An Analysis of the NBA's Change in Lineups

Max Sleek

2023-01-15

Basketball is one of the most lineup-dependent team sports in the world since each team can only have 5 players on the court at once. When working with basketball analytics, lineup data is crucial to understanding what makes a basketball team successful, and what players to keep on the court. Consequently, I have centered this project around deep lineup analytics in the NBA. I look to provide a rigorous model for the most important factors in predicting lineup success while also exploring and visualizing how the definition of success has changed over the past 19 NBA seasons. However, before I started, I had to first decide how to measure the success of a particular lineup. I chose to use efficiency differential, which is a lineup's offensive points per possession minus its defensive points allowed per possession. The database with the most widely available data around efficiency differential was CleaningTheGlass.com. This website was created by Ben Falk, a notable statistician and data scientist who previously worked with several NBA teams. The data itself was collected by sportradar, which is renowned for its large databases of major sports. This lineup data is available for every year of NBA basketball since 2004 and includes several divisions of basketball statistics including efficiency differential (points scored per offensive possession - points allowed per defensive possession), shooting efficiency, shooting locations, halfcourt offense, transition/fastbreaks, and more – for both offense and defense. As for the observations in my dataset, each row represents a single lineup that played at least 100 possessions. In total, there are 6,284 lineups since 2004 that met this requirement. Here is an example of some of the variables I used for lineup data (To view the dull data set and all 114 variables, see my scripts):

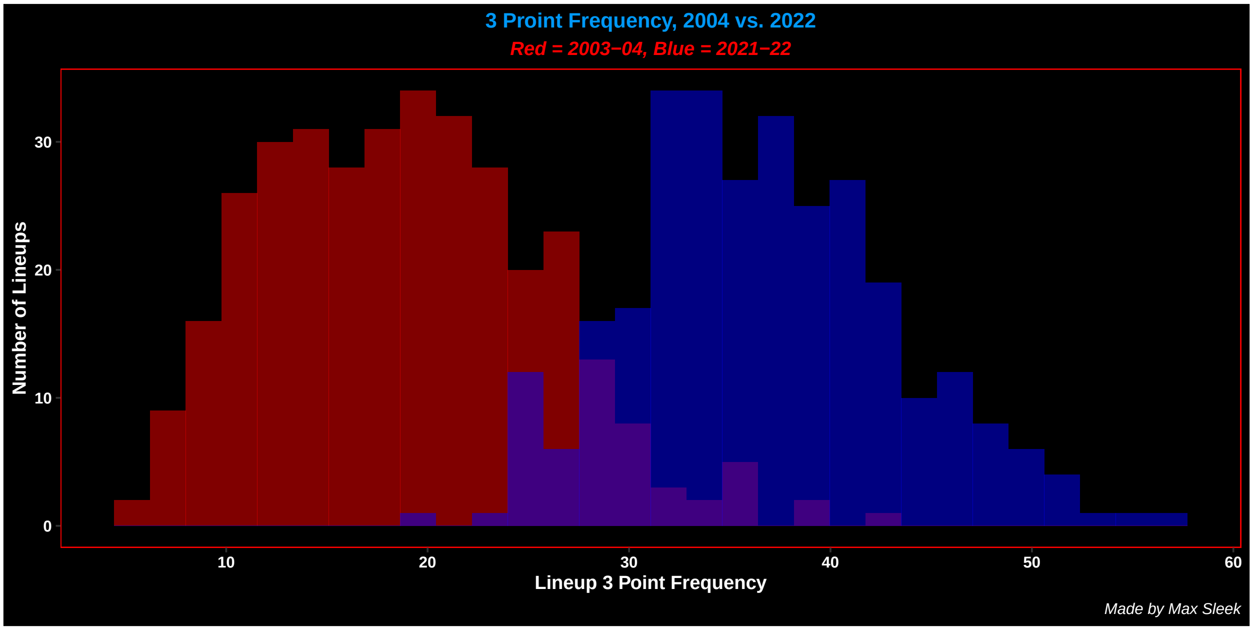
```
## # A tibble: 6 x 5
##   PG                 Poss  Diff OFFENSE..Pts.Poss OFFENSE..eFG.
##   <chr>             <dbl> <dbl>             <dbl>         <dbl>
## 1 Mike Bibby         2097  13.9             117.           56.8
## 2 Jason Kidd         1923   7.9             101.           47.7
## 3 Andre Miller       1901   8.2             106.           49.6
## 4 Steve Francis      1792  10.1             104.           50.7
## 5 Baron Davis        1310  -0.9              97.3          44.6
## 6 Chauncey Billups   1289   9.5             102.           45.2
```

Cleaning The Glass includes the names of all 5 players in the lineup, but doesn't include any biographical information about them (age, height, experience, etc). To remedy this, I scraped the players' biographical information from NBA.com and ran a match function to combine the two datasets. Please look at my lineup formations document to see how I formed my data set, cleaned it, and prepared it for analysis; This was a large part of my project and showcases a lot of my R skills. Once I accomplished this, my data for each player in each lineup was as follows:

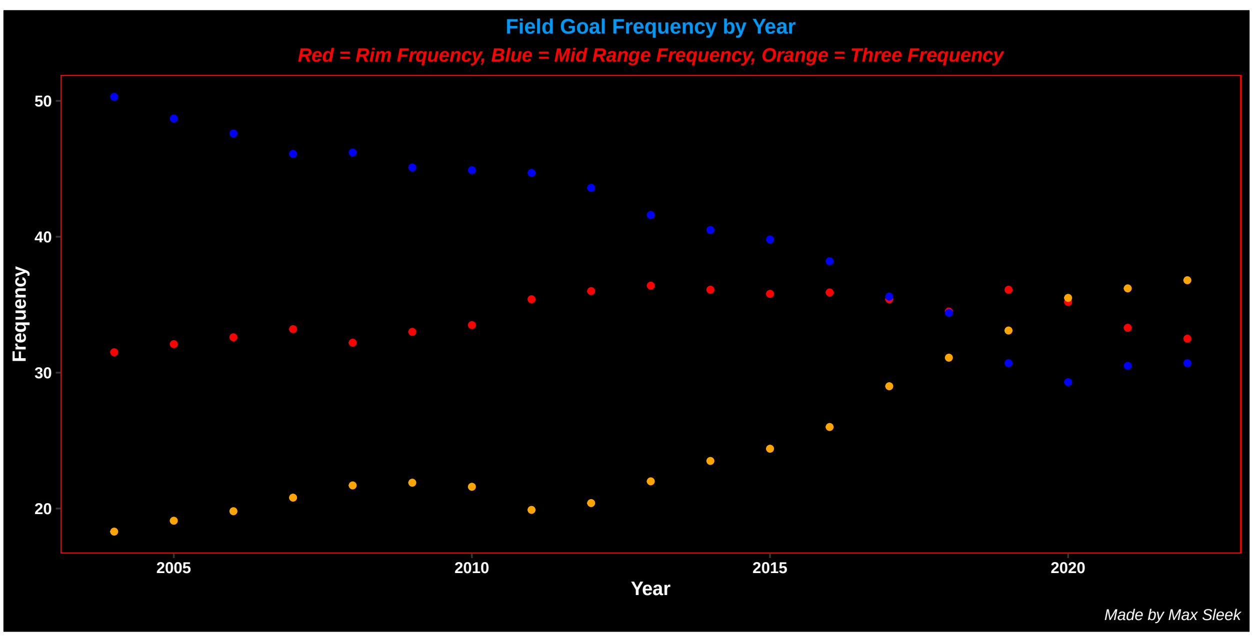
```
## # A tibble: 6 x 8
##   PG                PG.Height PG.Age PG.Experience PG.GP PG.Dra~1 PG.Dr~2 PG.Na~3
##   <chr>                 <dbl>  <dbl>         <dbl> <dbl> <chr>      <dbl> <chr>
## 1 Mike Bibby               74     25             5    82 1998           2 United~
## 2 Jason Kidd               76     30             9    67 1994           2 United~
## 3 Andre Miller             75     27             4    82 1999           8 United~
## 4 Steve Francis            75     27             4    79 1999           2 United~
## 5 Baron Davis              75     24             4    67 1999           3 United~
## 6 Chauncey Billups         75     27             6    78 1997           3 United~
## # ... with abbreviated variable names 1: PG.Draft.Year, 2: PG.Draft.Pick,
## #   3: PG.Nationality
```

I began with evaluating the changes in the league. It is virtually common knowledge that the NBA is moving away from the mid-range shot and moving toward the three-point shot. When we compare 2004 and 2022, we can see this massive increase in frequency.

