Jess, Anna and Seth Project

12/6/19

project r markdown document

```r
source('packages.R')
```

```
## Loading required package: optimx

## Loading required package: parallel

## Loading required package: minqa

## Loading required package: lme4

## Loading required package: Matrix

## Loading required package: segmented

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

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

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

## Loading required package: ggplot2

## Loading required package: ggcorrplot
```

```r
source('styleguide.R')
source('helpers.R')
source('cleaner.R')
source('models.R')
```

```
## [1] "Dim Check Successful"
## [1] "Dim Check Successful"

## refitting model(s) with ML (instead of REML)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00304692 (tol = 0.002, component 1)

## Loading required namespace: dfoptim

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00938467 (tol = 0.002, component 1)

## Loading required namespace: dfoptim

## refitting model(s) with ML (instead of REML)
```

Get the Data

```r
# Read in Clean DF
df.clean <- add_time("complete_data_clean.csv")
df.clean <- add_coach_change(df.clean)
df.tourney <- add_time("tourney_data_clean.csv")
df.tourney <- add_coach_change(df.tourney)
```

```r
# Check dimensions - len(unique schools) * len(unique years) must equal # of rows
dim_checker(df.clean)
```

```
## [1] "Dim Check Successful"
```

```r
dim_checker(df.tourney)
```

```
## [1] "Dim Check Successful"
```

```r
get_newprop = cbind(df.tourney$School, get_prop_df(df.tourney))
```

```r
# gets the minimum slope
sorted_alpha_school = df.tourney[order(df.tourney$School),]
sorted_alpha_prop = get_newprop[order(get_newprop$`df.tourney$School`),]
sorted_alpha_school$winner = (sorted_alpha_school$W.L. >= 0.5) * 1
```

```r
total_wins <- aggregate(sorted_alpha_school$winner, by=list(sorted_alpha_school$School), FUN=sum)[2]
total_3s_made <- aggregate(sorted_alpha_prop$X3P, by=list(sorted_alpha_prop$`df.tourney$School`), FUN=me
total_3s_attempted <- aggregate(sorted_alpha_prop$X3PA, by=list(sorted_alpha_prop$`df.tourney$School`),
total_sos <- aggregate(sorted_alpha_school$SOS, by=list(sorted_alpha_school$School), FUN=sum)[2]
```

```r
idx_min <- which.min(coef(lmer3d)$School[,2])
school_list <- sort(unique(df.tourney$School)) # alphabetically sort the list
school_list[idx_min]
```

```
## [1] Butler
## 232 Levels: Air Force Akron Alabama Alabama A&M ... Yale
```

