_ground_epa_min   1  0.037398  1.00809 -90.565
## - pressure_to_sack_rate_last  1  0.040696  1.01139 -90.437
## - ttt_max             1  0.042090  1.01278 -90.383
## - total_ground_epa_min  1  0.045149  1.01584 -90.266
## - scrambles           1  0.046009  1.01670 -90.233
## <none>                                0.97069 -90.039
## - tds                 1  0.086380  1.05707 -88.714
## - total_ground_epa_max  1  0.086387  1.05708 -88.714
## - pressure_to_sack_rate_max  1  0.099183  1.06988 -88.245
## - tds_max             1  0.100676  1.07137 -88.190
## - adot_max            1  0.111103  1.08180 -87.813
## - adot_first          1  0.161588  1.13228 -86.034
##
## Step:  AIC=-92.04
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##      adot_first + avg_ground_epa_last + total_ground_epa_last +
##      total_epa_max + total_ground_epa_min + total_epa_tot + avg_ground_epa_min +
##      mean_epa_max + tds + total_ground_epa_tot + total_ground_epa_first +
##      avg_sack_epa_last + scrambles + adot_max + avg_sack_epa_max +
##      rushes_tot + off_grade_max + tds_max + ttt_max + pressure_to_sack_rate_last +
##      pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##              Df Sum of Sq      RSS      AIC
## - mean_epa_max      1  0.000431  0.97113 -94.022
## - total_epa_tot      1  0.000528  0.97122 -94.018
## - avg_ground_epa_max  1  0.001079  0.97177 -93.996
## - total_ground_epa_last  1  0.001814  0.97251 -93.966
## - total_ground_epa_first  1  0.001863  0.97256 -93.964
## - off_grade_max      1  0.008707  0.97940 -93.691
## - total_ground_epa_tot  1  0.009694  0.98039 -93.651

```

```

## - avg_ground_epa_last      1  0.012702 0.98340 -93.532
## - avg_sack_epa_last       1  0.017524 0.98822 -93.341
## - avg_ground_epa_tot      1  0.019098 0.98979 -93.279
## - pressure_to_sack_rate_tot 1  0.033993 1.00469 -92.696
## - total_epa_max           1  0.037626 1.00832 -92.556
## - rushes_tot              1  0.039193 1.00989 -92.495
## - avg_sack_epa_max         1  0.042445 1.01314 -92.370
## - ttt_max                 1  0.045122 1.01582 -92.267
## - scrambles               1  0.046016 1.01671 -92.233
## - total_ground_epa_min     1  0.046174 1.01687 -92.226
## - pressure_to_sack_rate_last 1  0.046707 1.01740 -92.206
## <none>                     0.97070 -92.039
## - avg_ground_epa_min      1  0.058700 1.02940 -91.749
## - tds                     1  0.086564 1.05726 -90.707
## - total_ground_epa_max     1  0.086989 1.05768 -90.692
## - tds_max                 1  0.100722 1.07142 -90.189
## - pressure_to_sack_rate_max 1  0.105094 1.07579 -90.030
## - adot_max                1  0.112934 1.08363 -89.747
## - adot_first              1  0.207800 1.17850 -86.474
##
## Step: AIC=-94.02
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##   adot_first + avg_ground_epa_last + total_ground_epa_last +
##   total_epa_max + total_ground_epa_min + total_epa_tot + avg_ground_epa_min +
##   tds + total_ground_epa_tot + total_ground_epa_first + avg_sack_epa_last +
##   scrambles + adot_max + avg_sack_epa_max + rushes_tot + off_grade_max +
##   tds_max + ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##   pressure_to_sack_rate_tot
##
##               Df Sum of Sq    RSS    AIC
## - total_epa_tot      1  0.000383 0.97151 -96.006
## - avg_ground_epa_max  1  0.001435 0.97256 -95.964
## - total_ground_epa_last 1  0.001545 0.97267 -95.960
## - total_ground_epa_first 1  0.001982 0.97311 -95.942
## - total_ground_epa_tot 1  0.009680 0.98081 -95.635
## - off_grade_max      1  0.013266 0.98439 -95.492
## - avg_sack_epa_last   1  0.017948 0.98907 -95.307
## - avg_ground_epa_tot  1  0.018931 0.99006 -95.269
## - avg_ground_epa_last 1  0.030812 1.00194 -94.803
## - pressure_to_sack_rate_tot 1  0.033885 1.00501 -94.684
## - rushes_tot         1  0.039670 1.01080 -94.460
## - avg_sack_epa_max    1  0.042328 1.01345 -94.358
## - total_ground_epa_min 1  0.046439 1.01756 -94.200
## - pressure_to_sack_rate_last 1  0.046476 1.01760 -94.198
## - scrambles          1  0.048257 1.01938 -94.130
## - ttt_max            1  0.049713 1.02084 -94.075
## <none>                0.97113 -94.022
## - total_epa_max      1  0.052309 1.02344 -93.975
## - avg_ground_epa_min  1  0.072356 1.04348 -93.219
## - tds                1  0.087438 1.05856 -92.659
## - total_ground_epa_max 1  0.087469 1.05860 -92.658
## - tds_max            1  0.101576 1.07270 -92.142
## - pressure_to_sack_rate_max 1  0.110164 1.08129 -91.831
## - adot_max           1  0.113774 1.08490 -91.701

```

```

## - adot_first          1  0.221127 1.19225 -88.021
##
## Step: AIC=-96.01
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##   adot_first + avg_ground_epa_last + total_ground_epa_last +
##   total_epa_max + total_ground_epa_min + avg_ground_epa_min +
##   tds + total_ground_epa_tot + total_ground_epa_first + avg_sack_epa_last +
##   scrambles + adot_max + avg_sack_epa_max + rushes_tot + off_grade_max +
##   tds_max + ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##   pressure_to_sack_rate_tot
##
##
##              Df Sum of Sq    RSS    AIC
## - total_ground_epa_last      1  0.001204 0.97271 -97.958
## - avg_ground_epa_max         1  0.001251 0.97276 -97.956
## - total_ground_epa_first      1  0.001626 0.97313 -97.941
## - total_ground_epa_tot        1  0.010669 0.98218 -97.580
## - off_grade_max               1  0.013358 0.98487 -97.474
## - avg_sack_epa_last           1  0.018544 0.99005 -97.269
## - avg_ground_epa_tot          1  0.021520 0.99303 -97.152
## - avg_ground_epa_last         1  0.032282 1.00379 -96.731
## - pressure_to_sack_rate_tot    1  0.038136 1.00964 -96.505
## - rushes_tot                  1  0.041494 1.01300 -96.375
## - avg_sack_epa_max            1  0.042087 1.01360 -96.352
## - total_ground_epa_min        1  0.046559 1.01807 -96.181
## - pressure_to_sack_rate_last   1  0.047299 1.01881 -96.152
## <none>                        0.97151 -96.006
## - ttt_max                     1  0.056026 1.02753 -95.820
## - scrambles                   1  0.057497 1.02901 -95.764
## - total_epa_max               1  0.063036 1.03454 -95.554
## - avg_ground_epa_min          1  0.074548 1.04606 -95.123
## - total_ground_epa_max        1  0.087771 1.05928 -94.633
## - tds_max                     1  0.105447 1.07696 -93.987
## - tds                         1  0.115841 1.08735 -93.613
## - adot_max                   1  0.118606 1.09012 -93.514
## - pressure_to_sack_rate_max    1  0.125547 1.09706 -93.266
## - adot_first                  1  0.221302 1.19281 -90.003
##
## Step: AIC=-97.96
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##   adot_first + avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##   avg_ground_epa_min + tds + total_ground_epa_tot + total_ground_epa_first +
##   avg_sack_epa_last + scrambles + adot_max + avg_sack_epa_max +
##   rushes_tot + off_grade_max + tds_max + ttt_max + pressure_to_sack_rate_last +
##   pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##
##              Df Sum of Sq    RSS    AIC
## - avg_ground_epa_max          1  0.000885 0.97360 -99.922
## - total_ground_epa_first      1  0.006372 0.97908 -99.703
## - total_ground_epa_tot        1  0.015249 0.98796 -99.351
## - off_grade_max               1  0.015510 0.98822 -99.341
## - avg_sack_epa_last           1  0.018737 0.99145 -99.214
## - avg_ground_epa_tot          1  0.028407 1.00112 -98.835
## - pressure_to_sack_rate_tot    1  0.038714 1.01143 -98.436
## - rushes_tot                  1  0.040947 1.01366 -98.350

```

```

## - avg_sack_epa_max          1  0.046327  1.01904 -98.143
## - pressure_to_sack_rate_last 1  0.047160  1.01987 -98.111
## - avg_ground_epa_last      1  0.047864  1.02058 -98.085
## <none>                      0.97271 -97.958
## - scrambles                 1  0.056397  1.02911 -97.760
## - ttt_max                   1  0.060845  1.03356 -97.592
## - total_epa_max             1  0.062560  1.03527 -97.527
## - avg_ground_epa_min        1  0.076231  1.04894 -97.015
## - total_ground_epa_min      1  0.085973  1.05869 -96.655
## - tds_max                   1  0.114091  1.08680 -95.633
## - total_ground_epa_max      1  0.130571  1.10328 -95.046
## - tds                       1  0.132418  1.10513 -94.980
## - pressure_to_sack_rate_max 1  0.141935  1.11465 -94.646
## - adot_max                  1  0.160689  1.13340 -93.995
## - adot_first                1  0.221152  1.19386 -91.968
##
## Step:  AIC=-99.92
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##          avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##          avg_ground_epa_min + tds + total_ground_epa_tot + total_ground_epa_first +
##          avg_sack_epa_last + scrambles + adot_max + avg_sack_epa_max +
##          rushes_tot + off_grade_max + tds_max + ttt_max + pressure_to_sack_rate_last +
##          pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - total_ground_epa_first  1  0.007914  0.98151 -101.607
## - total_ground_epa_tot    1  0.016026  0.98962 -101.286
## - avg_sack_epa_last       1  0.017852  0.99145 -101.214
## - off_grade_max           1  0.021208  0.99481 -101.082
## - pressure_to_sack_rate_tot 1  0.040718  1.01432 -100.325
## - avg_ground_epa_last     1  0.047415  1.02101 -100.068
## - rushes_tot              1  0.047431  1.02103 -100.067
## - pressure_to_sack_rate_last 1  0.049250  1.02285 -99.998
## - avg_sack_epa_max        1  0.049817  1.02342 -99.976
## <none>                      0.97360 -99.922
## - scrambles               1  0.058610  1.03221 -99.643
## - total_epa_max           1  0.061815  1.03541 -99.522
## - ttt_max                 1  0.064310  1.03791 -99.428
## - avg_ground_epa_min      1  0.081138  1.05474 -98.801
## - avg_ground_epa_tot      1  0.098940  1.07254 -98.148
## - tds_max                 1  0.113928  1.08753 -97.607
## - total_ground_epa_min    1  0.117022  1.09062 -97.496
## - tds                     1  0.131655  1.10525 -96.976
## - pressure_to_sack_rate_max 1  0.150947  1.12454 -96.301
## - total_ground_epa_max    1  0.155828  1.12943 -96.132
## - adot_max                1  0.163593  1.13719 -95.865
## - adot_first              1  0.246809  1.22041 -93.111
##
## Step:  AIC=-101.61
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##          avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##          avg_ground_epa_min + tds + total_ground_epa_tot + avg_sack_epa_last +
##          scrambles + adot_max + avg_sack_epa_max + rushes_tot + off_grade_max +
##          tds_max + ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +

```