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```

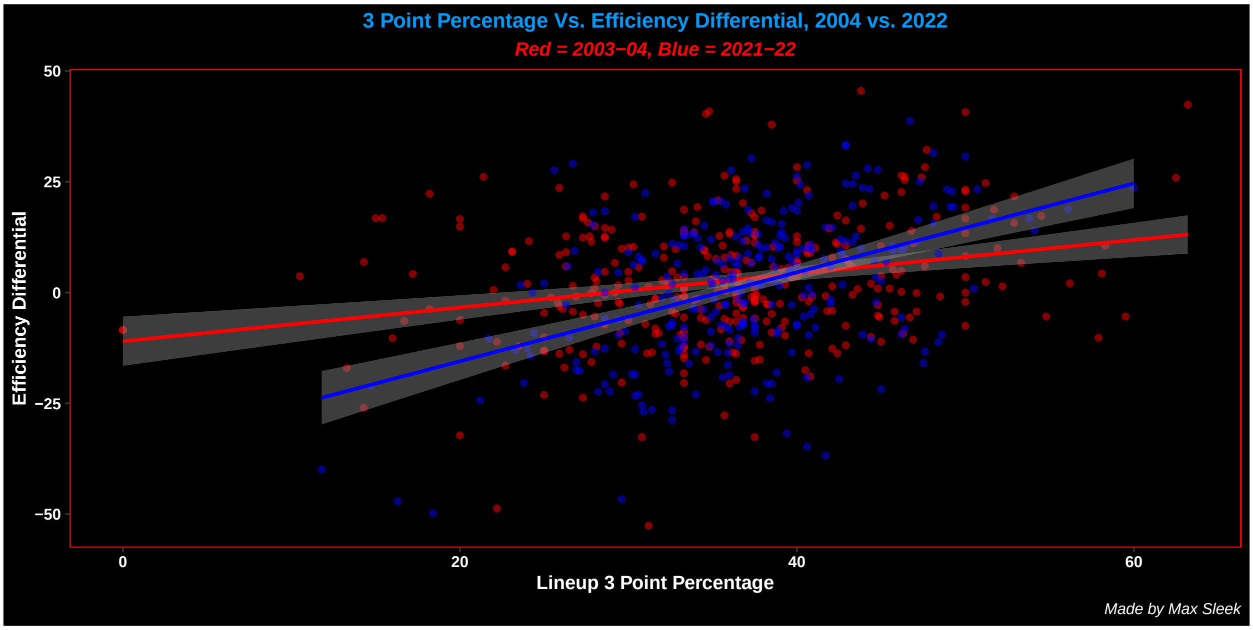


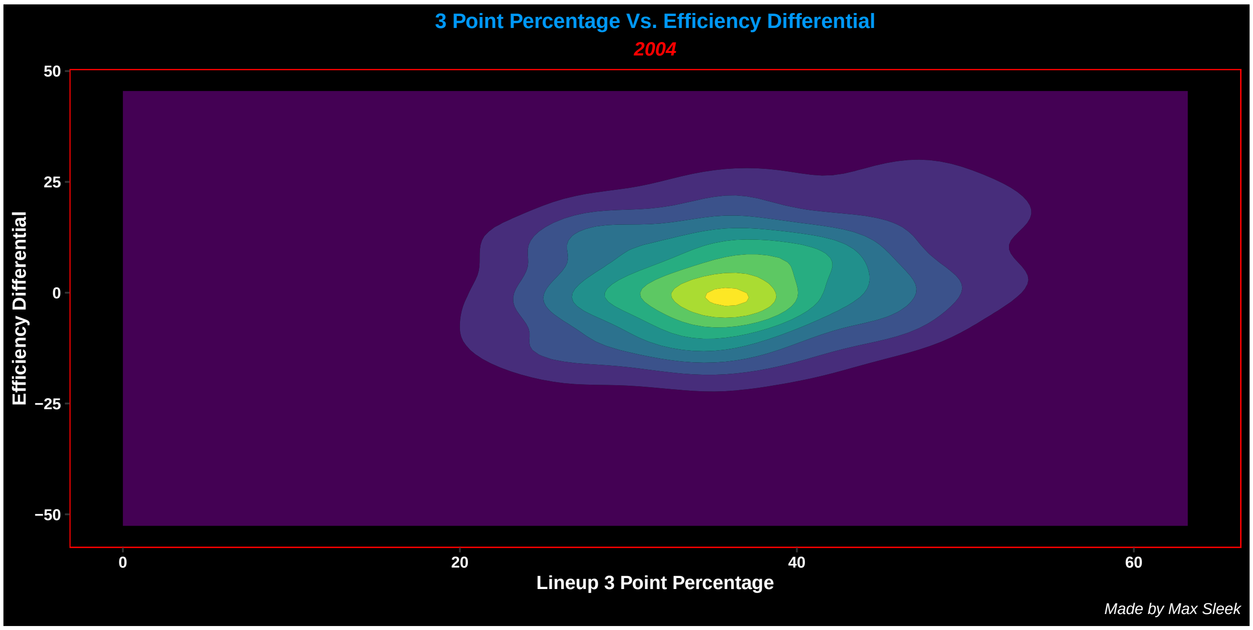
However, I wanted to visualize the change in distribution over the past 19 seasons. As seen below, it appears the frequency of mid-range shots in lineups was steadily decreasing after 2004, but the increase in 3-point frequency did not begin its soar until around 2014 (coincidentally, the beginning of Stephen Curry's influence).

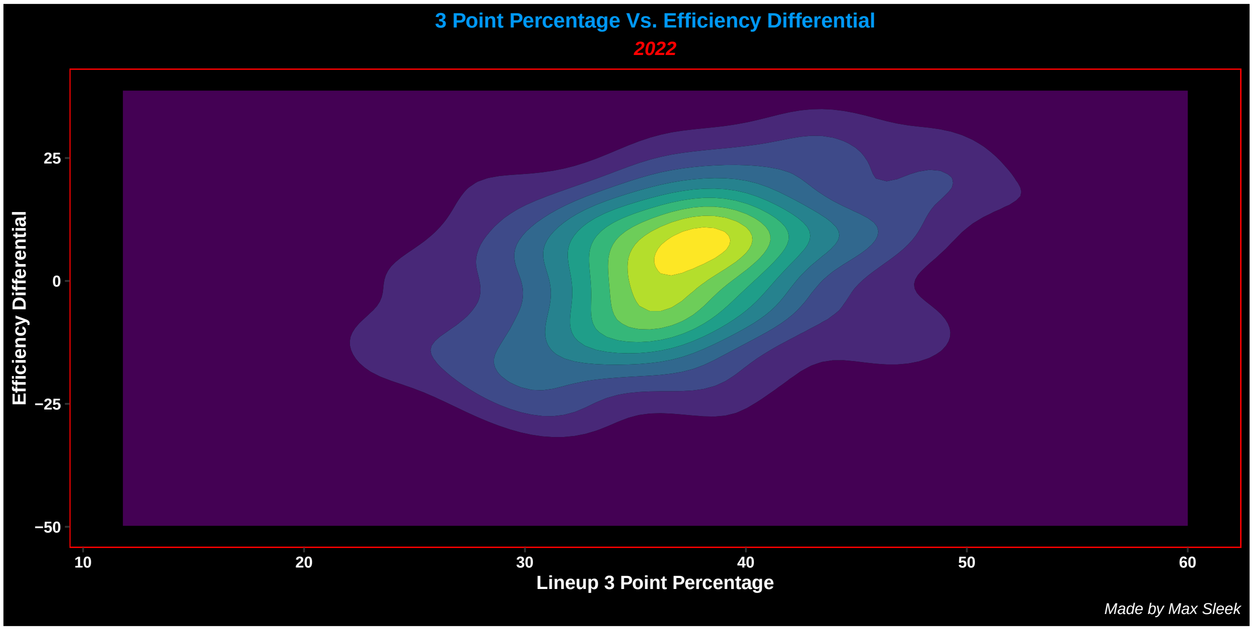


This is only a surface-level analysis though. Firstly, has the change in the distribution of shots impacted the efficiency differential of lineups throughout the past 19 seasons? To visualize this, I plotted the 3-point percentage by the efficiency differential of all lineups from 2004 and 2022. As seen below, there is a much clearer correlation between shooting percentage and the success of lineups in 2022 than in 2004. In so many words, teams that shoot well in 2022 are much more likely to have a high-efficiency differential (and vice versa). In 2004, lineups that were not centered around shooting threes could have much more success than they would today. This is clearly represented by the heatmaps I plotted comparing 2004 and 2022.