```r
coef(lmer3d)$School[idx_min, 2]
```

```
## [1] -0.005638882
```

```r
new_df <- data.frame(school = school_list, slope = coef(lmer3d)$School[,2], wins = total_wins, threes_m
new_df = new_df[order(new_df$slope),]
```

Find the Schools Most Prone and Least Prone by taking the Absolute value (slopes closer to 0 will be least impacted by the change) and we define most affected as a larger absolute value

```r
# check for with absolute value
abs_df <- new_df
abs_df$slope = abs(abs_df$slope)
abs_df = abs_df[order(abs_df$slope),]
abs_df
```

```
##                          school        slope  x      x.1     x.2
```

```
## 62                George Mason 2.790870e-05 12 5.721800   12.11
## 192                     Temple 4.965195e-05 14 7.185309   78.46
## 193                  Tennessee 5.843993e-05 14 6.813296  131.42
## 223           Western Kentucky 9.733913e-05 14 6.897856  -12.55
## 59               Florida State 1.097211e-04 15 6.462685  125.03
## 127                   Nebraska 1.219053e-04 10 6.310638  114.91
## 139 North Carolina-Wilmington 1.276611e-04  7 6.744223  -13.77
## 6          Alabama-Birmingham 1.326660e-04 13 6.522186   26.18
## 135        North Carolina A&T 1.548420e-04  4 6.284508 -127.78
## 29  Central Connecticut State 1.712815e-04  4 6.017776 -109.53
## 161                 Princeton 1.775000e-04 11 7.689089  -47.77
## 153                    Oregon 3.229018e-04 14 7.516776  111.73
## 27         Cal State Fullerton 3.261316e-04  8 7.021127  -38.99
## 151              Old Dominion 3.668240e-04 15 5.791018   -6.92
## 148                Ohio State 3.708165e-04 15 6.761041  126.37
## 183        Southern Illinois 3.806682e-04 10 5.606899   38.59
## 231                    Xavier 4.581981e-04 16 6.679811   99.84
## 77                    Indiana 5.206040e-04 12 7.008277  129.17
## 167             Robert Morris 5.720029e-04 11 7.052139  -92.65
## 155                   Pacific 5.836199e-04  9 6.494608    5.00
## 205                 UC-Irvine 5.994744e-04 10 6.647123  -22.18
## 168            Saint Joseph's 6.363121e-04 11 6.897087   57.38
## 30            Central Florida 6.419951e-04 11 6.862061   18.65
## 186          St. Bonaventure 6.996270e-04  9 6.064039   22.12
## 145                Notre Dame 7.377932e-04 14 8.085865  118.65
## 65                    Georgia 7.423079e-04 10 5.953399  119.52
## 96     Long Island University 7.861274e-04  9 7.053123 -118.59
## 67                Georgia Tech 7.952019e-04  7 5.608172  