```

##      pressure_to_sack_rate_tot
##
##              Df Sum of Sq      RSS       AIC
## - total_ground_epa_tot      1  0.009264  0.99078 -103.240
## - off_grade_max              1  0.013507  0.99502 -103.074
## - avg_sack_epa_last          1  0.015454  0.99697 -102.997
## - pressure_to_sack_rate_tot  1  0.034036  1.01555 -102.277
## - pressure_to_sack_rate_last 1  0.044622  1.02613 -101.873
## - rushes_tot                 1  0.047122  1.02863 -101.778
## - avg_sack_epa_max           1  0.048530  1.03004 -101.725
## <none>                       0.98151 -101.607
## - scrambles                  1  0.058799  1.04031 -101.338
## - ttt_max                    1  0.059585  1.04110 -101.308
## - avg_ground_epa_last        1  0.066670  1.04818 -101.044
## - avg_ground_epa_min         1  0.073247  1.05476 -100.800
## - total_epa_max              1  0.083052  1.06456 -100.439
## - avg_ground_epa_tot         1  0.092554  1.07407 -100.092
## - tds_max                    1  0.125339  1.10685  -98.920
## - tds                        1  0.132285  1.11380  -98.676
## - pressure_to_sack_rate_max   1  0.149795  1.13131  -98.067
## - total_ground_epa_max       1  0.150704  1.13222  -98.036
## - adot_max                   1  0.179356  1.16087  -97.061
## - total_ground_epa_min       1  0.237574  1.21909  -95.153
## - adot_first                 1  0.301406  1.28292  -93.163
##
## Step:  AIC=-103.24
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##      avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##      avg_ground_epa_min + tds + avg_sack_epa_last + scrambles +
##      adot_max + avg_sack_epa_max + rushes_tot + off_grade_max +
##      tds_max + ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##      pressure_to_sack_rate_tot
##
##              Df Sum of Sq      RSS       AIC
## - off_grade_max      1  0.014198  1.00497 -104.685
## - avg_sack_epa_last   1  0.016346  1.00712 -104.602
## - pressure_to_sack_rate_tot  1  0.042894  1.03367 -103.587
## - avg_sack_epa_max    1  0.049682  1.04046 -103.332
## - scrambles           1  0.051044  1.04182 -103.281
## <none>                 0.99078 -103.240
## - pressure_to_sack_rate_last 1  0.055755  1.04653 -103.105
## - rushes_tot          1  0.058543  1.04932 -103.001
## - avg_ground_epa_min  1  0.064357  1.05513 -102.786
## - ttt_max              1  0.074088  1.06486 -102.428
## - total_epa_max        1  0.074415  1.06519 -102.416
## - avg_ground_epa_last  1  0.077492  1.06827 -102.303
## - avg_ground_epa_tot   1  0.084382  1.07516 -102.053
## - tds_max              1  0.116298  1.10707 -100.912
## - tds                   1  0.124552  1.11533 -100.622
## - total_ground_epa_max 1  0.146877  1.13765  -99.849
## - adot_max             1  0.174394  1.16517  -98.917
## - pressure_to_sack_rate_max 1  0.190401  1.18118  -98.385
## - total_ground_epa_min 1  0.290784  1.28156  -95.204
## - adot_first           1  0.304149  1.29492  -94.799

```



```

##
## Step: AIC=-104.69
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##     avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##     avg_ground_epa_min + tds + avg_sack_epa_last + scrambles +
##     adot_max + avg_sack_epa_max + rushes_tot + tds_max + ttt_max +
##     pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##     pressure_to_sack_rate_tot
##
##
##           Df Sum of Sq   RSS   AIC
## - avg_sack_epa_last      1   0.01136 1.0163 -106.247
## - pressure_to_sack_rate_tot 1   0.03602 1.0410 -105.312
## - rushes_tot             1   0.04541 1.0504 -104.962
## - avg_sack_epa_max       1   0.04908 1.0540 -104.826
## - pressure_to_sack_rate_last 1   0.05255 1.0575 -104.698
## <none>                     1.0050 -104.685
## - avg_ground_epa_min     1   0.05365 1.0586 -104.657
## - avg_ground_epa_last    1   0.06554 1.0705 -104.222
## - scrambles              1   0.06852 1.0735 -104.113
## - avg_ground_epa_tot     1   0.07464 1.0796 -103.891
## - ttt_max                1   0.07746 1.0824 -103.790
## - tds_max                1   0.17923 1.1842 -100.285
## - pressure_to_sack_rate_max 1   0.18071 1.1857 -100.236
## - total_ground_epa_max   1   0.18186 1.1868 -100.199
## - adot_max               1   0.19112 1.1961 -99.895
## - total_epa_max          1   0.19113 1.1961 -99.895
## - tds                    1   0.22975 1.2347 -98.656
## - total_ground_epa_min   1   0.31942 1.3244 -95.922
## - adot_first             1   0.34099 1.3460 -95.292
##
## Step: AIC=-106.25
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##     avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##     avg_ground_epa_min + tds + scrambles + adot_max + avg_sack_epa_max +
##     rushes_tot + tds_max + ttt_max + pressure_to_sack_rate_last +
##     pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##
##           Df Sum of Sq   RSS   AIC
## - rushes_tot             1   0.03652 1.0529 -106.870
## - avg_sack_epa_max       1   0.04197 1.0583 -106.669
## - avg_ground_epa_min     1   0.04387 1.0602 -106.599
## - pressure_to_sack_rate_tot 1   0.05182 1.0681 -106.308
## <none>                     1.0163 -106.247
## - pressure_to_sack_rate_last 1   0.06048 1.0768 -105.993
## - avg_ground_epa_tot     1   0.07245 1.0888 -105.562
## - ttt_max                1   0.07677 1.0931 -105.407
## - avg_ground_epa_last    1   0.08889 1.1052 -104.977
## - scrambles              1   0.09150 1.1078 -104.885
## - total_ground_epa_max   1   0.19049 1.2068 -101.547
## - tds_max                1   0.20145 1.2178 -101.195
## - adot_max               1   0.20559 1.2219 -101.062
## - tds                    1   0.22278 1.2391 -100.518
## - total_epa_max          1   0.22563 1.2420 -100.428
## - pressure_to_sack_rate_max 1   0.23597 1.2523 -100.105

```

```

## - total_ground_epa_min      1    0.31533 1.3316 -97.708
## - adot_first                1    0.46005 1.4764 -93.685
##
## Step: AIC=-106.87
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##         avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##         avg_ground_epa_min + tds + scrambles + adot_max + avg_sack_epa_max +
##         tds_max + ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##         pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - avg_sack_epa_max      1    0.02199 1.0748 -108.064
## - avg_ground_epa_min    1    0.02503 1.0779 -107.954
## - pressure_to_sack_rate_tot 1    0.04703 1.0999 -107.166
## - ttt_max               1    0.05184 1.1047 -106.996
## <none>                  1    0.0529 1.1052 -106.870
## - pressure_to_sack_rate_last 1    0.05629 1.1091 -106.839
## - avg_ground_epa_last    1    0.07566 1.1285 -106.164
## - avg_ground_epa_tot     1    0.08015 1.1330 -106.009
## - scrambles              1    0.16431 1.2172 -103.215
## - total_ground_epa_max   1    0.16880 1.2216 -103.071
## - adot_max               1    0.18871 1.2416 -102.440
## - pressure_to_sack_rate_max 1    0.21138 1.2642 -101.735
## - tds_max                1    0.23034 1.2832 -101.154
## - total_epa_max          1    0.23481 1.2877 -101.019
## - tds                    1    0.27148 1.3243 -99.924
## - total_ground_epa_min   1    0.27947 1.3323 -99.689
## - adot_first             1    0.43927 1.4921 -95.271
##
## Step: AIC=-108.06
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##         avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##         avg_ground_epa_min + tds + scrambles + adot_max + tds_max +
##         ttt_max + pressure_to_sack_rate_last + pressure_to_sack_rate_max +
##         pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - avg_ground_epa_min    1    0.01872 1.0936 -109.39
## - ttt_max               1    0.04038 1.1152 -108.63
## - pressure_to_sack_rate_last 1    0.04435 1.1192 -108.49
## - pressure_to_sack_rate_tot 1    0.04641 1.1213 -108.42
## <none>                  1    0.0748 1.1074 -108.06
## - avg_ground_epa_last    1    0.07231 1.1471 -107.53
## - avg_ground_epa_tot     1    0.07432 1.1492 -107.46
## - total_ground_epa_max   1    0.16864 1.2435 -104.38
## - scrambles              1    0.17245 1.2473 -104.26
## - adot_max               1    0.17991 1.2548 -104.03
## - tds_max                1    0.22185 1.2967 -102.75
## - total_epa_max          1    0.22962 1.3045 -102.51
## - pressure_to_sack_rate_max 1    0.25334 1.3282 -101.81
## - tds                    1    0.25632 1.3312 -101.72
## - total_ground_epa_min   1    0.26219 1.3370 -101.55
## - adot_first             1    0.41732 1.4922 -97.27
##

```

```

## Step: AIC=-109.39
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##     avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##     tds + scrambles + adot_max + tds_max + ttt_max + pressure_to_sack_rate_last +
##     pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##           Df Sum of Sq   RSS   AIC
## - ttt_max      1    0.04247 1.1360 -109.905
## <none>                        1.0936 -109.391
## - pressure_to_sack_rate_tot  1    0.05845 1.1520 -109.360
## - pressure_to_sack_rate_last 1    0.07052 1.1641 -108.953
## - avg_ground_epa_tot      1    0.08486 1.1784 -108.476
## - avg_ground_epa_last     1    0.11479 1.2084 -107.498
## - total_ground_epa_max    1    0.15744 1.2510 -106.145
## - scrambles              1    0.17414 1.2677 -105.628
## - adot_max               1    0.17923 1.2728 -105.472
## - tds_max                1    0.20653 1.3001 -104.644
## - total_epa_max          1    0.21090 1.3045 -104.513
## - pressure_to_sack_rate_max 1    0.25649 1.3500 -103.173
## - tds                    1    0.26332 1.3569 -102.977
## - total_ground_epa_min    1    0.26977 1.3633 -102.792
## - adot_first             1    0.43682 1.5304  -98.284
##
## Step: AIC=-109.9
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##     avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##     tds + scrambles + adot_max + tds_max + pressure_to_sack_rate_last +
##     pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##           Df Sum of Sq   RSS   AIC
## - pressure_to_sack_rate_tot  1    0.04683 1.1829 -110.329
## <none>                        1.1360 -109.905
## - pressure_to_sack_rate_last 1    0.06656 1.2026 -109.684
## - avg_ground_epa_tot      1    0.07015 1.2062 -109.568
## - avg_ground_epa_last     1    0.10937 1.2454 -108.320
## - total_ground_epa_max    1    0.12837 1.2644 -107.730
## - tds_max                 1    0.16839 1.3044 -106.514
## - adot_max                1    0.17426 1.3103 -106.339
## - total_epa_max           1    0.18486 1.3209 -106.025
## - pressure_to_sack_rate_max 1    0.21659 1.3526 -105.099
## - tds                     1    0.22087 1.3569 -104.976
## - total_ground_epa_min    1    0.27357 1.4096 -103.490
## - scrambles               1    0.35215 1.4882 -101.374
## - adot_first              1    0.42769 1.5637  -99.443
##
## Step: AIC=-110.33
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##     avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##     tds + scrambles + adot_max + tds_max + pressure_to_sack_rate_last +
##     pressure_to_sack_rate_max
##
##           Df Sum of Sq   RSS   AIC
## - pressure_to_sack_rate_last 1    0.01986 1.2027 -111.68
## <none>                        1.1829 -110.33

```