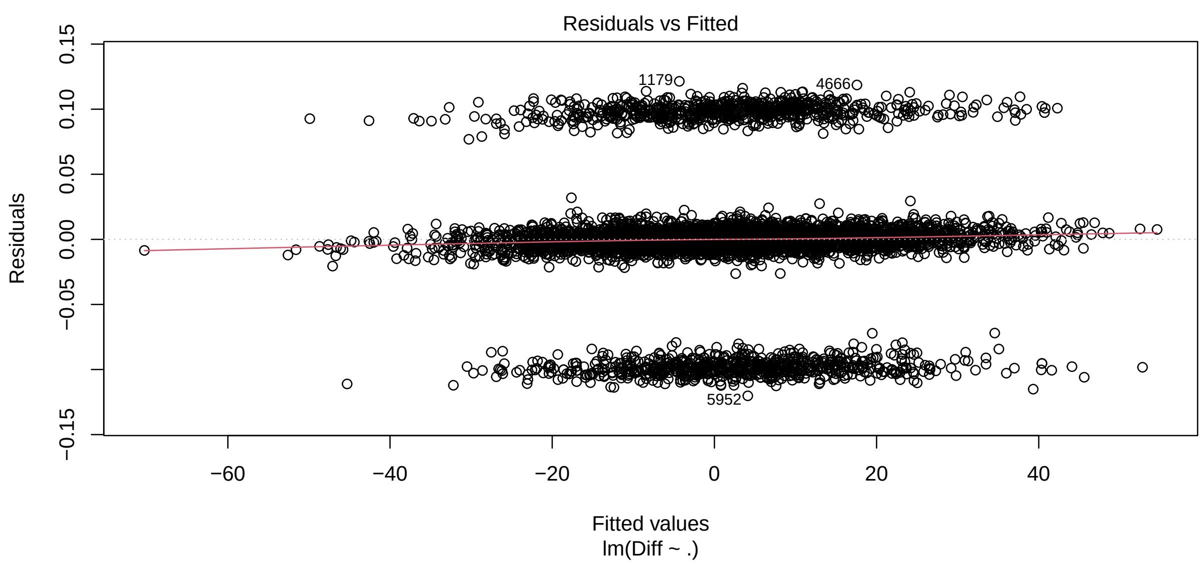
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

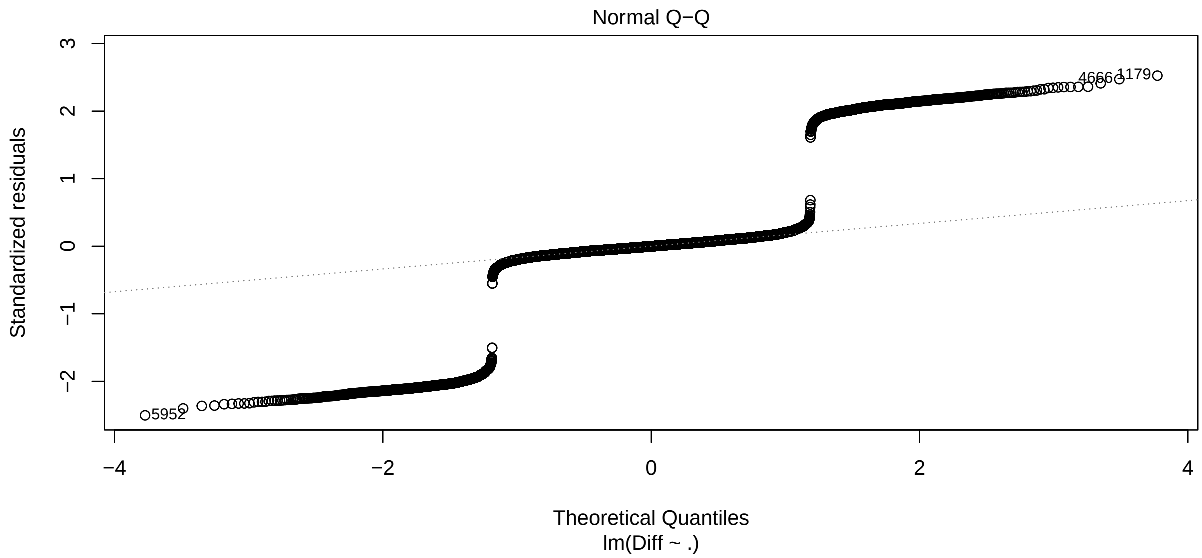






Shot distribution is only one of the ways the NBA has evolved over the past 20 years (Note that I have done research on these "other" changes, but chose not to include them in this paper for sake of length). How do these evolutions impact how we predict the success of a lineup today? To answer this question, I compared models created using all lineup data vs. only recent lineup data. I decided to use stepwise linear regression to create these models. Before creating a full model, I removed variables that would have no bearing on the prediction of efficiency differential. These included player names, teams, heights, draft year, games played, draft position, and nationality. This full model had very few significant predictors and a horrific Normal Q-Q Plot, as was to be expected.



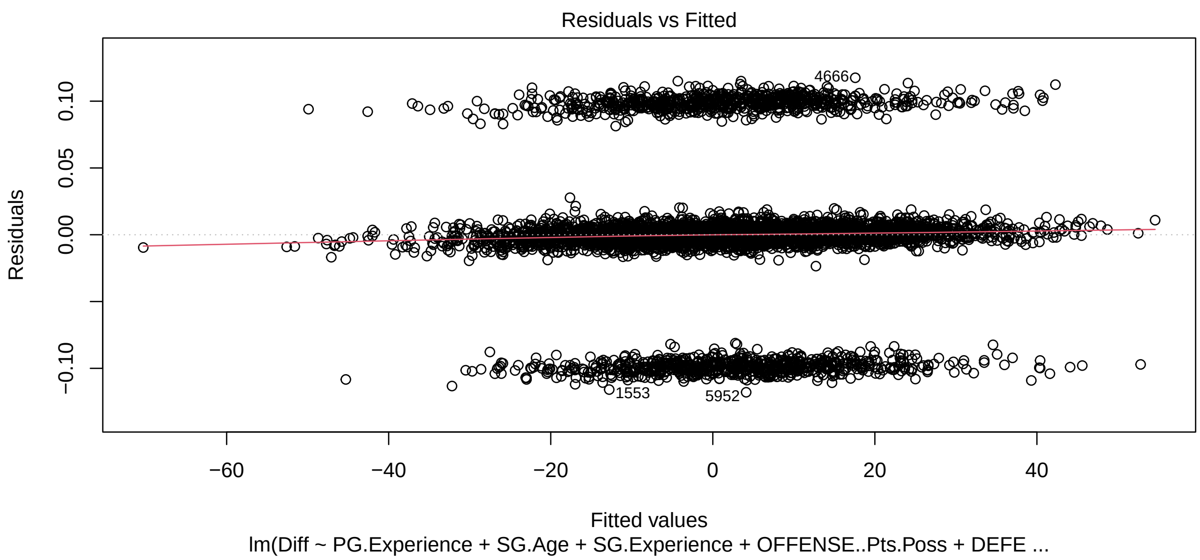


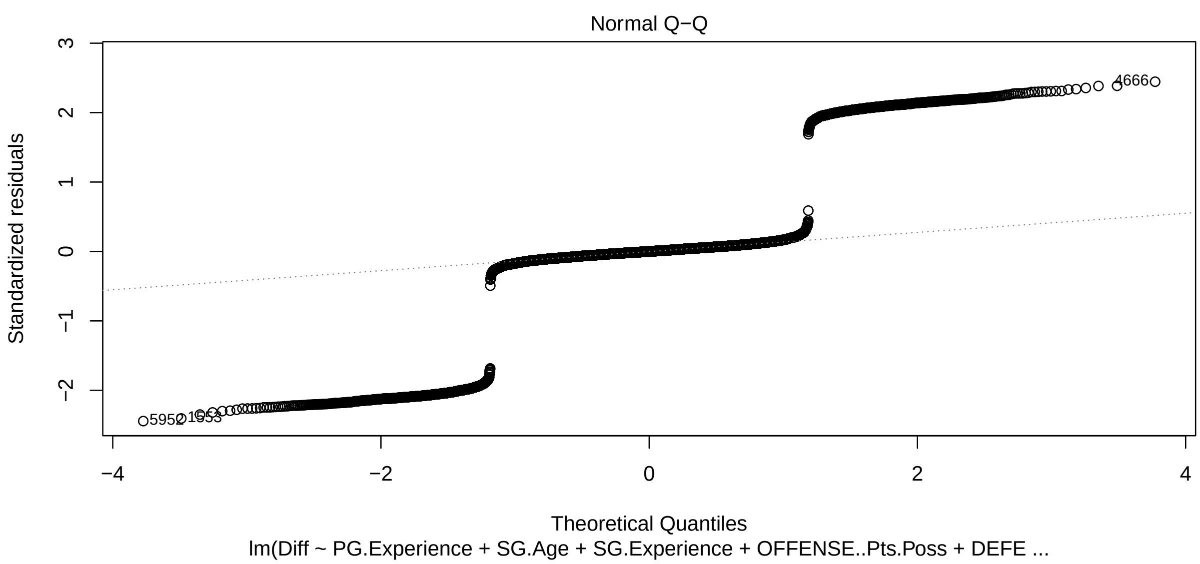
So, I ran stepwise regression (in both directions) on this full model to establish a base model. This regression selected 23 predictors, each of which was significant or close to significant at the .01 level.