123.91
## 21                    Bradley 8.013806e-04  8 6.167268   43.95
## 64                 Georgetown 8.138741e-04 13 6.675569  124.15
## 14                Austin Peay 8.166123e-04 10 6.229611  -54.87
## 48             Delaware State 8.344733e-04  6 6.385202 -114.52
## 178            South Carolina 8.441718e-04  9 6.655275  110.81
## 38      College of Charleston 8.900277e-04 13 7.226049  -45.47
## 91                      Lamar 9.010590e-04  9 6.142502  -83.81
## 215                  Virginia 9.427137e-04 13 6.483533  128.79
## 90                  Lafayette 9.736049e-04  5 8.156764  -73.30
## 39                   Colorado 9.935891e-04 10 6.365386  105.60
## 162                Providence 9.961643e-04 10 6.544888  122.92
## 130                New Mexico 1.066940e-03 13 7.314909   50.60
## 124             Morgan State 1.134026e-03  5 5.015951 -133.87
## 58                Florida A&M 1.139391e-03  1 5.988670 -128.30
## 83         Jacksonville State 1.165320e-03  6 6.407183  -85.46
## 177             South Alabama 1.190902e-03  8 6.773379  -39.59
## 47                   Delaware 1.193842e-03  5 6.198627  -24.96
## 28       Cal State Northridge 1.253521e-03  3 5.511947  -29.70
## 197           Texas Christian 1.258382e-03  6 6.459154   85.72
## 166                  Richmond 1.278710e-03 11 7.023925   45.12
## 80                       Iowa 1.280903e-03 11 6.650355  120.12
## 147                      Ohio 1.307250e-03 11 7.244677   -3.65
## 171             Saint Peter's 1.319750e-03  6 6.359713  -58.27
## 132                   Niagara 1.330204e-03  8 7.162730  -50.53
## 69                  Green Bay 1.352620e-03 13 6.269094   -4.07
## 22              Brigham Young 1.371272e-03 15 7.156491   59.49
```

```
## 211                Valparaiso 1.389770e-03 11 6.976101   -4.98
## 101               Loyola (MD) 1.402737e-03  6 5.840579  -58.56
## 74                    Houston 1.406796e-03 13 7.688357   32.48
## 78              Indiana State 1.456388e-03  5 6.684037   48.49
## 11                   Arkansas 1.469489e-03 13 6.539985   97.80
## 70                    Hampton 1.491714e-03  9 5.823553 -137.64
## 61               Gardner-Webb 1.493835e-03  9 7.225362  -75.49
## 60               Fresno State 1.515991e-03  8 7.215728   28.25
## 88                   Kentucky 1.524159e-03 16 6.154633  137.21
## 129          Nevada-Las Vegas 1.529315e-03 15 6.885973   64.19
## 33                 Cincinnati 1.530182e-03 14 6.602163  100.89
## 181                  Southern 