```

## - avg_ground_epa_last      1  0.07365 1.2565 -109.97
## - avg_ground_epa_tot      1  0.08943 1.2723 -109.49
## - adot_max                1  0.14734 1.3302 -107.75
## - total_ground_epa_max    1  0.16407 1.3469 -107.26
## - tds_max                 1  0.17738 1.3602 -106.88
## - total_epa_max           1  0.20876 1.3916 -105.99
## - pressure_to_sack_rate_max 1  0.21094 1.3938 -105.93
## - tds                     1  0.21563 1.3985 -105.80
## - total_ground_epa_min    1  0.25124 1.4341 -104.82
## - scrambles               1  0.30583 1.4887 -103.36
## - adot_first              1  0.38150 1.5644 -101.43
##
## Step: AIC=-111.68
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##          avg_ground_epa_last + total_epa_max + total_ground_epa_min +
##          tds + scrambles + adot_max + tds_max + pressure_to_sack_rate_max
##
##              Df Sum of Sq    RSS    AIC
## - avg_ground_epa_last      1  0.05861 1.2613 -111.82
## <none>                      1.2027 -111.68
## - avg_ground_epa_tot      1  0.08018 1.2829 -111.16
## - adot_max                 1  0.13365 1.3364 -109.57
## - total_ground_epa_max    1  0.18475 1.3875 -108.11
## - tds_max                  1  0.19816 1.4009 -107.73
## - total_epa_max           1  0.22446 1.4272 -107.01
## - tds                     1  0.25050 1.4532 -106.30
## - total_ground_epa_min    1  0.27946 1.4822 -105.53
## - scrambles                1  0.32677 1.5295 -104.31
## - adot_first              1  0.36166 1.5644 -103.43
## - pressure_to_sack_rate_max 1  0.48205 1.6848 -100.53
##
## Step: AIC=-111.82
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + adot_first +
##          total_epa_max + total_ground_epa_min + tds + scrambles +
##          adot_max + tds_max + pressure_to_sack_rate_max
##
##              Df Sum of Sq    RSS    AIC
## - avg_ground_epa_tot      1  0.04843 1.3098 -112.36
## <none>                      1.2613 -111.82
## - adot_max                 1  0.12217 1.3835 -110.22
## - total_ground_epa_max    1  0.16965 1.4310 -108.90
## - total_epa_max           1  0.20549 1.4668 -107.94
## - tds_max                  1  0.21262 1.4740 -107.75
## - total_ground_epa_min    1  0.22599 1.4873 -107.40
## - scrambles                1  0.26855 1.5299 -106.30
## - tds                     1  0.28586 1.5472 -105.86
## - adot_first              1  0.30667 1.5680 -105.34
## - pressure_to_sack_rate_max 1  0.44153 1.7029 -102.12
##
## Step: AIC=-112.35
## nfl_pct ~ total_ground_epa_max + adot_first + total_epa_max +
##          total_ground_epa_min + tds + scrambles + adot_max + tds_max +
##          pressure_to_sack_rate_max
##

```

```

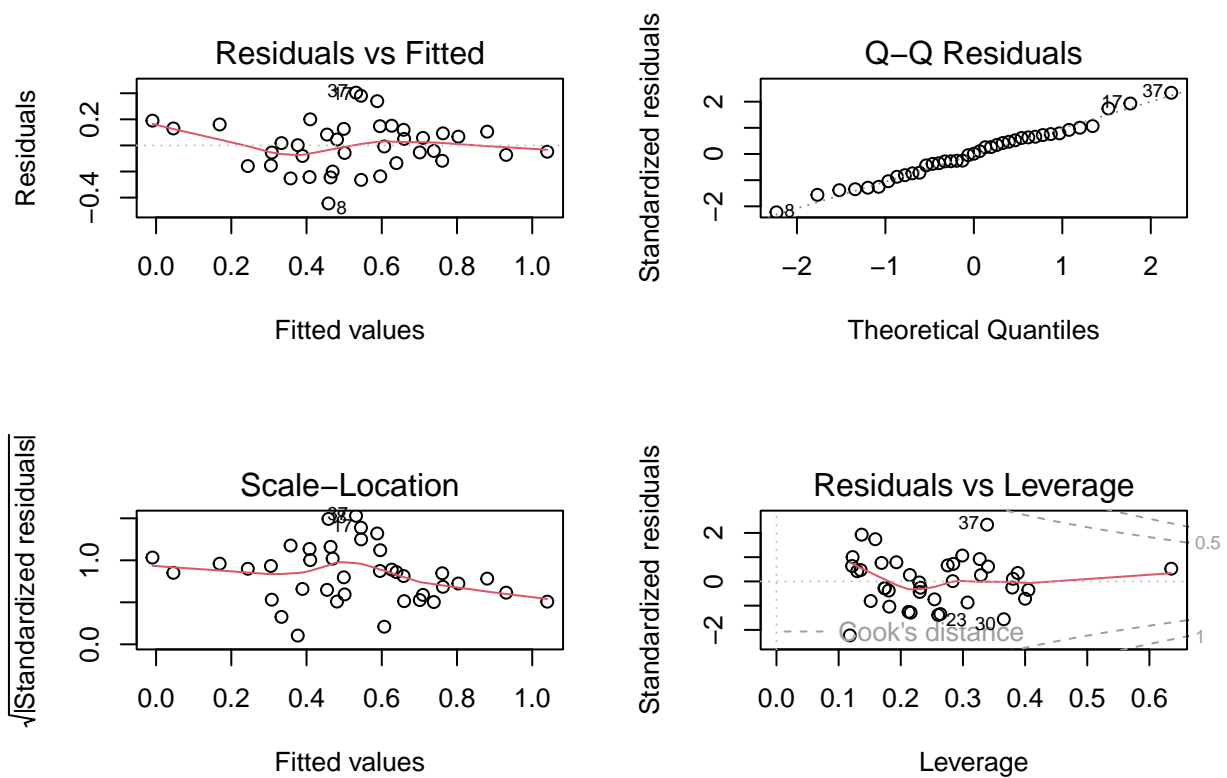
##              Df Sum of Sq    RSS    AIC
## <none>              1.3098 -112.36
## - adot_max          1  0.12381 1.4336 -110.83
## - total_ground_epa_max 1  0.12387 1.4336 -110.83
## - total_epa_max      1  0.17000 1.4798 -109.59
## - tds_max            1  0.17277 1.4825 -109.52
## - tds                1  0.24001 1.5498 -107.79
## - adot_first        1  0.34403 1.6538 -105.26
## - total_ground_epa_min 1  0.38989 1.6997 -104.19
## - pressure_to_sack_rate_max 1  0.40804 1.7178 -103.78
## - scrambles         1  0.51763 1.8274 -101.37

##
## Call:
## lm(formula = nfl_pct ~ total_ground_epa_max + adot_first + total_epa_max +
##   total_ground_epa_min + tds + scrambles + adot_max + tds_max +
##   pressure_to_sack_rate_max, data = train_data[, c(top_predictors,
## "nfl_pct")])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44486 -0.12669  0.00194  0.12304  0.40446
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.643913   0.383692   1.678  0.10405
## total_ground_epa_max -0.007598   0.004588  -1.656  0.10849
## adot_first      0.136423   0.049430   2.760  0.00991 **
## total_epa_max     0.002358   0.001215   1.940  0.06214 .
## total_ground_epa_min 0.011382   0.003874   2.938  0.00641 **
## tds              0.008139   0.003531   2.305  0.02850 *
## scrambles       0.004772   0.001410   3.385  0.00206 **
## adot_max        -0.089268   0.053916  -1.656  0.10857
## tds_max         -0.020382   0.010421  -1.956  0.06017 .
## pressure_to_sack_rate_max -0.030473   0.010138  -3.006  0.00542 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2125 on 29 degrees of freedom
## Multiple R-squared:  0.5925, Adjusted R-squared:  0.4661
## F-statistic: 4.685 on 9 and 29 DF,  p-value: 0.0006855

## [1] 0.2324643

##      total_ground_epa_max      adot_first      total_epa_max
##      2.856258      1.690276      2.642465
##      total_ground_epa_min      tds      scrambles
##      2.300496      2.478417      1.524338
##      adot_max      tds_max pressure_to_sack_rate_max
##      1.855259      3.453850      1.360051

```



```
##          37          30          23          8          32          17          13
## 0.28076460 0.14046255 0.06731103 0.06648024 0.06486225 0.05909015 0.05731830
##          34          35          31
## 0.04857198 0.04775582 0.04575045
```

```
##
## studentized Breusch-Pagan test
##
## data: m2a
## BP = 4.0033, df = 9, p-value = 0.9112
```

```
##
## Shapiro-Wilk normality test
##
## data: m2a$residuals
## W = 0.98264, p-value = 0.7975
```

For the linear regression we can't fit a model with all the potential predictors, because of dimensional issues. Therefore, we start by fitting one with predictors where their absolute correlations are 0.275 or more. Then, by optimizing using the greedy algorithm this gives us the following: an adjusted r^2 of 0.466 and a rmse of 0.232. The adjusted r^2 is very good and the error is not bad, but collinearity is a problem. Hence, it is not smart to consider this model in the model selection later now.

```
##
## Call:
```

```

## lm(formula = nfl_pct ~ ., data = train_data[, c(top_predictors,
##      "nfl_pct")]))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46997 -0.13028  0.02453  0.14884  0.34716
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.129814   0.839743   1.345   0.193
## total_ground_epa_max -0.003757   0.010730  -0.350   0.730
## avg_ground_epa_tot    1.362066   1.093232   1.246   0.227
## avg_ground_epa_max   -0.418235   0.744521  -0.562   0.580
## adot_first          0.007356   0.064383   0.114   0.910
## avg_ground_epa_last  -0.022655   0.476190  -0.048   0.963
## total_ground_epa_last  0.003023   0.007620   0.397   0.696
## total_epa_max       -0.001232   0.002194  -0.561   0.580
## total_ground_epa_min  0.001234   0.009800   0.126   0.901
## total_epa_tot        0.001032   0.001213   0.851   0.404
## avg_ground_epa_min   -0.186818   0.556330  -0.336   0.740
## mean_epa_max         0.411449   1.313263   0.313   0.757
## tds                  0.000339   0.004551   0.074   0.941
## total_ground_epa_tot  -0.002160   0.003132  -0.690   0.498
## total_ground_epa_first -0.001341   0.006305  -0.213   0.834
## avg_sack_epa_last     0.050758   0.246856   0.206   0.839
## pressure_to_sack_rate_max -0.009777   0.024012  -0.407   0.688
## pressure_to_sack_rate_tot -1.989347   2.876903  -0.691   0.497
##
## Residual standard error: 0.2811 on 21 degrees of freedom
## Multiple R-squared:  0.4837, Adjusted R-squared:  0.06583
## F-statistic: 1.158 on 17 and 21 DF, p-value: 0.3708