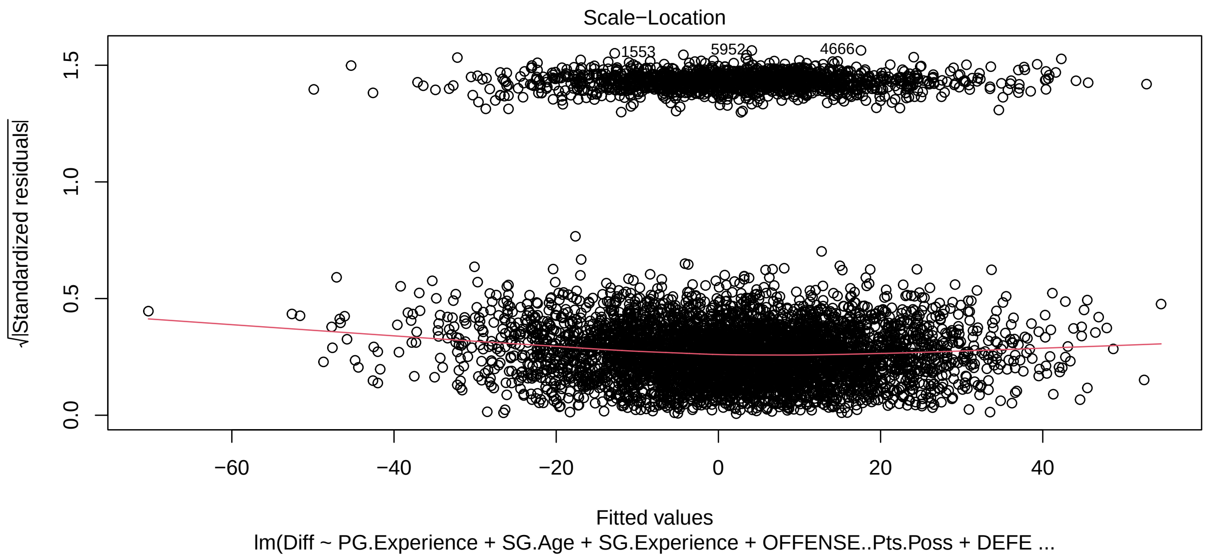
```
##
## Call:
## lm(formula = Diff ~ PG.Experience + SG.Age + SG.Experience +
##     OFFENSE..Pts.Poss + DEFENSE..Pts.Poss + DEFENSE..eFG. + DEFENSE..ORB. +
##     OFFENSE..Corner.Three.Frequency + OFFENSE..Rim.FG. + OFFENSE..Corner.Three.FG. +
##     OFFENSE..Non.Corner.Three.FG. + OFFENSE..All.Three.FG. +
##     PUTBACKS..Pts.Play + OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Pts.Play +
##     DEFENSE..Rim.Frequency + DEFENSE..Short.Mid.Frequency + DEFENSE..Long.Mid.Frequency +
##     DEFENSE..All.Mid.Frequency + DEFENSE..All.Mid.FG. + DEFENSE..All.Three.FG. +
##     DEFENSE.PUTBACKS..Plays.Miss + DEFENSE.TRANSITION..Pts..Poss +
##     DEFENSE.OFF.STEALS.TRANSITION..Freq, data = full$model)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.117935 -0.004549 -0.000023  0.004428  0.117526
##
## Coefficients:
##                                                 Estimate Std. Error   t value
## (Intercept)                                    2.931e-02  1.677e-02     1.748
## PG.Experience                                 -3.390e-04  1.660e-04    -2.042
## SG.Age                                        -8.278e-04  3.795e-04    -2.181
## SG.Experience                                  1.058e-03  4.021e-04     2.632
## OFFENSE..Pts.Poss                              9.998e-01  7.965e-05 12552.472
## DEFENSE..Pts.Poss                             -9.998e-01  1.299e-04 -7698.702
## DEFENSE..eFG.                                 -8.102e-04  2.972e-04    -2.726
## DEFENSE..ORB.                                 -3.994e-04  1.723e-04    -2.319
## OFFENSE..Corner.Three.Frequency                3.645e-04  2.050e-04     1.778
## OFFENSE..Rim.FG.                               2.388e-04  8.867e-05     2.693
## OFFENSE..Corner.Three.FG.                     -2.320e-04  6.858e-05    -3.382
## OFFENSE..Non.Corner.Three.FG.                 -5.275e-04  1.902e-04    -2.774
## OFFENSE..All.Three.FG.                         7.564e-04  2.639e-04     2.866
## PUTBACKS..Pts.Play                             3.087e-05  1.925e-05     1.604
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Pts.Play  2.903e-05  1.973e-05     1.471
## DEFENSE..Rim.Frequency                        -2.923e-04  1.470e-04    -1.988
## DEFENSE..Short.Mid.Frequency                   1.959e-02  1.237e-02     1.583
## DEFENSE..Long.Mid.Frequency                    1.962e-02  1.237e-02     1.586
## DEFENSE..All.Mid.Frequency                    -1.968e-02  1.237e-02    -1.591
## DEFENSE..All.Mid.FG.                           1.977e-04  1.325e-04     1.491
## DEFENSE..All.Three.FG.                         2.062e-04  1.112e-04     1.853
## DEFENSE.PUTBACKS..Plays.Miss                   4.811e-04  1.846e-04     2.605
## DEFENSE.TRANSITION..Pts..Poss                  5.215e-04  2.094e-04     2.490
## DEFENSE.OFF.STEALS.TRANSITION..Freq            7.864e-05  4.309e-05     1.825
##                                               Pr(>|t|)
## (Intercept)                                   0.080535 .
## PG.Experience                                 0.041213 *
## SG.Age                                        0.029209 *
## SG.Experience                                 0.008521 **
## OFFENSE..Pts.Poss                              < 2e-16 ***
## DEFENSE..Pts.Poss                              < 2e-16 ***
## DEFENSE..eFG.                                 0.006424 **
## DEFENSE..ORB.                                 0.020437 *
## OFFENSE..Corner.Three.Frequency               0.075441 .
## OFFENSE..Rim.FG.                              0.007100 **
## OFFENSE..Corner.Three.FG.                     0.000723 ***
## OFFENSE..Non.Corner.Three.FG.                 0.005558 **
## OFFENSE..All.Three.FG.                        0.004167 **
## PUTBACKS..Pts.Play                            0.108785
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Pts.Play 0.141289
## DEFENSE..Rim.Frequency                        0.046838 *
## DEFENSE..Short.Mid.Frequency                  0.113422
## DEFENSE..Long.Mid.Frequency                   0.112813
## DEFENSE..All.Mid.Frequency                    0.111598
```

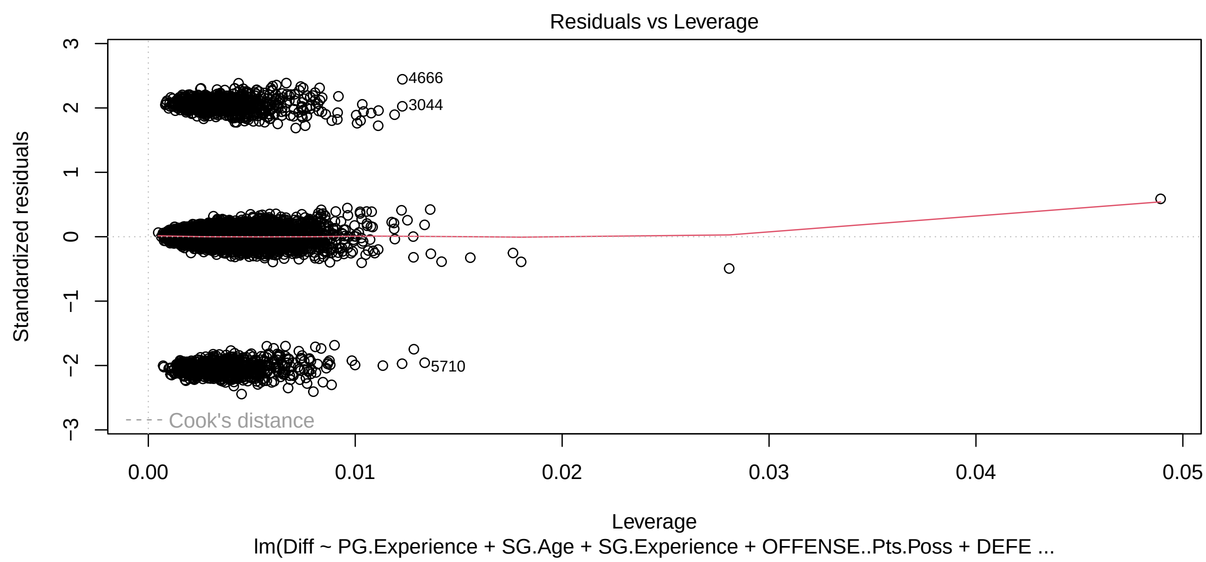
```
## DEFENSE..All.Mid.FG.                            0.135969
## DEFENSE..All.Three.FG.                          0.063868 .
## DEFENSE.PUTBACKS..Plays.Miss                    0.009203 **
## DEFENSE.TRANSITION..Pts..Poss                   0.012790 *
## DEFENSE.OFF.STEALS.TRANSITION..Freq             0.068055 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04837 on 6163 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 2.275e+07 on 23 and 6163 DF,  p-value: < 2.2e-16
```

However, the Normal Q-Q and residuals vs. fitted plots still had not improved. Additionally, there were alarming variance inflation factors for all frequency-related statistics. This was likely due to multicollinearity or non-essential columns.