1.557363e-03  6 5.211291 -161.81
## 143              Northwestern 1.562525e-03  6 7.670868  115.15
## 108                   Memphis 1.625274e-03 16 6.569505   68.71
## 15                     Baylor 1.632899e-03 12 7.124214  126.00
## 97                  Louisiana 1.653640e-03 10 6.676045  -32.86
## 159            Portland State 1.696784e-03  9 6.892471  -68.40
## 7                 Albany (NY) 1.697746e-03 11 6.131447  -89.52
## 152              Oral Roberts 1.718486e-03 12 6.156634  -22.52
## 156                Penn State 1.722025e-03  6 6.484239  119.87
## 224         Western Michigan 1.727121e-03 11 6.394350   -6.74
## 75                   Illinois 1.729081e-03 12 6.995798  133.43
## 31                  Charlotte 1.771142e-03  7 7.041463   25.91
## 54          Eastern Kentucky 1.800401e-03  8 8.211955  -74.73
## 119  Mississippi Valley State 1.856378e-03  4 5.889948 -141.20
## 36          Coastal Carolina 1.875451e-03 10 6.724475  -86.19
## 16                    Belmont 1.877112e-03 15 9.586143  -52.17
## 118         Mississippi State 1.881707e-03 12 6.823209   87.91
## 206          UC-Santa Barbara 1.895579e-03 13 6.725307  -20.67
## 122                   Montana 1.902487e-03 14 6.873324  -53.46
## 4                  Alabama A&M 1.910670e-03  2 5.397392 -177.09
## 87                 Kent State 1.915808e-03 16 6.676689   -3.86
## 116                 Minnesota 1.986769e-03 12 6.003933  126.07
## 52                       Duke 2.011376e-03 16 7.887467  155.52
## 107             Massachusetts 2.019466e-03  8 6.887404   44.09
## 213                   Vermont 2.026286e-03 15 6.061072  -74.68
## 26                   Cal Poly 2.070102e-03  4 7.171318  -24.21
## 149                  Oklahoma 2.086706e-03 12 6.866157  137.27
## 56       Fairleigh Dickinson 2.123456e-03  5 6.329318  -93.07
## 225             Wichita State 2.204181e-03 15 7.019215   60.38
## 5               Alabama State 2.213276e-03  7 5.898918 -168.43
## 184        Southern Methodist 2.215674e-03  7 6.250370   38.12
## 49                     DePaul 2.273036e-03  4 6.047734  112.75
## 170         Saint Mary's (CA) 2.280819e-03 16 7.471634   33.82
## 66              Georgia State 2.326856e-03  8 6.626472  -34.88
## 111                 Miami (OH) 2.333651e-03  6 6.361490   18.82
## 32                Chattanooga 2.363351e-03 10 7.350013  -55.39
## 182       Southern California 2.379242e-03  9 6.079078  113.76
## 227                 Wisconsin 2.387344e-03 15 6.979702  141.29
## 228                   Wofford 2.412120e-03 10 7.551247  -44.25
## 179             South Florida 2.439797e-03  3 5.047055   78.34
## 103                 Marquette 2.440084e-03 15 7.204356  119.20
## 99                 Louisville 2.466948e-03 16 7.496409  130.88
## 196  Texas A&M-Corpus Christi 2.470178e-03 10 5.321892  -73.10
```