## Start:  AIC=-87.13
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##      adot_first + avg_ground_epa_last + total_ground_epa_last +
##      total_epa_max + total_ground_epa_min + total_epa_tot + avg_ground_epa_min +
##      mean_epa_max + tds + total_ground_epa_tot + total_ground_epa_first +
##      avg_sack_epa_last + pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - avg_ground_epa_last      1  0.000179 1.6596 -89.124
## - tds                      1  0.000438 1.6598 -89.117
## - adot_first               1  0.001032 1.6604 -89.103
## - total_ground_epa_min      1  0.001252 1.6606 -89.098
## - avg_sack_epa_last         1  0.003341 1.6627 -89.049
## - total_ground_epa_first    1  0.003576 1.6629 -89.044
## - mean_epa_max              1  0.007756 1.6671 -88.946
## - avg_ground_epa_min        1  0.008910 1.6683 -88.919
## - total_ground_epa_max      1  0.009688 1.6691 -88.901
## - total_ground_epa_last     1  0.012435 1.6718 -88.837
## - pressure_to_sack_rate_max  1  0.013100 1.6725 -88.821
## - total_epa_max             1  0.024911 1.6843 -88.547
## - avg_ground_epa_max        1  0.024935 1.6843 -88.546
## - total_ground_epa_tot      1  0.037570 1.6969 -88.255

```

```

## - pressure_to_sack_rate_tot 1 0.037783 1.6972 -88.250
## - total_epa_tot 1 0.057269 1.7166 -87.804
## <none> 1.6594 -87.128
## - avg_ground_epa_tot 1 0.122658 1.7820 -86.346
##
## Step: AIC=-89.12
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
## adot_first + total_ground_epa_last + total_epa_max + total_ground_epa_min +
## total_epa_tot + avg_ground_epa_min + mean_epa_max + tds +
## total_ground_epa_tot + total_ground_epa_first + avg_sack_epa_last +
## pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
## Df Sum of Sq RSS AIC
## - tds 1 0.000499 1.6601 -91.112
## - adot_first 1 0.000858 1.6604 -91.103
## - total_ground_epa_min 1 0.001105 1.6607 -91.098
## - total_ground_epa_first 1 0.003398 1.6629 -91.044
## - avg_sack_epa_last 1 0.003575 1.6631 -91.040
## - total_ground_epa_max 1 0.009939 1.6695 -90.891
## - total_ground_epa_last 1 0.013282 1.6728 -90.813
## - mean_epa_max 1 0.014390 1.6739 -90.787
## - pressure_to_sack_rate_max 1 0.014792 1.6743 -90.777
## - avg_ground_epa_min 1 0.018072 1.6776 -90.701
## - avg_ground_epa_max 1 0.027149 1.6867 -90.491
## - total_epa_max 1 0.033189 1.6927 -90.351
## - total_ground_epa_tot 1 0.039514 1.6991 -90.206
## - pressure_to_sack_rate_tot 1 0.044431 1.7040 -90.093
## - total_epa_tot 1 0.059166 1.7187 -89.757
## <none> 1.6596 -89.124
## - avg_ground_epa_tot 1 0.132765 1.7923 -88.122
##
## Step: AIC=-91.11
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
## adot_first + total_ground_epa_last + total_epa_max + total_ground_epa_min +
## total_epa_tot + avg_ground_epa_min + mean_epa_max + total_ground_epa_tot +
## total_ground_epa_first + avg_sack_epa_last + pressure_to_sack_rate_max +
## pressure_to_sack_rate_tot
##
## Df Sum of Sq RSS AIC
## - adot_first 1 0.001003 1.6611 -93.088
## - total_ground_epa_min 1 0.001339 1.6614 -93.080
## - avg_sack_epa_last 1 0.003081 1.6631 -93.039
## - total_ground_epa_first 1 0.005343 1.6654 -92.986
## - total_ground_epa_max 1 0.009592 1.6696 -92.887
## - total_ground_epa_last 1 0.013194 1.6732 -92.803
## - pressure_to_sack_rate_max 1 0.014758 1.6748 -92.767
## - mean_epa_max 1 0.014864 1.6749 -92.764
## - avg_ground_epa_min 1 0.019902 1.6800 -92.647
## - avg_ground_epa_max 1 0.029017 1.6891 -92.436
## - total_epa_max 1 0.036340 1.6964 -92.267
## - pressure_to_sack_rate_tot 1 0.051425 1.7115 -91.922
## - total_ground_epa_tot 1 0.066455 1.7265 -91.581
## <none> 1.6601 -91.112
## - total_epa_tot 1 0.101400 1.7614 -90.800

```



```

## - avg_ground_epa_tot          1  0.157620 1.8177 -89.574
##
## Step:  AIC=-93.09
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##      total_ground_epa_last + total_epa_max + total_ground_epa_min +
##      total_epa_tot + avg_ground_epa_min + mean_epa_max + total_ground_epa_tot +
##      total_ground_epa_first + avg_sack_epa_last + pressure_to_sack_rate_max +
##      pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq    RSS    AIC
## - total_ground_epa_min      1  0.000820 1.6619 -95.069
## - avg_sack_epa_last          1  0.004270 1.6653 -94.988
## - total_ground_epa_first     1  0.004886 1.6659 -94.974
## - total_ground_epa_max       1  0.009915 1.6710 -94.856
## - pressure_to_sack_rate_max   1  0.015329 1.6764 -94.730
## - total_ground_epa_last      1  0.015537 1.6766 -94.725
## - mean_epa_max               1  0.017649 1.6787 -94.676
## - avg_ground_epa_min         1  0.021125 1.6822 -94.595
## - avg_ground_epa_max         1  0.031400 1.6925 -94.358
## - total_epa_max              1  0.040235 1.7013 -94.155
## - pressure_to_sack_rate_tot   1  0.050830 1.7119 -93.913
## - total_ground_epa_tot       1  0.067041 1.7281 -93.545
## <none>                      1.6611 -93.088
## - total_epa_tot              1  0.107833 1.7689 -92.635
## - avg_ground_epa_tot         1  0.174346 1.8354 -91.196
##
## Step:  AIC=-95.07
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##      total_ground_epa_last + total_epa_max + total_epa_tot + avg_ground_epa_min +
##      mean_epa_max + total_ground_epa_tot + total_ground_epa_first +
##      avg_sack_epa_last + pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq    RSS    AIC
## - avg_sack_epa_last          1  0.003480 1.6654 -96.987
## - total_ground_epa_first     1  0.005668 1.6675 -96.936
## - pressure_to_sack_rate_max   1  0.014539 1.6764 -96.729
## - total_ground_epa_max       1  0.017305 1.6792 -96.665
## - mean_epa_max               1  0.022734 1.6846 -96.539
## - avg_ground_epa_min         1  0.023365 1.6852 -96.524
## - avg_ground_epa_max         1  0.031694 1.6936 -96.332
## - total_ground_epa_last      1  0.040529 1.7024 -96.129
## - total_epa_max              1  0.045482 1.7074 -96.016
## - pressure_to_sack_rate_tot   1  0.051671 1.7135 -95.875
## <none>                      1.6619 -95.069
## - total_ground_epa_tot       1  0.093129 1.7550 -94.943
## - total_epa_tot              1  0.107465 1.7693 -94.625
## - avg_ground_epa_tot         1  0.194203 1.8561 -92.759
##
## Step:  AIC=-96.99
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##      total_ground_epa_last + total_epa_max + total_epa_tot + avg_ground_epa_min +
##      mean_epa_max + total_ground_epa_tot + total_ground_epa_first +
##      pressure_to_sack_rate_max + pressure_to_sack_rate_tot
##
##

```

```

##              Df Sum of Sq    RSS    AIC
## - total_ground_epa_first      1  0.004880 1.6702 -98.873
## - pressure_to_sack_rate_max    1  0.013797 1.6791 -98.666
## - total_ground_epa_max        1  0.021718 1.6871 -98.482
## - avg_ground_epa_min          1  0.023588 1.6889 -98.439
## - mean_epa_max                1  0.028840 1.6942 -98.318
## - avg_ground_epa_max          1  0.033972 1.6993 -98.200
## - total_ground_epa_last       1  0.050546 1.7159 -97.821
## - pressure_to_sack_rate_tot    1  0.050759 1.7161 -97.816
## - total_epa_max               1  0.057099 1.7225 -97.673
## <none>                        1.6654 -96.987
## - total_ground_epa_tot        1  0.094570 1.7599 -96.833
## - total_epa_tot               1  0.124227 1.7896 -96.182
## - avg_ground_epa_tot          1  0.205537 1.8709 -94.449
##
## Step:  AIC=-98.87
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##           total_ground_epa_last + total_epa_max + total_epa_tot + avg_ground_epa_min +
##           mean_epa_max + total_ground_epa_tot + pressure_to_sack_rate_max +
##           pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - pressure_to_sack_rate_max    1  0.011340 1.6816 -100.609
## - avg_ground_epa_min          1  0.021665 1.6919 -100.371
## - mean_epa_max                1  0.025296 1.6955 -100.287
## - avg_ground_epa_max          1  0.030922 1.7012 -100.158
## - total_ground_epa_max        1  0.037472 1.7077 -100.008
## - total_epa_max               1  0.054933 1.7252  -99.611
## - pressure_to_sack_rate_tot    1  0.054943 1.7252  -99.611
## - total_ground_epa_last       1  0.069992 1.7402  -99.272
## <none>                        1.6702  -98.873
## - total_ground_epa_tot        1  0.112177 1.7824  -98.338
## - total_epa_tot               1  0.144577 1.8148  -97.636
## - avg_ground_epa_tot          1  0.200857 1.8711  -96.445
##
## Step:  AIC=-100.61
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##           total_ground_epa_last + total_epa_max + total_epa_tot + avg_ground_epa_min +
##           mean_epa_max + total_ground_epa_tot + pressure_to_sack_rate_tot
##
##              Df Sum of Sq    RSS    AIC
## - mean_epa_max                1  0.01915 1.7007 -102.168
## - avg_ground_epa_min          1  0.02069 1.7023 -102.132
## - avg_ground_epa_max          1  0.02918 1.7108 -101.938
## - total_ground_epa_max        1  0.04264 1.7242 -101.633
## - total_epa_max               1  0.04844 1.7300 -101.502
## - total_ground_epa_last       1  0.06906 1.7506 -101.040
## <none>                        1.6816 -100.609
## - total_ground_epa_tot        1  0.10299 1.7846 -100.291
## - total_epa_tot               1  0.16124 1.8428  -99.038
## - avg_ground_epa_tot          1  0.19299 1.8746  -98.372
## - pressure_to_sack_rate_tot    1  0.33057 2.0121  -95.610
##
## Step:  AIC=-102.17

```