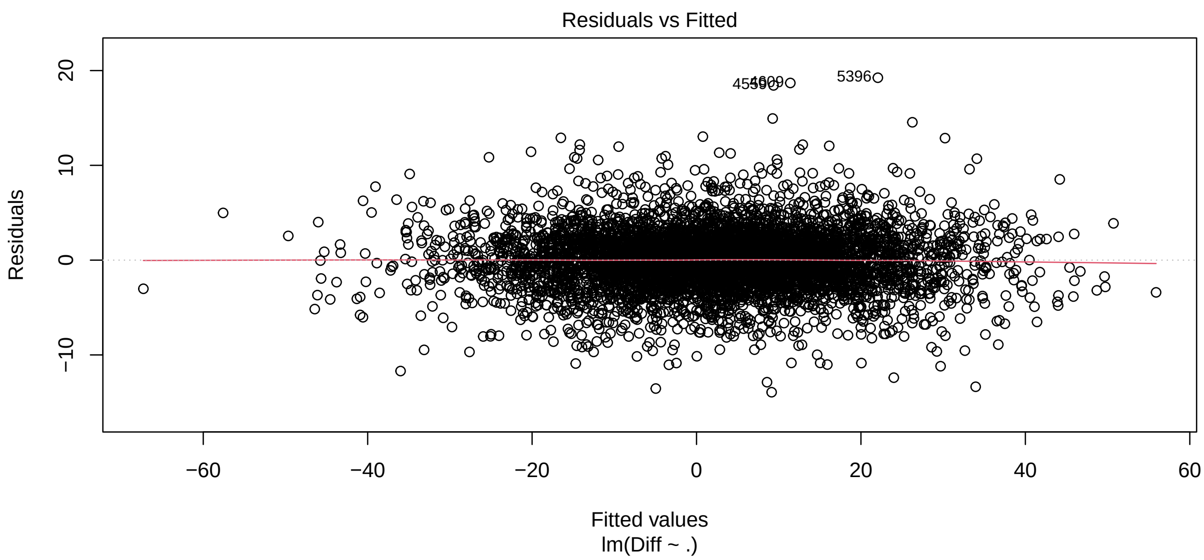


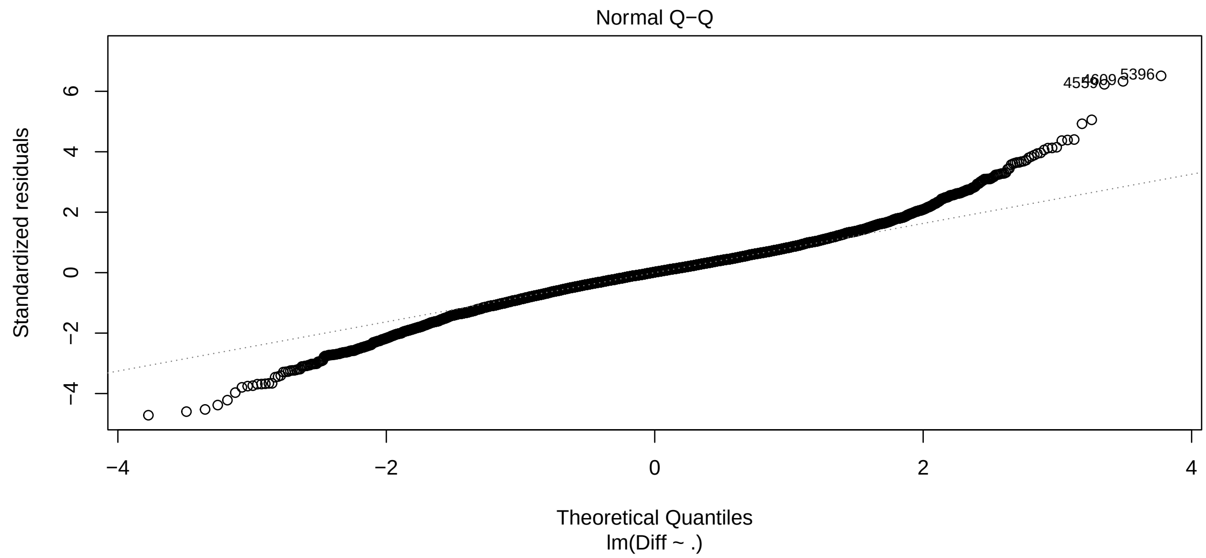



```
##                          PG.Experience
##                               1.070537
##                                 SG.Age
##                               5.824298
##                          SG.Experience
##                               5.798728
##                      OFFENSE..Pts.Poss
##                               1.827672
##                      DEFENSE..Pts.Poss
##                               4.456894
##                          DEFENSE..eFG.
##                               6.438050
##                          DEFENSE..ORB.
##                               2.565231
##         OFFENSE..Corner.Three.Frequency
##                               1.187837
##                        OFFENSE..Rim.FG.
##                               1.361553
##               OFFENSE..Corner.Three.FG.
##                               3.931931
##           OFFENSE..Non.Corner.Three.FG.
##                               9.440024
##                 OFFENSE..All.Three.FG.
##                              12.917689
##                      PUTBACKS..Pts.Play
##                               1.094687
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Pts.Play
```

```
##                          1.101666
##               DEFENSE..Rim.Frequency
##                          1.603177
##         DEFENSE..Short.Mid.Frequency
##                       6656.924898
##          DEFENSE..Long.Mid.Frequency
##                      26984.640436
##           DEFENSE..All.Mid.Frequency
##                      30225.761854
##                  DEFENSE..All.Mid.FG.
##                          2.015590
##                DEFENSE..All.Three.FG.
##                          2.174721
##          DEFENSE.PUTBACKS..Plays.Miss
##                          2.317205
##         DEFENSE.TRANSITION..Pts..Poss
##                          1.254024
##  DEFENSE.OFF.STEALS.TRANSITION..Freq
##                          1.055774
```

To absolve this, I removed effective field goal percentage, offensive and defensive points per possession, and the columns that were a sum of location frequencies (i.e. All Three Frequency being a sum of Corner and Non-Corner Three Frequency). After doing so, we are much closer to meeting the conditions for a linear model. The residuals seem to be evenly clustered around zero, and the linearity is acceptable enough to proceed with a second stepwise function.



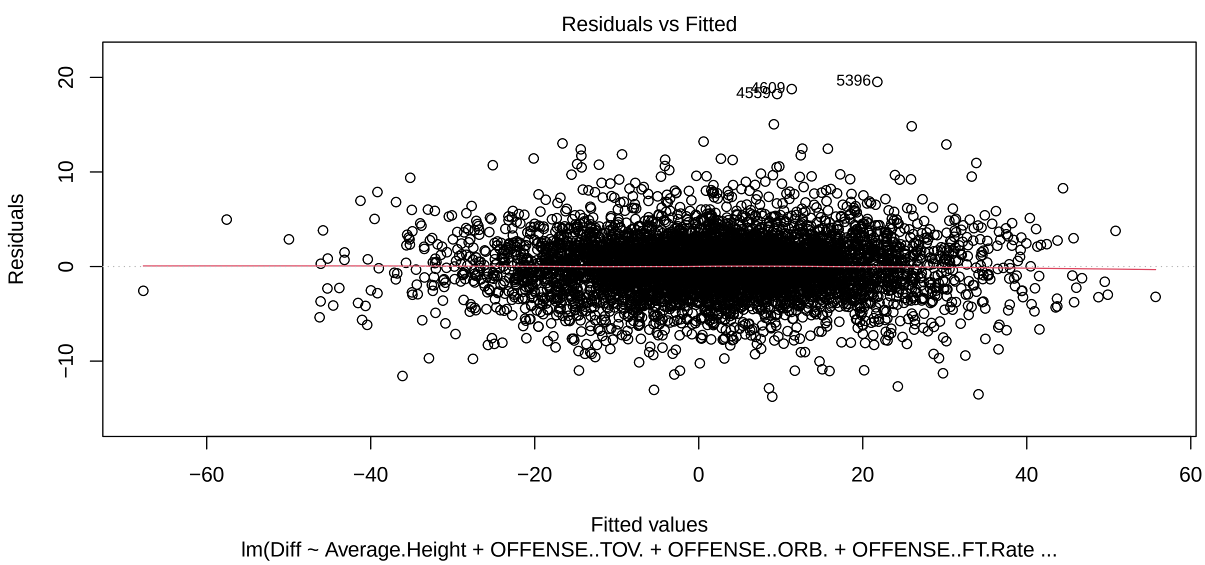


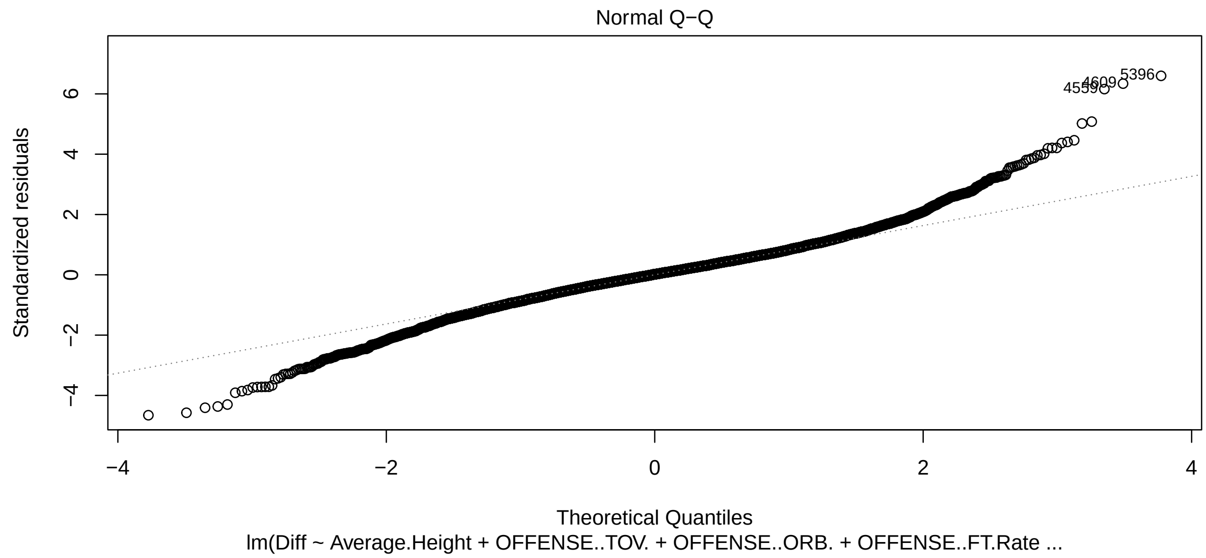
The predictors of this model are as follows:

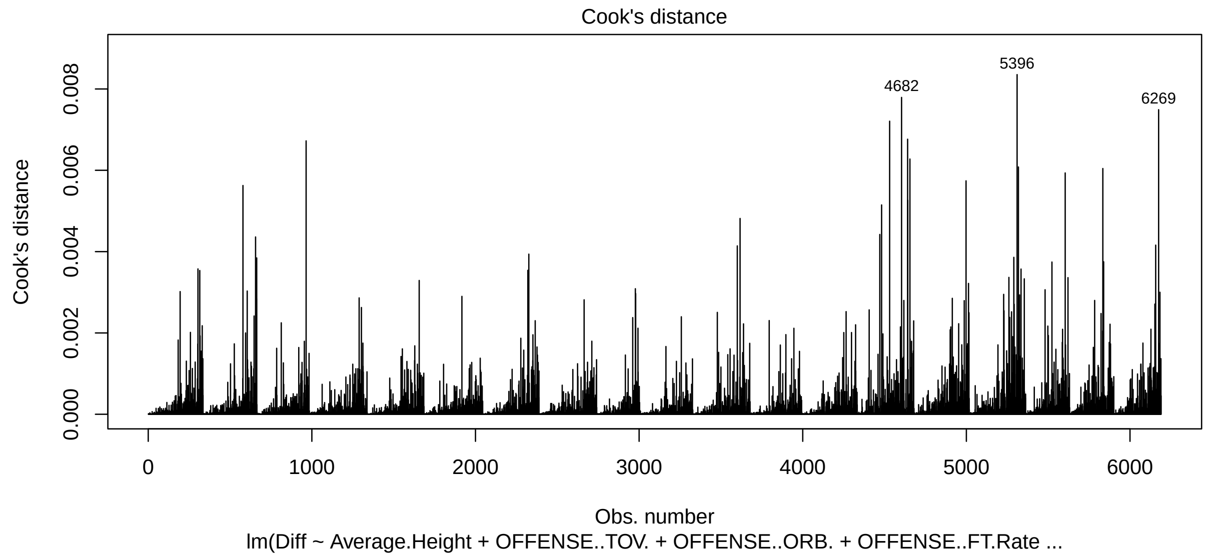
```
##
## Call:
## lm(formula = Diff ~ Average.Height + OFFENSE..TOV. + OFFENSE..ORB. +
##     OFFENSE..FT.Rate + DEFENSE..TOV. + DEFENSE..ORB. + DEFENSE..FT.Rate +
##     OFFENSE..Rim.Frequency + OFFENSE..Short.Mid.Frequency + OFFENSE..All.Mid.Frequency +
##     OFFENSE..Rim.FG. + OFFENSE..All.Mid.FG. + OFFENSE..All.Three.FG. +
##     HALFCOURT..Pts.Play + HALFCOURT..OREB. + HALFCOURT....of.Plays +
##     PUTBACKS..Pts.Miss + PUTBACKS..Pts.Play + OFFENSE.TRANSITION..Pts..Poss +
##     OFFENSE.TRANSITION..Freq + OFFENSE.OFF.STEALS.TRANSITION..Freq +
##     OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq + DEFENSE..Long.Mid.Frequency +
##     DEFENSE..Corner.Three.Frequency + DEFENSE..Non.Corner.Three.Frequency +
##     DEFENSE..Corner.Three.FG. + DEFENSE..All.Three.FG. + DEFENSE.HALFCOURT..Pts.Play +
##     DEFENSE.HALFCOURT..OREB. + DEFENSE.HALFCOURT....of.Plays +
##     DEFENSE.PUTBACKS..Pts.Miss + DEFENSE.PUTBACKS..Pts.Play +
##     DEFENSE.TRANSITION..Pts..Poss + DEFENSE.TRANSITION..Freq +
##     DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq, data = full2$model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.7622  -1.6225   0.0462   1.6343  19.5243
##
## Coefficients:
##                                             Estimate Std. Error  t value
## (Intercept)                                 8.157979   5.174468    1.577
## Average.Height                             -0.098839   0.049448   -1.999
## OFFENSE..TOV.                              -0.304617   0.018525  -16.443
## OFFENSE..ORB.                               0.309228   0.017165   18.015
## OFFENSE..FT.Rate                            0.043652   0.006350    6.874
## DEFENSE..TOV.                               0.293840   0.015064   19.506
## DEFENSE..ORB.                              -0.235189   0.017148  -13.715
## DEFENSE..FT.Rate                           -0.018421   0.006452   -2.855
## OFFENSE..Rim.Frequency                     -0.055634   0.007750   -7.178
## OFFENSE..Short.Mid.Frequency               -0.016569   0.008945   -1.852
## OFFENSE..All.Mid.Frequency                 -0.039083   0.006921   -5.647
## OFFENSE..Rim.FG.                            0.017560   0.007940    2.212
## OFFENSE..All.Mid.FG.                       -0.018991   0.009204   -2.063
## OFFENSE..All.Three.FG.                      0.025187   0.007926    3.178
## HALFCOURT..Pts.Play                         0.764985   0.010263   74.542
## HALFCOURT..OREB.                            0.111090   0.014530    7.646
## HALFCOURT....of.Plays                       0.374101   0.028372   13.185
## PUTBACKS..Pts.Miss                          0.225380   0.010385   21.702
## PUTBACKS..Pts.Play                          0.016273   0.001900    8.565
## OFFENSE.TRANSITION..Pts..Poss               0.910574   0.016259   56.005
## OFFENSE.TRANSITION..Freq                    0.379275   0.038912    9.747
## OFFENSE.OFF.STEALS.TRANSITION..Freq        -0.009218   0.003410   -2.703
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq -0.032301   0.011650   -2.773
## DEFENSE..Long.Mid.Frequency                -0.012899   0.008428   -1.531
## DEFENSE..Corner.Three.Frequency            -0.067875   0.016848   -4.029
## DEFENSE..Non.Corner.Three.Frequency        -0.046769   0.009745   -4.799
## DEFENSE..Corner.Three.FG.                  -0.004159   0.002521   -1.650
## DEFENSE..All.Three.FG.                     -0.031066   0.006346   -4.896
## DEFENSE.HALFCOURT..Pts.Play                -0.758672   0.006056 -125.279
## DEFENSE.HALFCOURT..OREB.                   -0.169516   0.014667  -11.557
## DEFENSE.HALFCOURT....of.Plays              -0.312737   0.029221  -10.703
## DEFENSE.PUTBACKS..Pts.Miss                 -0.244653   0.011001  -22.239
## DEFENSE.PUTBACKS..Pts.Play                 -0.017403   0.002004   -8.684
## DEFENSE.TRANSITION..Pts..Poss              -0.947171   0.013218  -71.659
## DEFENSE.TRANSITION..Freq                   -0.354061   0.034558  -10.245
## DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq  0.030615   0.009881    3.098
##                                            Pr(>|t|)
## (Intercept)                                 0.11494
## Average.Height                              0.04567 *
## OFFENSE..TOV.                               < 2e-16 ***
## OFFENSE..ORB.                               < 2e-16 ***
```

```
## OFFENSE..FT.Rate                          6.84e-12 ***
## DEFENSE..TOV.                              < 2e-16 ***
## DEFENSE..ORB.                              < 2e-16 ***
## DEFENSE..FT.Rate                           0.00431 **
## OFFENSE..Rim.Frequency                    7.88e-13 ***
## OFFENSE..Short.Mid.Frequency               0.06403 .
## OFFENSE..All.Mid.Frequency                1.70e-08 ***
## OFFENSE..Rim.FG.                           0.02703 *
## OFFENSE..All.Mid.FG.                       0.03912 *
## OFFENSE..All.Three.FG.                     0.00149 **
## HALFCOURT..Pts.Play                        < 2e-16 ***
## HALFCOURT..OREB.                          2.40e-14 ***
## HALFCOURT....of.Plays                      < 2e-16 ***
## PUTBACKS..Pts.Miss                         < 2e-16 ***
## PUTBACKS..Pts.Play                         < 2e-16 ***
## OFFENSE.TRANSITION..Pts..Poss              < 2e-16 ***
## OFFENSE.TRANSITION..Freq                   < 2e-16 ***
## OFFENSE.OFF.STEALS.TRANSITION..Freq        0.00689 **
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq  0.00558 **
## DEFENSE..Long.Mid.Frequency                0.12592
## DEFENSE..Corner.Three.Frequency           5.68e-05 ***
## DEFENSE..Non.Corner.Three.Frequency       1.63e-06 ***
## DEFENSE..Corner.Three.FG.                  0.09905 .
## DEFENSE..All.Three.FG.                    1.01e-06 ***
## DEFENSE.HALFCOURT..Pts.Play                < 2e-16 ***
## DEFENSE.HALFCOURT..OREB.                   < 2e-16 ***
## DEFENSE.HALFCOURT....of.Plays              < 2e-16 ***
## DEFENSE.PUTBACKS..Pts.Miss                 < 2e-16 ***
## DEFENSE.PUTBACKS..Pts.Play                 < 2e-16 ***
## DEFENSE.TRANSITION..Pts..Poss              < 2e-16 ***
## DEFENSE.TRANSITION..Freq                   < 2e-16 ***
## DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq  0.00196 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.969 on 6154 degrees of freedom
## Multiple R-squared:  0.9557, Adjusted R-squared:  0.9554
## F-statistic:  3792 on 35 and 6154 DF,  p-value: < 2.2e-16
```