```
## 195                   Texas A&M 2.494641e-03 13 6.044516  104.01
## 212                  Vanderbilt 2.523020e-03 12 7.993621  118.38
## 114           Middle Tennessee 2.538578e-03 14 6.000197  -18.79
## 165               Rhode Island 2.543603e-03 12 5.964717   44.10
## 20          Boston University 2.607326e-03 10 7.259560  -69.89
## 23                   Bucknell 2.621217e-03 12 6.741496  -58.56
## 121                  Monmouth 2.626323e-03  6 6.447372  -69.77
## 115                 Milwaukee 2.714735e-03  9 7.419218    1.10
## 109                    Mercer 2.722499e-03  8 7.094293  -60.00
## 134            North Carolina 2.729929e-03 16 6.271221  157.06
## 10             Arizona State 2.758011e-03  9 7.383960  104.62
## 207                      UCLA 2.813067e-03 13 6.447906  121.42
## 17                Binghamton 2.816150e-03  2 6.259588  -85.43
## 216    Virginia Commonwealth 2.835503e-03 16 7.245661   29.08
## 57                   Florida 2.847409e-03 15 7.588434  126.60
## 144       Northwestern State 2.879030e-03  8 6.211679  -69.89
## 51                     Drake 2.883595e-03  6 7.336464   40.55
## 169              Saint Louis 2.885301e-03 11 5.734167   51.13
## 141             Northeastern 2.911363e-03 10 7.076050  -12.96
## 46                    Dayton 2.965604e-03 14 6.703542   55.20
## 218              Wake Forest 2.997841e-03  8 6.216456  130.39
## 137 North Carolina-Asheville 3.012658e-03  8 6.429142  -65.19
## 154             Oregon State 3.019166e-03  8 5.747199   89.39
## 123           Morehead State 3.023707e-03  9 6.133421  -55.72
## 172        Sam Houston State 3.024168e-03 14 7.054085  -79.98
## 208 University of California 3.026062e-03 11 5.676435  114.80
## 34                   Clemson 3.053811e-03 14 6.581742  119.01
## 98           Louisiana State 3.118372e-03 12 6.086522  102.33
## 201            Texas-El Paso 3.145531e-03 11 5.901999    9.68
## 176                    Siena 3.168709e-03  9 5.891237  -36.43
## 8                   American 3.217724e-03 10 6.462683  -79.77
## 131         New Mexico State 3.225442e-03 13 6.253432  -10.87
## 120                 Missouri 3.231862e-03 10 7.055662  115.65
## 63         