```

## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##   total_ground_epa_last + total_epa_max + total_epa_tot + avg_ground_epa_min +
##   total_ground_epa_tot + pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq  RSS    AIC
## - avg_ground_epa_min      1  0.01182 1.7125 -103.898
## - avg_ground_epa_max      1  0.01831 1.7190 -103.750
## - total_epa_max           1  0.02970 1.7304 -103.493
## - total_ground_epa_max    1  0.03777 1.7385 -103.311
## - total_ground_epa_last   1  0.06432 1.7651 -102.720
## <none>                     1.7007 -102.168
## - total_ground_epa_tot    1  0.09829 1.7990 -101.977
## - avg_ground_epa_tot      1  0.17414 1.8749 -100.366
## - total_epa_tot           1  0.19567 1.8964  -99.921
## - pressure_to_sack_rate_tot 1  0.32917 2.0299  -97.268
##
## Step:  AIC=-103.9
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + avg_ground_epa_max +
##   total_ground_epa_last + total_epa_max + total_epa_tot + total_ground_epa_tot +
##   pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq  RSS    AIC
## - avg_ground_epa_max      1  0.01317 1.7257 -105.599
## - total_ground_epa_max    1  0.03016 1.7427 -105.217
## - total_epa_max           1  0.04168 1.7542 -104.960
## - total_ground_epa_last   1  0.08486 1.7974 -104.012
## <none>                     1.7125 -103.898
## - total_ground_epa_tot    1  0.13671 1.8493 -102.903
## - total_epa_tot           1  0.20556 1.9181 -101.477
## - avg_ground_epa_tot      1  0.29616 2.0087  -99.677
## - pressure_to_sack_rate_tot 1  0.37126 2.0838  -98.245
##
## Step:  AIC=-105.6
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + total_ground_epa_last +
##   total_epa_max + total_epa_tot + total_ground_epa_tot + pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq  RSS    AIC
## - total_epa_max           1  0.04984 1.7755 -106.489
## <none>                     1.7257 -105.599
## - total_ground_epa_max    1  0.11581 1.8415 -105.066
## - total_ground_epa_tot    1  0.12375 1.8495 -104.898
## - total_ground_epa_last   1  0.12574 1.8515 -104.856
## - total_epa_tot           1  0.21690 1.9426 -102.982
## - pressure_to_sack_rate_tot 1  0.37379 2.0995  -99.953
## - avg_ground_epa_tot      1  0.48408 2.2098  -97.956
##
## Step:  AIC=-106.49
## nfl_pct ~ total_ground_epa_max + avg_ground_epa_tot + total_ground_epa_last +
##   total_epa_tot + total_ground_epa_tot + pressure_to_sack_rate_tot
##
##
##      Df Sum of Sq  RSS    AIC
## - total_ground_epa_max    1  0.07002 1.8456 -106.980
## - total_ground_epa_last   1  0.07737 1.8529 -106.825
## <none>                     1.7755 -106.489

```

```

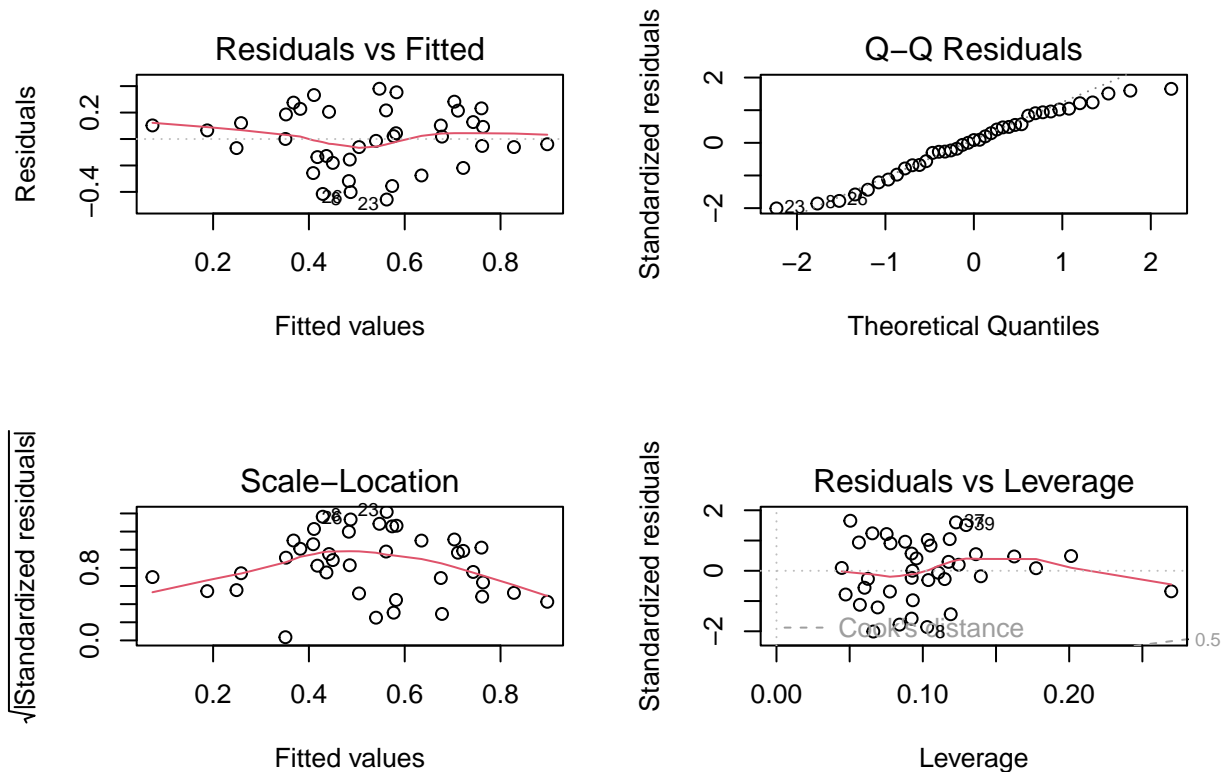
## - total_ground_epa_tot      1    0.10971 1.8853 -106.150
## - total_epa_tot             1    0.24624 2.0218 -103.424
## - pressure_to_sack_rate_tot  1    0.37005 2.1456 -101.106
## - avg_ground_epa_tot        1    0.43448 2.2100 -99.952
##
## Step: AIC=-106.98
## nfl_pct ~ avg_ground_epa_tot + total_ground_epa_last + total_epa_tot +
##   total_ground_epa_tot + pressure_to_sack_rate_tot
##
##               Df Sum of Sq    RSS    AIC
## - total_ground_epa_last      1    0.03011 1.8757 -108.35
## <none>                        1.8456 -106.98
## - total_ground_epa_tot      1    0.09731 1.9429 -106.98
## - total_epa_tot              1    0.17628 2.0219 -105.42
## - pressure_to_sack_rate_tot  1    0.32006 2.1656 -102.74
## - avg_ground_epa_tot        1    0.42047 2.2660 -100.98
##
## Step: AIC=-108.35
## nfl_pct ~ avg_ground_epa_tot + total_epa_tot + total_ground_epa_tot +
##   pressure_to_sack_rate_tot
##
##               Df Sum of Sq    RSS    AIC
## - total_ground_epa_tot      1    0.07096 1.9466 -108.901
## <none>                        1.8757 -108.349
## - total_epa_tot              1    0.21708 2.0928 -106.078
## - pressure_to_sack_rate_tot  1    0.31382 2.1895 -104.316
## - avg_ground_epa_tot        1    0.58283 2.4585 -99.796
##
## Step: AIC=-108.9
## nfl_pct ~ avg_ground_epa_tot + total_epa_tot + pressure_to_sack_rate_tot
##
##               Df Sum of Sq    RSS    AIC
## <none>                        1.9466 -108.901
## - total_epa_tot              1    0.23393 2.1806 -106.475
## - pressure_to_sack_rate_tot  1    0.25496 2.2016 -106.101
## - avg_ground_epa_tot        1    0.68536 2.6320 -99.137
##
## Call:
## lm(formula = nfl_pct ~ avg_ground_epa_tot + total_epa_tot + pressure_to_sack_rate_tot,
##     data = train_data[, c(top_predictors, "nfl_pct")])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45767 -0.14615  0.01819  0.19555  0.37994
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.9172063   0.2007661    4.569 5.86e-05 ***
## avg_ground_epa_tot  0.4497127   0.1281106    3.510  0.00125 **
## total_epa_tot      0.0006793   0.0003312    2.051  0.04783 *
## pressure_to_sack_rate_tot -2.0842937   0.9734921   -2.141  0.03931 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## Residual standard error: 0.2358 on 35 degrees of freedom
## Multiple R-squared:  0.3944, Adjusted R-squared:  0.3425
## F-statistic: 7.597 on 3 and 35 DF,  p-value: 0.000486
```

```
## [1] 0.2600129
```

```
##      avg_ground_epa_tot      total_epa_tot pressure_to_sack_rate_tot
##      1.016623            1.019326            1.022333
```



```
##      8      37      39      26      23      31      32
## 0.09927003 0.08941573 0.08481908 0.07264365 0.07145808 0.06972676 0.06363350
##      30      12      13
## 0.04254930 0.03675492 0.03631633
```

```
##
## studentized Breusch-Pagan test
##
## data: m2b
## BP = 1.2, df = 3, p-value = 0.753
```

```
##
## Shapiro-Wilk normality test
##
## data: m2b$residuals
## W = 0.97001, p-value = 0.3758
```

This model considers predictors with an absolute correlation above 0.30. When optimizing, again with the greedy algorithm, the adjusted r^2 is 0.343 and the error is 0.260. These numbers are both good. The diagnostics pass here, so this is an improvement over the last model. The predictors selected here are average ground EPA, total EPA, and PTSR.

```
##
## Call:
## lm(formula = nfl_pct ~ total_ground_epa_max, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48885 -0.18940 -0.04705  0.18958  0.45828
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.585030   0.045802  12.773 3.92e-15 ***
## total_ground_epa_max 0.006195   0.001981   3.127 0.00343 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2621 on 37 degrees of freedom
## Multiple R-squared:  0.209, Adjusted R-squared:  0.1876
## F-statistic: 9.776 on 1 and 37 DF, p-value: 0.003435

##
## Call:
## lm(formula = nfl_pct ~ pressure_to_sack_rate_tot, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59282 -0.21907 -0.04646  0.22741  0.50258
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8593     0.2012   4.271 0.00013 ***
## pressure_to_sack_rate_tot -1.9394     1.1466  -1.691 0.09915 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.284 on 37 degrees of freedom
## Multiple R-squared:  0.07178, Adjusted R-squared:  0.04669
## F-statistic: 2.861 on 1 and 37 DF, p-value: 0.09915

##
## Call:
## lm(formula = nfl_pct ~ pick, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55571 -0.23922  0.02502  0.19826  0.55014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 0.5744824 0.0601556 9.550 1.59e-11 ***
## pick -0.0005298 0.0004354 -1.217 0.231
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.289 on 37 degrees of freedom
## Multiple R-squared: 0.03847, Adjusted R-squared: 0.01248
## F-statistic: 1.48 on 1 and 37 DF, p-value: 0.2315

## [1] 0.3310179
```

Using the maximum total ground EPA as the only predictor, we get an RMSE of 0.34. The adjusted r^2 is 0.19, which isn't awful. If we instead use PTSR as the only predictor, the r^2 drops to 0.05, but the error improves to 0.31. When using draft pick, both r^2 and RMSE are bad – 0.01 and 0.33 respectively. Considering these factors, I will try to construct a model which has a high adjusted r^2 , low RMSE, and low collinearity error. This will not be perfect, and is not recommended as it will introduce bias, although it will be an interesting exercise.