Nearly all of the predictors are significant at a .05 level. The multiple R-Squared of 0.9557 is very encouraging as well. There are no VIFs above 9, the ANOVA suggests that all predictors carry importance, and we have no significant outliers within reach of Cook's Distance.








```
##                      Average.Height
##                            1.163168
##                       OFFENSE..TOV.
##                            2.335170
##                       OFFENSE..ORB.
##                            8.375231
##                    OFFENSE..FT.Rate
##                            1.828540
##                       DEFENSE..TOV.
##                            1.686373
##                       DEFENSE..ORB.
##                            6.747549
##                    DEFENSE..FT.Rate
##                            1.405469
##              OFFENSE..Rim.Frequency
##                            1.788624
##        OFFENSE..Short.Mid.Frequency
##                            1.544913
##          OFFENSE..All.Mid.Frequency
##                            3.634248
##                    OFFENSE..Rim.FG.
##                            2.896753
##                OFFENSE..All.Mid.FG.
##                            2.829897
##              OFFENSE..All.Three.FG.
##                            3.100248
##                 HALFCOURT..Pts.Play
```

```
##                                   8.669065
##                           HALFCOURT..OREB.
##                                   7.103150
##                      HALFCOURT....of.Plays
##                                   8.726058
##                         PUTBACKS..Pts.Miss
##                                   5.093040
##                         PUTBACKS..Pts.Play
##                                   2.832098
##              OFFENSE.TRANSITION..Pts..Poss
##                                   2.079771
##                   OFFENSE.TRANSITION..Freq
##                                  15.413510
##        OFFENSE.OFF.STEALS.TRANSITION..Freq
##                                   1.945518
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq
##                                   7.065334
##                DEFENSE..Long.Mid.Frequency
##                                   3.325985
##            DEFENSE..Corner.Three.Frequency
##                                   1.361172
##        DEFENSE..Non.Corner.Three.Frequency
##                                   2.647425
##                  DEFENSE..Corner.Three.FG.
##                                   1.292887
##                     DEFENSE..All.Three.FG.
##                                   1.880705
##                DEFENSE.HALFCOURT..Pts.Play
##                                   2.680297
##                   DEFENSE.HALFCOURT..OREB.
##                                   5.892095
##              DEFENSE.HALFCOURT....of.Plays
##                                   6.657870
##                 DEFENSE.PUTBACKS..Pts.Miss
##                                   5.147919
##                 DEFENSE.PUTBACKS..Pts.Play
##                                   3.054012
##              DEFENSE.TRANSITION..Pts..Poss
##                                   1.325842
##                   DEFENSE.TRANSITION..Freq
##                                   8.353917
## DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq
##                                   3.493737
```

```
## Analysis of Variance Table
##
## Response: Diff
##                                              Df Sum Sq Mean Sq    F value
## Average.Height                                1    847     847    96.0929
## OFFENSE..TOV.                                 1  74190   74190  8414.7506
## OFFENSE..ORB.                                 1  61686   61686  6996.5474
## OFFENSE..FT.Rate                              1  39160   39160  4441.6121
## DEFENSE..TOV.                                 1  71295   71295  8086.3464
## DEFENSE..ORB.                                 1  78640   78640  8919.4534
## DEFENSE..FT.Rate                              1  27041   27041  3067.0728
## OFFENSE..Rim.Frequency                        1     61      61     6.9683
## OFFENSE..Short.Mid.Frequency                  1  14630   14630  1659.4083
## OFFENSE..All.Mid.Frequency                    1  19672   19672  2231.2617
## OFFENSE..Rim.FG.                              1 110581  110581 12542.2397
## OFFENSE..All.Mid.FG.                          1  93966   93966 10657.7283
## OFFENSE..All.Three.FG.                        1 132325  132325 15008.5256
## HALFCOURT..Pts.Play                           1  15491   15491  1757.0056
## HALFCOURT..OREB.                              1    543     543    61.6235
## HALFCOURT....of.Plays                         1   8961    8961  1016.3339
## PUTBACKS..Pts.Miss                            1     16      16     1.7952
## PUTBACKS..Pts.Play                            1   5040    5040   571.6696
## OFFENSE.TRANSITION..Pts..Poss                 1  25594   25594  2902.9178
## OFFENSE.TRANSITION..Freq                      1   4538    4538   514.6539
## OFFENSE.OFF.STEALS.TRANSITION..Freq           1   6859    6859   777.9337
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq    1  86259   86259  9783.5890
## DEFENSE..Long.Mid.Frequency                   1  22373   22373  2537.5745
## DEFENSE..Corner.Three.Frequency               1   5112    5112   579.7962
## DEFENSE..Non.Corner.Three.Frequency           1   2628    2628   298.1265
## DEFENSE..Corner.Three.FG.                     1  18825   18825  2135.1399
## DEFENSE..All.Three.FG.                        1  68857   68857  7809.8506
## DEFENSE.HALFCOURT..Pts.Play                   1 107694  107694 12214.8023
## DEFENSE.HALFCOURT..OREB.                      1    289     289    32.7436
## DEFENSE.HALFCOURT....of.Plays                 1   1842    1842   208.9344
## DEFENSE.PUTBACKS..Pts.Miss                    1  13134   13134  1489.7024
## DEFENSE.PUTBACKS..Pts.Play                    1   3413    3413   387.1025
## DEFENSE.TRANSITION..Pts..Poss                 1  47540   47540  5392.0598
## DEFENSE.TRANSITION..Freq                      1    981     981   111.3217
## DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq    1     85      85     9.5994
## Residuals                                  6154  54258       9
##                                               Pr(>F)
## Average.Height                             < 2.2e-16 ***
## OFFENSE..TOV.                              < 2.2e-16 ***
## OFFENSE..ORB.                              < 2.2e-16 ***
## OFFENSE..FT.Rate                           < 2.2e-16 ***
## DEFENSE..TOV.                              < 2.2e-16 ***
## DEFENSE..ORB.                              < 2.2e-16 ***
## DEFENSE..FT.Rate                           < 2.2e-16 ***
## OFFENSE..Rim.Frequency                      0.008317 **
## OFFENSE..Short.Mid.Frequency               < 2.2e-16 ***
## OFFENSE..All.Mid.Frequency                 < 2.2e-16 ***
## OFFENSE..Rim.FG.                           < 2.2e-16 ***
## OFFENSE..All.Mid.FG.                       < 2.2e-16 ***
## OFFENSE..All.Three.FG.                     < 2.2e-16 ***
## HALFCOURT..Pts.Play                        < 2.2e-16 ***
## HALFCOURT..OREB.                           4.871e-15 ***
## HALFCOURT....of.Plays                      < 2.2e-16 ***
## PUTBACKS..Pts.Miss                          0.180345
## PUTBACKS..Pts.Play                         < 2.2e-16 ***
## OFFENSE.TRANSITION..Pts..Poss              < 2.2e-16 ***
## OFFENSE.TRANSITION..Freq                   < 2.2e-16 ***
## OFFENSE.OFF.STEALS.TRANSITION..Freq        < 2.2e-16 ***
```

```
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq < 2.2e-16 ***
## DEFENSE..Long.Mid.Frequency              < 2.2e-16 ***
## DEFENSE..Corner.Three.Frequency          < 2.2e-16 ***
## DEFENSE..Non.Corner.Three.Frequency      < 2.2e-16 ***
## DEFENSE..Corner.Three.FG.                < 2.2e-16 ***
## DEFENSE..All.Three.FG.                   < 2.2e-16 ***
## DEFENSE.HALFCOURT..Pts.Play              < 2.2e-16 ***
## DEFENSE.HALFCOURT..OREB.                 1.101e-08 ***
## DEFENSE.HALFCOURT....of.Plays            < 2.2e-16 ***
## DEFENSE.PUTBACKS..Pts.Miss               < 2.2e-16 ***
## DEFENSE.PUTBACKS..Pts.Play               < 2.2e-16 ***
## DEFENSE.TRANSITION..Pts..Poss            < 2.2e-16 ***
## DEFENSE.TRANSITION..Freq                 < 2.2e-16 ***
## DEFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq  0.001955 **
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I then preceded to make the model for recent NBA history, which I considered to be from 2014 to 2022. I used the same process as above - stepwise regression. Shockingly, the models ended up being very similar aside from two major predictors: offensive 3-point percentage and offensive rim field goal percentage were absent from the "recent" NBA model.