George Washington 3.258254e-03 10 5.648510   34.79
## 117              Mississippi 3.289159e-03 12 7.074049   96.19
## 136     North Carolina State 3.318133e-03 12 6.563238  124.84
## 194                    Texas 3.351889e-03 14 6.603725  140.57
## 92                    Lehigh 3.378968e-03 12 6.770819  -86.33
## 175               Seton Hall 3.430851e-03 12 6.559728  113.11
## 45                  Davidson 3.441282e-03 16 8.925587   -4.21
## 157             Pennsylvania 3.444370e-03  7 6.806416  -30.58
## 9                    Arizona 3.455211e-03 16 6.456764  127.18
## 13                   Auburn 3.470796e-03  6 7.512625   97.08
## 100              Loyola (IL) 3.601446e-03  7 5.874828    0.23
## 43                   Cornell 3.645366e-03  4 7.596776  -46.09
## 140              North Texas 3.717216e-03 11 6.054695  -45.62
## 203                     Troy 3.816750e-03  4 8.779607  -52.43
## 174          San Diego State 3.828083e-03 14 6.031051   67.69
## 89                  La Salle 3.909470e-03  6 6.832841   36.17
## 126             Murray State 3.922218e-03 15 7.033542  -63.31
## 142            Northern Iowa 4.059802e-03 14 7.250434   60.82
## 71                   Harvard 4.069929e-03 10 6.375169  -40.46
## 110               Miami (FL) 4.106440e-03 13 7.058740  117.45
## 3                    Alabama 4.111946e-03 15 6.144707  117.73
```

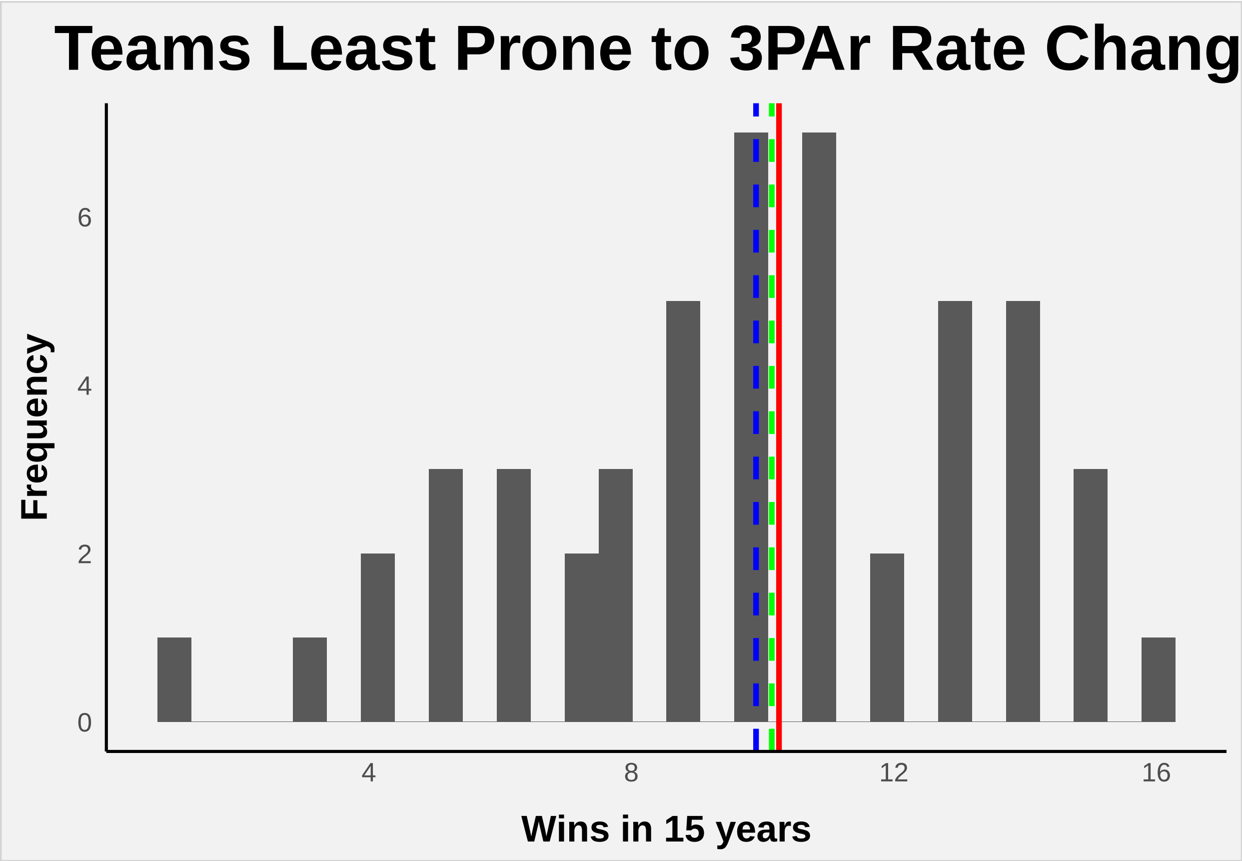
```
## 84              James Madison 4.124645e-03  5 6.370749  -27.90
## 18                Boise State 4.139649e-03 11 7.599539   28.63
## 222             West Virginia 4.257394e-03 14 7.405296  135.42
## 94                   Lipscomb 4.273298e-03 10 7.756064  -70.72
## 44                  Creighton 4.305770e-03 15 8.211294   81.65
## 85                     Kansas 4.321536e-03 16 6.806275  155.83
## 82              Jackson State 4.322197e-03  5 5.397949 -156.81
## 160             Prairie View 4.403416e-03  3 5.471237 -148.17
## 95                Little Rock 4.423617e-03  7 5.804571  -41.89
## 188                  Stanford 4.620269e-03 12 5.993873  114.23
## 35            Cleveland State 4.682915e-03  6 6.536270    1.94
## 204                     Tulsa 4.724622e-03 12 6.203486   45.07
## 1                   Air Force 4.791179e-03  7 7.551358   28.56
## 133            Norfolk State 4.849900e-03  8 5.633081 -135.23
## 138 North Carolina-Greensboro 4.873969e-03  6 6.754625  -49.34
## 202         Texas-San Antonio 4.927869e-03  8 6.895634  -51.29
## 219                Washington 