```
##
## Call:
## lm(formula = nfl_pct ~ pressure_to_sack_rate_tot + avg_ground_epa_tot +
##     adot_first + total_epa_max, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41182 -0.15302  0.03689  0.17617  0.45771
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5431539   0.3808835   1.426  0.16298
## pressure_to_sack_rate_tot -2.3781499   0.9807389  -2.425  0.02078 *
## avg_ground_epa_tot    0.4041735   0.1352739   2.988  0.00519 **
## adot_first      0.0488942   0.0368342   1.327  0.19322
## total_epa_max    0.0005028   0.0005809   0.866  0.39273
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2397 on 34 degrees of freedom
## Multiple R-squared: 0.3924, Adjusted R-squared: 0.3209
## F-statistic: 5.489 on 4 and 34 DF, p-value: 0.00161

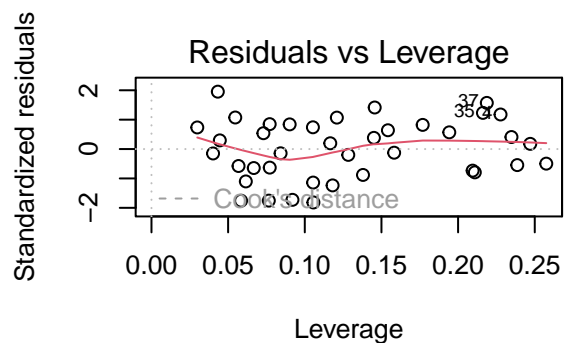
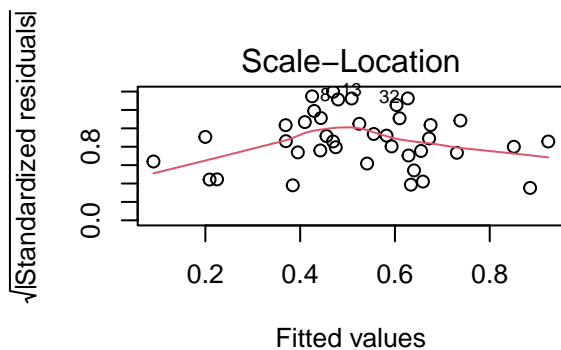
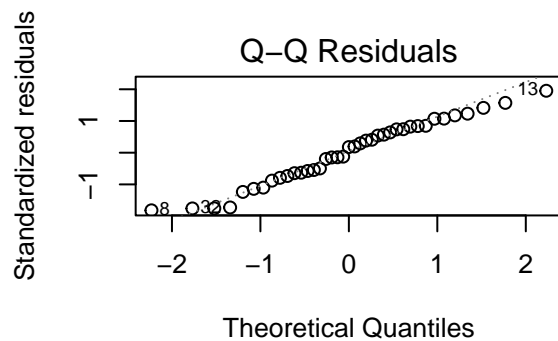
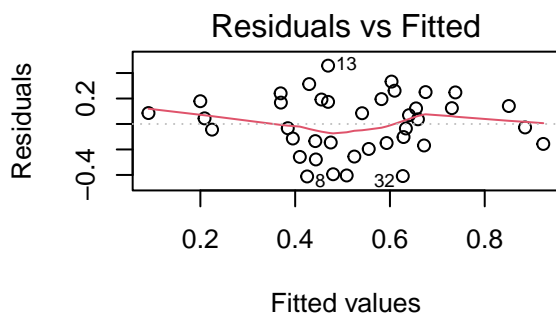
## Start: AIC=-106.77
## nfl_pct ~ pressure_to_sack_rate_tot + avg_ground_epa_tot + adot_first +
##     total_epa_max
##
##              Df Sum of Sq    RSS    AIC
## - total_epa_max      1    0.04305 1.9962 -107.921
## - adot_first         1    0.10122 2.0543 -106.801
## <none>                  1.9531 -106.771
## - pressure_to_sack_rate_tot 1    0.33777 2.2909 -102.550
## - avg_ground_epa_tot      1    0.51281 2.4659  -99.679
##
## Step: AIC=-107.92
```

```
## nfl_pct ~ pressure_to_sack_rate_tot + avg_ground_epa_tot + adot_first
##
##              Df Sum of Sq    RSS    AIC
## <none>                1.9962 -107.921
## - adot_first          1  0.18440  2.1806 -106.475
## - pressure_to_sack_rate_tot 1  0.36437  2.3605 -103.382
## - avg_ground_epa_tot   1  0.58442  2.5806  -99.906

##
## Call:
## lm(formula = nfl_pct ~ pressure_to_sack_rate_tot + avg_ground_epa_tot +
##     adot_first, data = train_data)
##
## Coefficients:
##             (Intercept)  pressure_to_sack_rate_tot
##                0.50834                -2.45884
##      avg_ground_epa_tot                adot_first
##                0.42476                 0.06103

## [1] 0.2400422

## pressure_to_sack_rate_tot      avg_ground_epa_tot      adot_first
##                1.013438                1.056266                1.116849
##                total_epa_max
##                1.119939
```



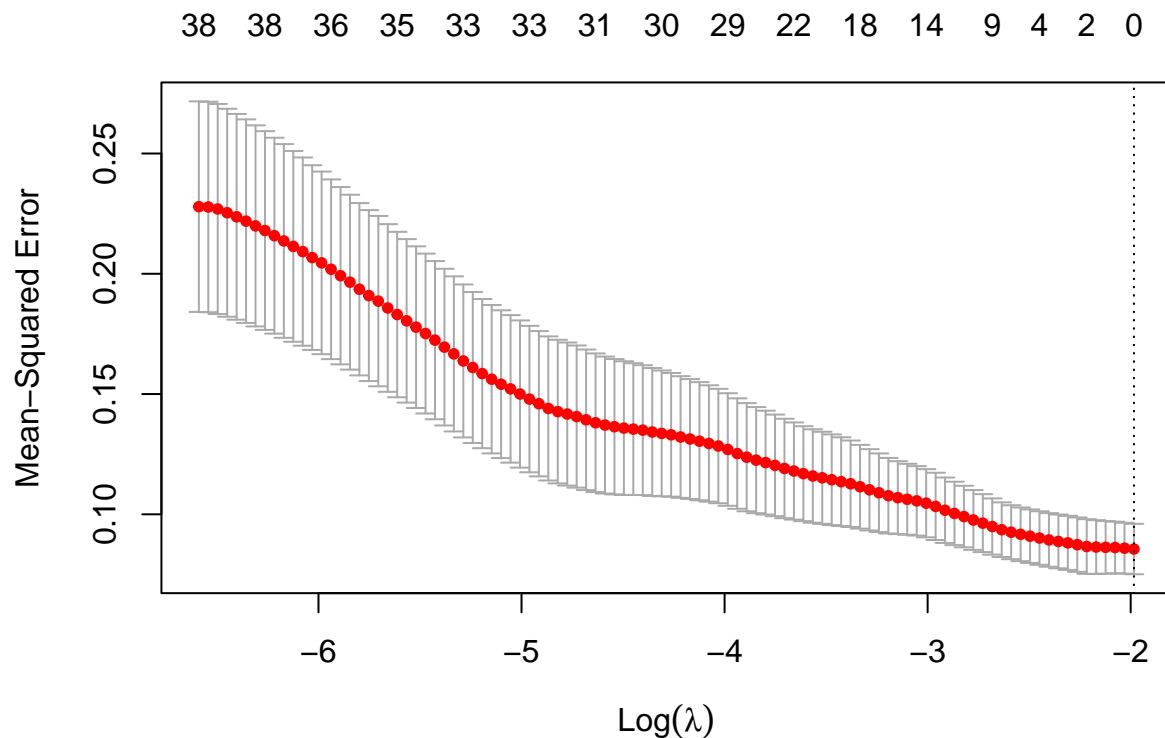

```
##          37          35          4          8          39          26          23
## 0.13812573 0.08354113 0.08157471 0.07797299 0.06811447 0.06043411 0.05102909
##          31          32          13
## 0.04097076 0.03834156 0.03449723

##
## studentized Breusch-Pagan test
##
## data: m3
## BP = 0.37709, df = 4, p-value = 0.9843

##
## Shapiro-Wilk normality test
##
## data: m3$residuals
## W = 0.96811, p-value = 0.3279
```

Again, this process is not ideal, but I looked at the top correlations and did trial and error to optimize r^2 and the prediction error. I ended up with PTSR, average ground EPA, ADOT in first year, and total EPA in best year as the predictors. (Total EPA in best year had a high correlation with round and a better trade off with the comparison metrics.) This model has a adjusted r^2 of 0.321, rmse of 0.240, and low collinearity. If I run the greedy algorithm on this model, it suggests to drop total EPA in best year. This would improve the r^2 to 0.326, but the rmse would worsen to 0.251. Thus, we will keep this predictor considering the better rmse and the tell of overall play, on top of pocket awareness and arm strength.

We will next consider a lasso regression, as not all of these predictors should be included due to their massive collinearity.



```

## [1] 0.1375255

## 194 x 1 sparse Matrix of class "dgCMatrix"
##                               s1
## (Intercept)                0.5277272
## off_grade_avg                .
## pass_grade_avg                .
## run_grade_avg                .
## fum_grade_avg                .
## aimed_passes                 .
## attempts                     .
## dropbacks                    .
## completions                  .
## yards                        .
## tds                          .
## ints                         .
## sacks                        .
## pressures                    .
## scrambles                    .
## btts                         .
## twps                         .
## first_downs                  .
## penalties                    .
## bats                         .
## hats                         .
## tas                          .
## cp                           .
## adj_cp                       .
## ypa_tot                      .
## sack_rate                    .
## pressure_to_sack_rate_tot    .
## pressure_to_scramble_rate_tot .
## scramble_rate_tot           .
## ttt                          .
## btt_rate_tot                 .
## twp_rate_tot                 .
## adot_avg                     .
## nfl_pr                       .
## td_rate_tot                  .
## int_rate_tot                 .
## bat_rate_tot                 .
## hat_rate_tot                 .
## ta_rate_tot                  .
## fd_rate_tot                  .
## pen_rate_tot                 .
## tds_max                      .
## off_grade_max                .
## pass_grade_max                .
## run_grade_max                .
## fum_grade_max                .
## cp_max                       .
## adj_cp_max                   .
## ypa_max                      .
## sack_rate_max                .

```

```

## pressure_to_sack_rate_max      .
## pressure_to_scramble_rate_max  .
## scramble_rate_max              .
## ttt_max                        .
## btt_rate_max                   .
## twp_rate_max                   .
## adot_max                       .
## nfl_pr_max                     .
## td_rate_max                    .
## int_rate_max                   .
## bat_rate_max                   .
## hat_rate_max                   .
## ta_rate_max                    .
## fd_rate_max                    .
## pen_rate_max                   .
## tds_min                        .
## off_grade_min                  .
## pass_grade_min                 .
## run_grade_min                  .
## fum_grade_min                  .
## cp_min                         .
## adj_cp_min                     .
## ypa_min                        .
## sack_rate_min                  .
## pressure_to_sack_rate_min      .
## pressure_to_scramble_rate_min  .
## scramble_rate_min              .
## ttt_min                        .
## btt_rate_min                   .
## twp_rate_min                   .
## adot_min                       .
## nfl_pr_min                     .
## td_rate_min                    .
## int_rate_min                   .
## bat_rate_min                   .
## hat_rate_min                   .
## ta_rate_min                    .
## fd_rate_min                    .
## pen_rate_min                   .
## tds_last                       .
## off_grade_last                 .
## pass_grade_last                .
## run_grade_last                 .
## fum_grade_last                 .
## cp_last                        .
## adj_cp_last                    .
## ypa_last                       .
## sack_rate_last                 .
## pressure_to_sack_rate_last     .
## pressure_to_scramble_rate_last .
## scramble_rate_last             .
## ttt_last                       .
## btt_rate_last                  .
## twp_rate_last                  .

```