```
##
## Call:
## lm(formula = Diff ~ Average.Height + OFFENSE..TOV. + OFFENSE..ORB. +
##     OFFENSE..FT.Rate + DEFENSE..TOV. + DEFENSE..ORB. + DEFENSE..FT.Rate +
##     OFFENSE..All.Mid.Frequency + OFFENSE..Corner.Three.Frequency +
##     OFFENSE..Non.Corner.Three.Frequency + OFFENSE..All.Mid.FG. +
##     HALFCOURT..Pts.Play + HALFCOURT..OREB. + HALFCOURT....of.Plays +
##     PUTBACKS..Pts.Miss + PUTBACKS..Pts.Play + OFFENSE.TRANSITION..Pts..Poss +
##     OFFENSE.TRANSITION..Freq + OFFENSE.OFF.STEALS.TRANSITION..Freq +
##     OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq + DEFENSE..All.Mid.Frequency +
##     DEFENSE..Corner.Three.Frequency + DEFENSE..Non.Corner.Three.Frequency +
##     DEFENSE..Rim.FG. + DEFENSE..All.Mid.FG. + DEFENSE..Non.Corner.Three.FG. +
##     DEFENSE.HALFCOURT..Pts.Play + DEFENSE.HALFCOURT..OREB. +
##     DEFENSE.HALFCOURT....of.Plays + DEFENSE.PUTBACKS..Pts.Miss +
##     DEFENSE.PUTBACKS..Pts.Play + DEFENSE.TRANSITION..Pts..Poss +
##     DEFENSE.TRANSITION..Freq + DEFENSE.OFF.STEALS.TRANSITION..Freq,
##     data = full2022$model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9020  -1.8406  -0.0277   1.8238  17.6474
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 8.818170   8.717104   1.012
## Average.Height                             -0.181850   0.079620  -2.284
## OFFENSE..TOV.                              -0.257360   0.024511 -10.500
## OFFENSE..ORB.                               0.365099   0.026581  13.735
## OFFENSE..FT.Rate                            0.046641   0.009787   4.766
## DEFENSE..TOV.                               0.171974   0.028803   5.971
## DEFENSE..ORB.                              -0.215007   0.028699  -7.492
## DEFENSE..FT.Rate                            0.029344   0.011802   2.486
## OFFENSE..All.Mid.Frequency                  0.025381   0.011847   2.143
## OFFENSE..Corner.Three.Frequency             0.050711   0.023093   2.196
## OFFENSE..Non.Corner.Three.Frequency         0.086307   0.013880   6.218
## OFFENSE..All.Mid.FG.                       -0.031777   0.009906  -3.208
## HALFCOURT..Pts.Play                         0.805408   0.008207  98.132
## HALFCOURT..OREB.                            0.085430   0.023386   3.653
## HALFCOURT....of.Plays                       0.548122   0.045993  11.918
## PUTBACKS..Pts.Miss                          0.253328   0.016218  15.620
## PUTBACKS..Pts.Play                          0.015078   0.002898   5.203
## OFFENSE.TRANSITION..Pts..Poss               0.981422   0.019394  50.603
## OFFENSE.TRANSITION..Freq                    0.554798   0.065072   8.526
## OFFENSE.OFF.STEALS.TRANSITION..Freq        -0.009383   0.005455  -1.720
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq -0.047136   0.020180  -2.336
## DEFENSE..All.Mid.Frequency                 -0.065315   0.015369  -4.250
## DEFENSE..Corner.Three.Frequency            -0.101302   0.026969  -3.756
## DEFENSE..Non.Corner.Three.Frequency        -0.073974   0.016763  -4.413
## DEFENSE..Rim.FG.                            0.040175   0.011728   3.426
## DEFENSE..All.Mid.FG.                        0.058738   0.012939   4.540
## DEFENSE..Non.Corner.Three.FG.               0.019351   0.012015   1.611
## DEFENSE.HALFCOURT..Pts.Play                -0.822141   0.014242 -57.728
## DEFENSE.HALFCOURT..OREB.                   -0.181262   0.024029  -7.544
## DEFENSE.HALFCOURT....of.Plays              -0.494196   0.048791 -10.129
## DEFENSE.PUTBACKS..Pts.Miss                 -0.306698   0.018452 -16.621
## DEFENSE.PUTBACKS..Pts.Play                 -0.010821   0.003153  -3.432
## DEFENSE.TRANSITION..Pts..Poss              -1.008466   0.023568 -42.789
## DEFENSE.TRANSITION..Freq                   -0.438013   0.046454  -9.429
## DEFENSE.OFF.STEALS.TRANSITION..Freq        -0.010503   0.004760  -2.207
##                                            Pr(>|t|)
## (Intercept)                                0.311819
## Average.Height                             0.022447 *
## OFFENSE..TOV.                               < 2e-16 ***
## OFFENSE..ORB.                               < 2e-16 ***
## OFFENSE..FT.Rate                           1.98e-06 ***
## DEFENSE..TOV.                              2.66e-09 ***
## DEFENSE..ORB.                              9.04e-14 ***
## DEFENSE..FT.Rate                           0.012964 *
## OFFENSE..All.Mid.Frequency                 0.032237 *
## OFFENSE..Corner.Three.Frequency            0.028178 *
## OFFENSE..Non.Corner.Three.Frequency        5.77e-10 ***
## OFFENSE..All.Mid.FG.                       0.001352 **
## HALFCOURT..Pts.Play                         < 2e-16 ***
## HALFCOURT..OREB.                           0.000264 ***
## HALFCOURT....of.Plays                       < 2e-16 ***
## PUTBACKS..Pts.Miss                          < 2e-16 ***
## PUTBACKS..Pts.Play                         2.10e-07 ***
```

```
## OFFENSE.TRANSITION..Pts..Poss               < 2e-16 ***
## OFFENSE.TRANSITION..Freq                    < 2e-16 ***
## OFFENSE.OFF.STEALS.TRANSITION..Freq        0.085501 .
## OFFENSE.OFF.LIVE.REBOUNDS.TRANSITION..Freq 0.019570 *
## DEFENSE..All.Mid.Frequency                 2.21e-05 ***
## DEFENSE..Corner.Three.Frequency            0.000176 ***
## DEFENSE..Non.Corner.Three.Frequency        1.06e-05 ***
## DEFENSE..Rim.FG.                           0.000622 ***
## DEFENSE..All.Mid.FG.                       5.87e-06 ***
## DEFENSE..Non.Corner.Three.FG.              0.107391
## DEFENSE.HALFCOURT..Pts.Play                 < 2e-16 ***
## DEFENSE.HALFCOURT..OREB.                   6.13e-14 ***
## DEFENSE.HALFCOURT....of.Plays               < 2e-16 ***
## DEFENSE.PUTBACKS..Pts.Miss                  < 2e-16 ***
## DEFENSE.PUTBACKS..Pts.Play                 0.000607 ***
## DEFENSE.TRANSITION..Pts..Poss               < 2e-16 ***
## DEFENSE.TRANSITION..Freq                    < 2e-16 ***
## DEFENSE.OFF.STEALS.TRANSITION..Freq        0.027425 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.279 on 2827 degrees of freedom
## Multiple R-squared:  0.9518, Adjusted R-squared:  0.9512
## F-statistic:  1641 on 34 and 2827 DF,  p-value: < 2.2e-16
```

I believe this is because many NBA teams now prioritize shooting, meaning that the spread of 3 point percentage between lineups has decreased. Therefore, its significance was removed from the model. Thus, the other predictors in the model gain more importance (as seen when comparing the estimate values between the models). While I thought 3 point percentage would see an increase in importance when predicting recent NBA history due to the spike in attempts per game, I can also see why its significance has lowered. As for the predictors that are used in my model, there is a heavy emphasis on offensive rebounds, turnovers, and transition. Shot locations are significant but have a smaller importance to the model. One surprising predictor present in both models that surprised me was average height; the model favors lineups with smaller average height levels. A few NBA teams have experimented with "small-ball" to increase efficiency, but I believe it has not been done enough, and the model agrees. However, using smaller players may also subject the lineup to a lack of rebounding, and therefore opportunities in transition.

Lastly, I wanted to evaluate the effectiveness of the two models in predicting current-day NBA lineups. When predicting on data from 2004-2022, the first model had an average residual of +/- 2.17 with a median of +/-1.63 (overall very successful considering the distribution of efficiency differentials typically spreads from -15 to 15). However, when predicting on data from this current NBA season (2022-2023), the first model had an average residual +/- 2.66 with a median of +/-1.99, meaning the model performs slightly worse on current data. The second model, designed specifically for recent NBA history, had an average residual of 2.39 with a median residual of +/-1.83. Therefore, by a slim margin, the second model would be preferable to use on current data, even without the presence of 3 point percentage as a predictor. This suggests that the modeling procedures for analyzing the NBA must change, and researchers must be careful when using non-recent data to predict the future success of lineups. Not only can my model be used to analyze this success, but can be used as evidence for the shift in the dynamic of the NBA.