4.992617e-03 14 6.433361  109.75
## 221              Weber State 5.013173e-03 13 7.284673  -57.12
## 209                      Utah 5.060943e-03 10 6.760732   88.26
## 232                      Yale 5.079997e-03 10 5.948941  -44.96
## 163                    Purdue 5.099678e-03 12 6.741528  130.76
## 146                   Oakland 5.215012e-03  9 7.514128  -16.78
## 164                   Radford 5.256009e-03  8 6.228929  -82.25
## 113            Michigan State 5.265836e-03 16 6.583259  145.03
## 189         Stephen F. Austin 5.313053e-03 14 5.719587 -101.35
## 112                  Michigan 5.328781e-03 13 7.552010  141.21
## 12        Arkansas-Pine Bluff 5.352610e-03  2 5.302537 -141.64
## 180    Southeastern Louisiana 5.354288e-03 10 6.071915  -95.57
## 104                  Marshall 5.420877e-03  8 7.229566   17.83
## 68                    Gonzaga 5.447725e-03 16 6.716946   63.89
## 86              Kansas State 5.456449e-03 15 6.048941  126.17
## 158                Pittsburgh 5.519558e-03 13 6.027521  120.84
## 220          Washington State 5.589081e-03  6 6.726650   96.44
## 93                    Liberty 5.599490e-03  7 6.709843  -85.98
## 25                     Butler 5.638882e-03 13 7.592211   77.56
## 150            Oklahoma State 5.681481e-03 13 6.902494  134.27
## 190               Stony Brook 5.696688e-03  9 5.964597  -86.47
## 191                  Syracuse 5.703737e-03 16 6.402456  132.36
## 24                    Buffalo 5.785106e-03 13 6.801858   -2.92
## 187           St. John's (NY) 5.786628e-03  7 5.467189  122.86
## 229              Wright State 5.835519e-03 13 6.273270  -13.54
## 76           Illinois-Chicago 5.889671e-03  8 6.128775   -7.25
## 53       East Tennessee State 5.897911e-03 14 6.681445  -52.78
## 40            Colorado State 5.923076e-03 10 6.260372   55.15
## 210                Utah State 5.931023e-03 15 6.799605    9.68
## 73                 Holy Cross 6.048317e-03  7 6.021283  -71.05
## 55         Eastern Washington 6.074115e-03  7 7.771243  -46.53
## 50             Detroit Mercy 6.136221e-03  7 6.198857    0.05
## 198            Texas Southern 6.179689e-03  8 5.725888 -140.53
## 102                 Manhattan 6.204649e-03  7 5.983108  -47.39
## 37                    Colgate 6.251096e-03  4 7.063172  -78.87
## 125          Mount St. Mary's 6.280049e-03  7 6.767301  -90.44
## 72                     Hawaii 6.412211e-03 12 5.974273  -12.30
## 173                 San Diego 6.443620e-03  8 6.695405   26.83
```

```