```

## adot_last .
## nfl_pr_last .
## td_rate_last .
## int_rate_last .
## bat_rate_last .
## hat_rate_last .
## ta_rate_last .
## fd_rate_last .
## pen_rate_last .
## tds_first .
## off_grade_first .
## pass_grade_first .
## run_grade_first .
## fum_grade_first .
## cp_first .
## adj_cp_first .
## ypa_first .
## sack_rate_first .
## pressure_to_sack_rate_first .
## pressure_to_scramble_rate_first .
## scramble_rate_first .
## ttt_first .
## btt_rate_first .
## twp_rate_first .
## adot_first .
## nfl_pr_first .
## td_rate_first .
## int_rate_first .
## bat_rate_first .
## hat_rate_first .
## ta_rate_first .
## fd_rate_first .
## pen_rate_first .
## mean_epa_tot .
## total_epa_tot .
## avg_pass_epa_tot .
## total_pass_epa_tot .
## avg_rush_epa_tot .
## total_rush_epa_tot .
## avg_sack_epa_tot .
## total_sack_epa_tot .
## avg_ground_epa_tot .
## total_ground_epa_tot .
## passes_tot .
## rushes_tot .
## sacks_tot .
## plays_tot .
## mean_epa_last .
## total_epa_last .
## avg_pass_epa_last .
## total_pass_epa_last .
## avg_rush_epa_last .
## total_rush_epa_last .
## avg_sack_epa_last .

```

```
## total_sack_epa_last      .
## avg_ground_epa_last     .
## total_ground_epa_last   .
## mean_epa_first          .
## total_epa_first         .
## avg_pass_epa_first      .
## total_pass_epa_first    .
## avg_rush_epa_first      .
## total_rush_epa_first    .
## avg_sack_epa_first      .
## total_sack_epa_first    .
## avg_ground_epa_first    .
## total_ground_epa_first  .
## mean_epa_max            .
## total_epa_max           .
## avg_pass_epa_max        .
## total_pass_epa_max      .
## avg_rush_epa_max        .
## total_rush_epa_max      .
## avg_sack_epa_max        .
## total_sack_epa_max      .
## avg_ground_epa_max      .
## total_ground_epa_max    .
## mean_epa_min            .
## total_epa_min           .
## avg_pass_epa_min        .
## total_pass_epa_min      .
## avg_rush_epa_min        .
## total_rush_epa_min      .
## avg_sack_epa_min        .
## total_sack_epa_min      .
## avg_ground_epa_min      .
## total_ground_epa_min    .
## round                   .
## pick                    .
## age                     .
```

```
## [1] 0.3534774
```

After completing Lasso, this tells us that no predictors are worth keeping. The intercept (0.53) is what the model will predict each time. By doing so, the error is 0.353; this is worse than our average model.

```
## [1] 0.3463917
```

```
## [1] 22
```

The prediction error for KNN was optimized at $K = 22$. This led to an error of 0.343, again not great.

```
## # A tibble: 39 x 194
##   off_grade_avg pass_grade_avg run_grade_avg fum_grade_avg aimed_passes
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1      94.6      93.7      84.4      59.5      355
```

```

## 2      85      84.9      58.1      67.9      519
## 3      70.4      69.7      55.5      46.0      1472
## 4      81.8      81.1      63.4      58.5      1388
## 5      75.5      68.0      74.2      48.2      473
## 6      81.8      82.3      56.1      57.6      1165
## 7      90.6      90.6      60.1      58.6      581
## 8      81.1      79.5      63.7      49.4      1090
## 9      73.5      64      76.6      44.0      833
## 10     92.5      92.2      73.1      52.8      540
## # i 29 more rows
## # i 189 more variables: attempts <dbl>, dropbacks <dbl>, completions <dbl>,
## # yards <dbl>, tds <dbl>, ints <dbl>, sacks <dbl>, pressures <dbl>,
## # scrambles <dbl>, btts <dbl>, twps <dbl>, first_downs <dbl>,
## # penalties <dbl>, bats <dbl>, hats <dbl>, tas <dbl>, cp <dbl>, adj_cp <dbl>,
## # ypa_tot <dbl>, sack_rate <dbl>, pressure_to_sack_rate_tot <dbl>,
## # pressure_to_scramble_rate_tot <dbl>, scramble_rate_tot <dbl>, ...

## [1] 0.08622595

## [1] 0.08173585 0.08610857 0.08671414 0.08958390 0.09017767 0.08197296
## [7] 0.08191050 0.08495175 0.08776677 0.08692339 0.08301596 0.08197871
## [13] 0.08326210 0.08490308 0.08364818 0.08726259 0.08714498 0.08644656
## [19] 0.08717517 0.08646343 0.08632336 0.08684853 0.08718196 0.08731608
## [25] 0.08664191

## mtry nodesize
## 1 1 1

## Sample size: 39
## Number of trees: 500
## Forest terminal node size: 1
## Average no. of terminal nodes: 24.884
## No. of variables tried at each split: 1
## Total no. of variables: 193
## Resampling used to grow trees: swor
## Resample size used to grow trees: 25
## Analysis: RF-R
## Family: regr
## Splitting rule: mse *random*
## Number of random split points: 10
## (OOB) R squared: 0.03369837
## (OOB) Requested performance error: 0.08173585

## total_sack_epa_last      total_epa_first
## 1.524357e-03      1.352124e-03
## adot_last      pressure_to_sack_rate_first
## 1.323457e-03      1.289052e-03
## adot_avg      td_rate_max
## 1.280267e-03      1.256655e-03
## avg_ground_epa_tot      avg_ground_epa_max
## 1.169266e-03      1.158770e-03
## total_pass_epa_tot      tds_min
## 1.078048e-03      1.075729e-03

```

##	cp_min	pen_rate_max
##	1.067444e-03	1.020780e-03
##	scramble_rate_last	btt_rate_first
##	1.016429e-03	1.014190e-03
##	pass_grade_first	adot_min
##	9.879583e-04	9.863526e-04
##	cp	total_sack_epa_tot
##	9.835797e-04	9.800672e-04
##	bat_rate_tot	dropbacks
##	9.778945e-04	9.480977e-04
##	ypa_last	bat_rate_last
##	9.218531e-04	9.049825e-04
##	scrambles	fd_rate_tot
##	8.774556e-04	8.372752e-04
##	twp_rate_first	ttt_last
##	8.061905e-04	8.057180e-04
##	avg_sack_epa_min	avg_pass_epa_max
##	7.801248e-04	7.765753e-04
##	hat_rate_max	ypa_min
##	7.723496e-04	7.525044e-04
##	btt_rate_last	pen_rate_last
##	7.503934e-04	7.483034e-04
##	total_pass_epa_first	nfl_pr_last
##	7.480133e-04	7.467610e-04
##	avg_sack_epa_first	mean_epa_max
##	7.459428e-04	7.389558e-04
##	pressure_to_scramble_rate_tot	first_downs
##	7.195607e-04	7.163275e-04
##	pressure_to_sack_rate_last	total_sack_epa_min
##	7.115339e-04	7.101287e-04
##	total_sack_epa_first	total_rush_epa_tot
##	7.054970e-04	6.977701e-04
##	rushes_tot	pass_grade_last
##	6.835604e-04	6.656816e-04
##	int_rate_last	avg_pass_epa_first
##	6.550418e-04	6.516354e-04
##	pressure_to_scramble_rate_max	run_grade_avg
##	6.475234e-04	6.430684e-04
##	avg_rush_epa_max	sack_rate_min
##	6.367877e-04	6.337207e-04
##	fum_grade_min	completions
##	6.288956e-04	6.253546e-04
##	total_ground_epa_min	cp_max
##	6.190545e-04	6.170263e-04
##	total_rush_epa_min	total_ground_epa_last
##	6.157168e-04	6.133516e-04
##	nfl_pr_min	run_grade_min
##	6.131561e-04	6.123749e-04
##	total_pass_epa_min	tds_first
##	6.083507e-04	6.001787e-04
##	scramble_rate_max	ta_rate_last
##	5.956296e-04	5.924083e-04
##	total_sack_epa_max	avg_rush_epa_tot
##	5.902701e-04	5.900626e-04

##	total_epa_last	ints
##	5.861405e-04	5.839225e-04
##	attempts	total_pass_epa_max
##	5.832332e-04	5.813867e-04
##	fum_grade_max	total_rush_epa_last
##	5.794559e-04	5.633037e-04
##	adot_first	pressures
##	5.611717e-04	5.604421e-04
##	pen_rate_min	ttt_max
##	5.590616e-04	5.557558e-04
##	avg_sack_epa_max	penalties
##	5.547449e-04	5.388789e-04
##	sacks	total_ground_epa_max
##	5.349928e-04	5.319073e-04
##	tds	age
##	5.288884e-04	5.282885e-04
##	total_epa_min	ypa_tot
##	5.210917e-04	5.208715e-04
##	pen_rate_first	pass_grade_min
##	5.195182e-04	5.165780e-04
##	cp_last	ta_rate_min
##	5.152400e-04	5.024844e-04
##	pass_grade_avg	aimed_passes
##	5.003683e-04	4.944550e-04
##	avg_ground_epa_last	pen_rate_tot
##	4.915902e-04	4.890885e-04
##	tds_last	tds_max
##	4.878111e-04	4.843031e-04
##	twp_rate_min	adj_cp_last
##	4.752875e-04	4.475229e-04
##	yards	hat_rate_tot
##	4.416736e-04	4.394721e-04
##	off_grade_first	hat_rate_first
##	4.350095e-04	4.283682e-04
##	total_epa_tot	total_epa_max
##	4.247188e-04	4.241671e-04
##	btt_rate_max	int_rate_max
##	4.202703e-04	4.191436e-04
##	run_grade_first	total_ground_epa_tot
##	4.133742e-04	4.114017e-04
##	hat_rate_min	pressure_to_scramble_rate_last
##	4.044171e-04	4.021386e-04
##	int_rate_first	adj_cp_min
##	4.016215e-04	3.796175e-04
##	td_rate_tot	total_ground_epa_first
##	3.770441e-04	3.700776e-04
##	td_rate_first	avg_pass_epa_tot
##	3.680694e-04	3.630762e-04
##	fum_grade_avg	total_rush_epa_first
##	3.608970e-04	3.574689e-04
##	cp_first	ttt_first
##	3.566897e-04	3.555987e-04
##	sack_rate	fd_rate_max
##	3.472539e-04	3.359386e-04

##	off_grade_min	adj_cp_first
##	3.299468e-04	3.143140e-04
##	sack_rate_last	td_rate_last
##	3.113701e-04	3.103133e-04
##	btt_rate_tot	off_grade_max
##	3.037268e-04	3.012977e-04
##	twps	ypa_first
##	2.928840e-04	2.876416e-04
##	ta_rate_tot	bat_rate_first
##	2.835148e-04	2.827780e-04
##	twp_rate_last	hat_rate_last
##	2.785459e-04	2.781207e-04
##	pick	twp_rate_max
##	2.767676e-04	2.743358e-04
##	passes_tot	bat_rate_min
##	2.673677e-04	2.614028e-04
##	avg_rush_epa_first	pressure_to_sack_rate_tot
##	2.604811e-04	2.544412e-04
##	avg_pass_epa_min	td_rate_min
##	2.515788e-04	2.491558e-04
##	mean_epa_first	nfl_pr_first
##	2.475007e-04	2.446419e-04
##	scramble_rate_min	adot_max
##	2.427420e-04	2.332449e-04
##	avg_ground_epa_min	tas
##	2.283837e-04	2.272139e-04
##	fd_rate_min	scramble_rate_tot
##	2.259726e-04	2.255064e-04
##	bats	ta_rate_first
##	2.190620e-04	2.158609e-04
##	sacks_tot	pressure_to_sack_rate_min
##	2.150909e-04	2.141063e-04
##	avg_rush_epa_last	scramble_rate_first
##	2.111288e-04	2.051185e-04
##	fd_rate_last	plays_tot
##	1.991336e-04	1.972411e-04
##	bat_rate_max	fd_rate_first
##	1.920947e-04	1.883193e-04
##	off_grade_last	avg_sack_epa_tot
##	1.873064e-04	1.669650e-04
##	adj_cp_max	hats
##	1.557687e-04	1.524851e-04
##	mean_epa_last	ta_rate_max
##	1.367299e-04	1.346260e-04
##	nfl_pr_max	avg_rush_epa_min
##	1.316848e-04	1.309696e-04
##	avg_sack_epa_last	int_rate_min
##	1.230885e-04	1.152535e-04
##	twp_rate_tot	btts
##	1.150857e-04	1.139404e-04
##	nfl_pr	run_grade_last
##	1.116646e-04	1.017314e-04
##	sack_rate_first	ypa_max
##	9.785280e-05	9.611597e-05

```
##          btt_rate_min          ttt
##          9.341885e-05          9.293880e-05
##          total_rush_epa_max      pass_grade_max
##          8.082529e-05          7.901406e-05
##          sack_rate_max          adj_cp
##          7.811438e-05          7.426476e-05
##          off_grade_avg          int_rate_tot
##          6.657252e-05          6.104986e-05
##          pressure_to_sack_rate_max      fum_grade_last
##          4.927213e-05          4.746246e-05
##          mean_epa_tot          ttt_min
##          4.422635e-05          4.412032e-05
##          pressure_to_scramble_rate_min      avg_pass_epa_last
##          4.109822e-05          3.265985e-05
##          avg_ground_epa_first      mean_epa_min
##          3.251432e-05          2.690330e-05
##          pressure_to_scramble_rate_first      run_grade_max
##          2.494397e-05          1.957082e-05
##          fum_grade_first      total_pass_epa_last
##          1.518240e-05          1.454270e-05
##          round
##          8.638386e-06
```

```
## [1] 0.3843768
```

After tuning on mtry and nodesize, the selected model has an r^2 of 0.031, with an OOB rate of 0.082. The rmse is bad too at 0.386. So, while this model won't be selected, the top variables for this model include EPA totals, ADOT, and completion percentage (CP).

```
## [1] 0.422049
```

```
##      best_eta best_max_depth best_ntrees best_error
## 1      0.1           2           2 0.3676351
## 2      0.5           2           2 0.3968052
## 3      1.0           2           2 0.4551555
## 4      0.1           5           2 0.3661133
## 5      0.5           5           2 0.3929162
## 6      1.0           5           2 0.4553070
## 7      0.1          10           2 0.3661133
## 8      0.5          10           2 0.3929162
## 9      1.0          10           2 0.4553070
```

```
##      best_eta best_max_depth best_ntrees best_error
## 4      0.1           5           2 0.3661133
```

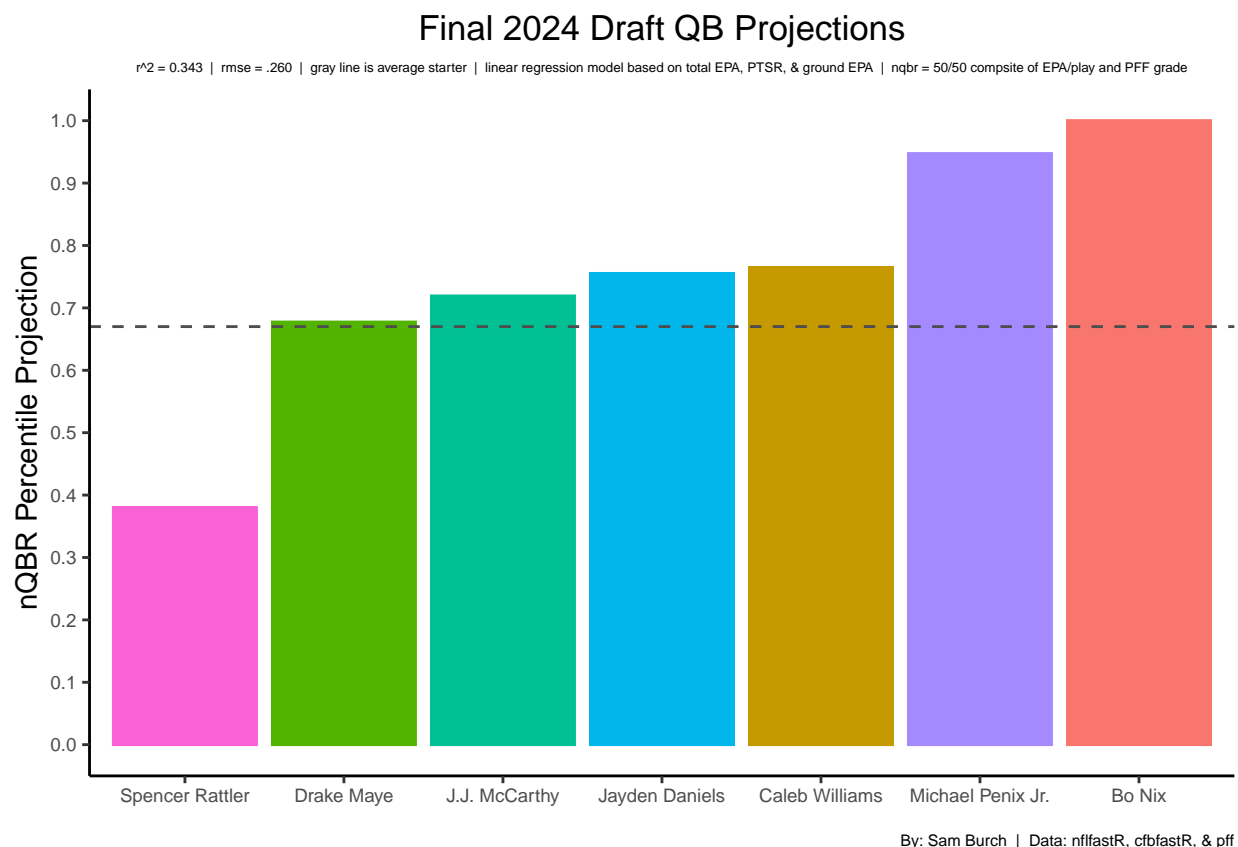
The best prediction error for XGBoost also does not perform with an error of 0.366. The optimal tuning here was 0.1 ETA, 5 max depth, and 2 ntrees.

Across the board, most models perform very similarly. Looking at adjusted r^2 , the first linear regression (which included all predictors with an abs. correlation of 0.275 or higher) is by far the best at 0.466. However, there are certainly overfitting issues as the model does not pass the collinearity diagnostics. Next tier here is the second and third linear regression respectively. These have good r^2 s, but not great at around a third.

For prediction error, the baseline is 0.348 for the average. Five models did perform better than this. The best here was the first linear regression, however note the collinearity issues. Next comes the other linear regression models, with this time the third performing slightly better. With the testing sample size being so small, this rmse statistic should be taken with a grain of salt though.

Therefore, we will select the second linear regression model, with hesitancy. This model has the second highest adjusted r^2 and the third lowest prediction error. The first LR model would have been chosen, but it did not pass the collinearity diagnostics. Meanwhile, the third linear regression model does have a better rmse, however the more stable stat (r^2) is worse than the second LR. On top of this, bias was introduced when modeling the third linear regression. The runners up are the average, the best predictor, and the KNN model. The best predictor and KNN model each had a slightly better prediction error than the average. Although that is the case, the errors are all close enough that the average is probably the best of the runner ups. The advanced models all didn't perform well. This can possibly be improved by transforming the predictors into principal components. However, it is not necessary as the advanced models all pass the diagnostics by tuning them.

Let's see how the selected model projects the 2024 NFL draft prospects!



Takeaways from projections

Nix and Penix are in their own tier due to their elite pocket awareness and experience. The issue with both is their consensus rankings have them as fringe first rounders. Considering this, if they fall out of the first round, they shouldn't necessarily be looked at as values. The top-4 on the consensus board are next at slightly above average. McCarthy is hurt by his limited number of plays, same with Williams and Maye, but Daniels is helped by it. Take on the fact that all have had elite seasons makes them good prospects. The biggest issue is Daniels needs a good surrounding as his PTSR is concerning low. Daniels' experience and rushing should help him though. Williams and Maye weren't elite in pocket awareness in college, but their elite production make them elite prospects regardless. McCarthy was great across the board, but it was on

a small sample, as stated earlier. On the other hand, Rattler projects poorly being worst in every category but PTSR.

Some late round QBs are interesting. Jordan Travis has the third highest projection (77th percentile) out of the 11 prospects considered, due to his experience and good (not great) pocket awareness. The other three (Hartman, Pratt, and Milton) all project as below average starter or backup. Hartman and Pratt are helped by their production & experience, but hurt by their pocket awareness, which is concerning. Milton might be worth a flyer with his strong arm and decent pocket awareness.

Conclusion

Next steps include using a larger sample, implementing principal components into the models, and adding even more predictors. The first is by far the most important, as having a larger sample size will almost certainly create a better model than using the average. Implementing PCs will help the advanced models perform better. More predictors, like performance in a clean pocket, on play action plays, or percentage of pressures assigned to the QB will likely have some significance. Regardless, this analysis is a big step forward from last year's and provides us with a more concrete projection of the QB prospects. One continued theme is the stability of pocket awareness. Metrics like PTSR and ground EPA need to be considered more when ranking prospects.

Thanks for reading!

OLD / Unnecessary

first aspect of building the model is determining what NFL success is. Ben Baldwin stated “[about] 95% of [how good a QB is can be determined from] looking at EPA (expected points added) and PFF (Pro Football Focus) grade.” So, we will use this as our indicator of QB play in the NFL.

PBP

PFF

Full NFL Dataset

PFF

Joining Datasets

```
# pca = prcomp(df_final_no_name |> dplyr::select(-nfl_pct), scale. = TRUE)
#
# new_x = pca$x[, 1:20]
#
# train_x = new_x[train_ind, ]
# test_x = new_x[-train_ind, ]
```

K-Means

```
set.seed(123)
```

```
kmean_result = kmeans(train_data |> select(-c(nfl_pct)), 3)
```

```
kmean_result$withinss
```

```
train_data_kmean = train_data |>  
  mutate(nfl_pct_cat = case_when((nfl_pct <= .6) ~ 1,  
                                  (nfl_pct > .6 & nfl_pct <= .85) ~ 2,  
                                  (nfl_pct > .85 ~ 3)))
```

```
# kmean_result$cluster
```

```
train_data_kmean$cluster = as.factor(kmean_result$cluster)
```

```
table(train_data_kmean$cluster, train_data_kmean$nfl_pct_cat)
```

```
### Ridge (Not as useful)
```

```
# library(caret)
```

```
# library(MASS)
```

```
# library(glmnet)
```

```
#
```

```
#
```

```
# set.seed(123)
```

```
#
```

```
# x_train = as.matrix(train_data |> select(-nfl_pct)) # Standardize predictor variables for training s
```

```
# y_train = as.matrix(train_data$nfl_pct)
```

```
#
```

```
# x_test = as.matrix(test_data |> select(-nfl_pct)) # Standardize predictor variables for testing set
```

```
# y_test = as.matrix(test_data$nfl_pct)
```

```
# ridge_model = cv.glmnet(x_train, y_train, alpha = 0)
```

```
#
```

```
# plot(ridge_model)
```

```
#
```

```
# best_lambda = ridge_model$lambda.min
```

```
#
```

```
# final_model = glmnet(x_train, y_train, alpha = 0, lambda = best_lambda)
```

```
#
```

```
# sort(abs(final_model$beta), decreasing = TRUE)
```

```
# final_model$beta
```

```
#
```

```
# pred = predict(final_model, newx = x_test)
```

```
#
```

```
# sqrt(mean((pred - y_test)^2))
```