## 81                Iowa State 6.528966e-03 11 7.404039  128.89
## 200           Texas-Arlington 6.685139e-03 11 6.889081  -51.57
## 2                       Akron 6.838891e-03 14 8.093001   -7.63
## 226                  Winthrop 6.847741e-03 12 7.143976  -79.14
## 214                 Villanova 6.990611e-03 15 7.944232  135.88
## 199                Texas Tech 7.033891e-03  9 5.538718  124.81
## 19             Boston College 7.400558e-03  7 6.588496  117.58
## 41                Connecticut 7.528914e-03 13 5.894859  114.57
## 128                    Nevada 7.869737e-03 12 6.260792   38.89
## 217             Virginia Tech 7.964500e-03 11 6.269554  102.13
## 106 Maryland-Baltimore County 7.983819e-03  4 6.211620  -85.24
## 185      Southern Mississippi 8.523459e-03  8 5.687583   15.37
## 42               Coppin State 8.868988e-03  2 6.734071  -99.70
## 79                       Iona 8.913573e-03 11 7.580094  -36.21
## 105                  Maryland 8.971031e-03 16 6.165858  124.85
## 230                   Wyoming 9.297598e-03  9 6.490703   43.16
```

```r
close_zero_least <- abs_df[1:50,]
close_zero_most <- abs_df[181:230,]
```

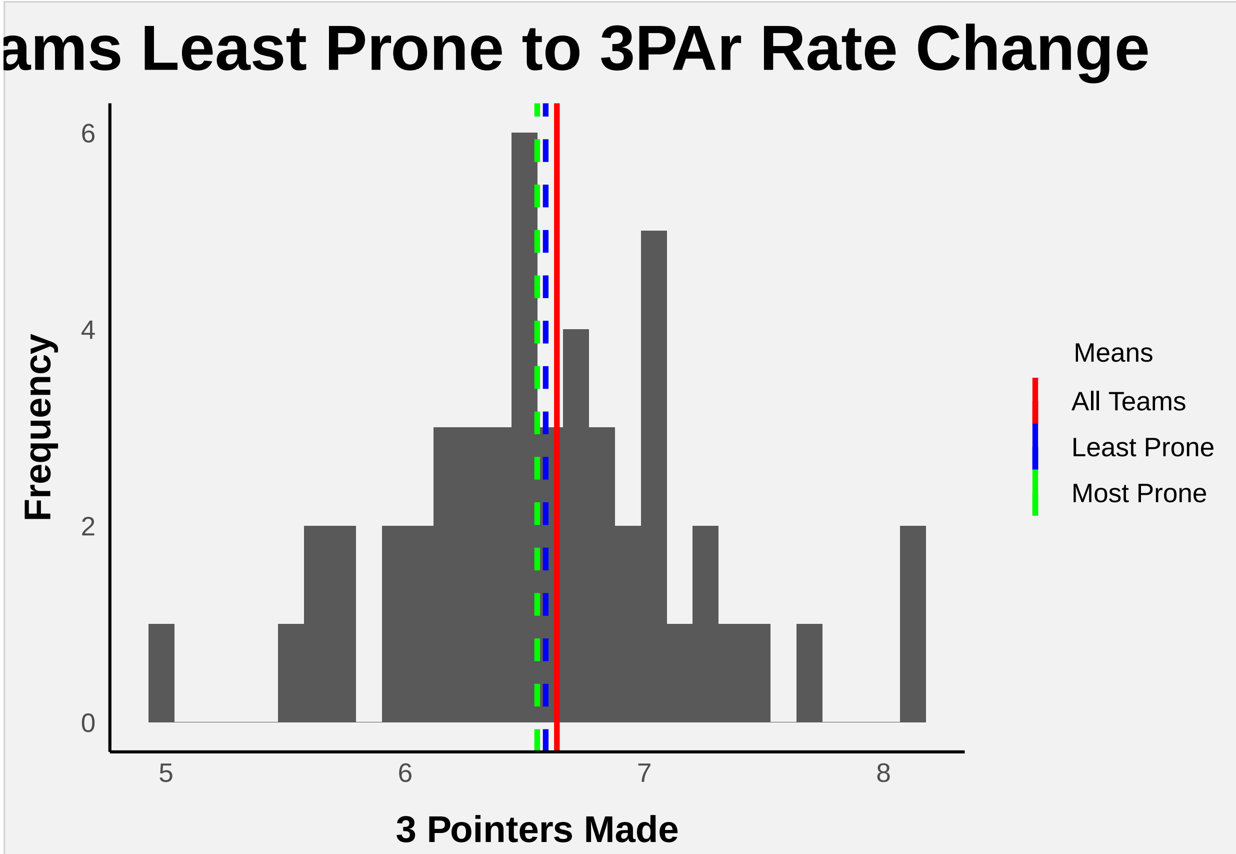
```r
# histograms - teams least prone

ggplot(close_zero_least, aes(x = x)) +
  geom_histogram() +
  geom_vline(aes(xintercept=mean(total_wins$x)),
           color="red", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_least$x)),
           color="blue",linetype="dashed", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_most$x)),
           color="green",linetype="dashed", size=1) +
  theme_hodp() +
  labs(title="Teams Least Prone to 3PAr Rate Change") +
  xlab("Wins in 15 years") +
  ylab("Frequency")
```

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



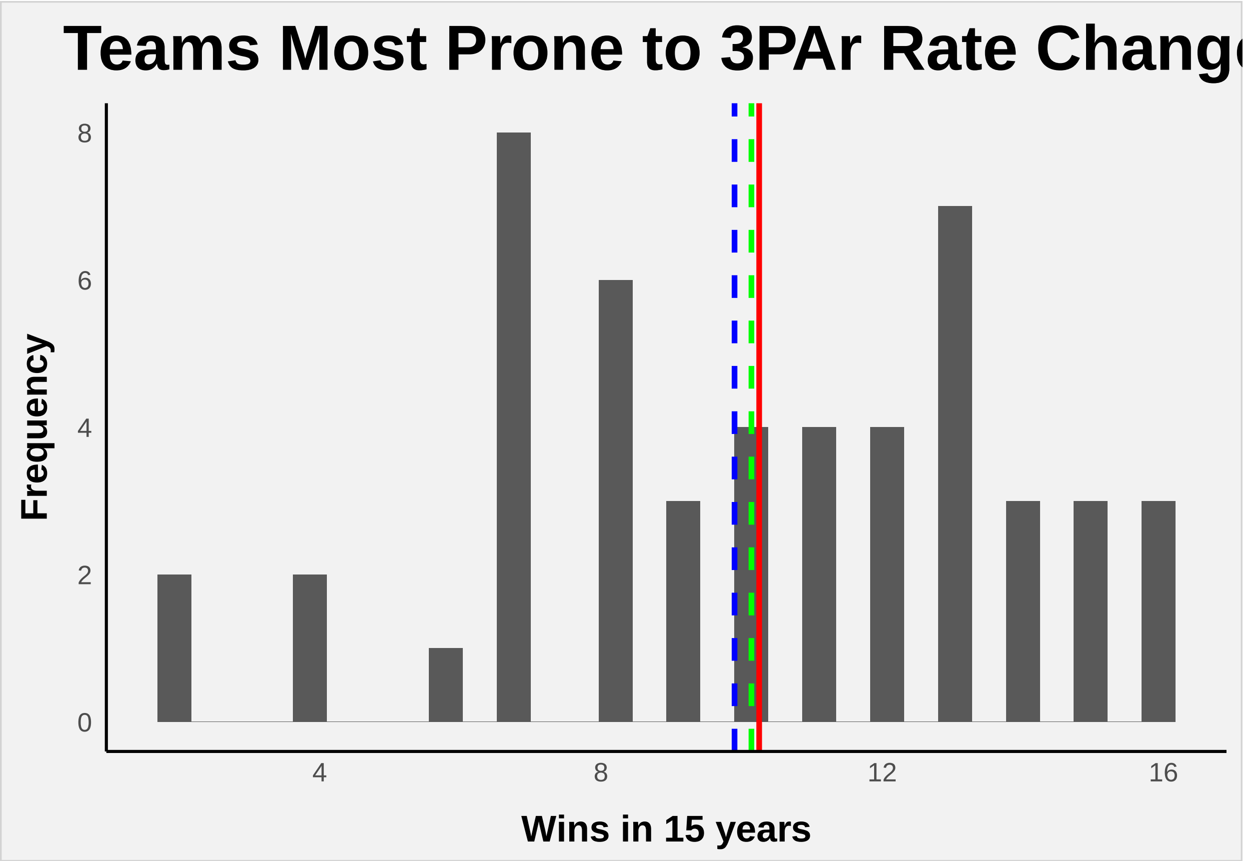

```
cols <- c("All Teams"="red","Most Prone"="green", "Least Prone"="blue")

ggplot(close_zero_least, aes(x = x.1)) +
  geom_histogram() +
  geom_vline(aes(xintercept=mean(total_3s_made$x), color="All Teams"), size=1) +
  geom_vline(aes(xintercept=mean(close_zero_least$x.1), color="Least Prone"), linetype="dashed", size=1)
  geom_vline(aes(xintercept=mean(close_zero_most$x.1), color="Most Prone"), linetype="dashed", size=1) +
  theme_hodp() +
  labs(title="Teams Least Prone to 3PAr Rate Change") +
  theme(legend.position = "right") +
  scale_color_manual("Means", values = cols) +
  xlab("3 Pointers Made") +
  ylab("Frequency")
```

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



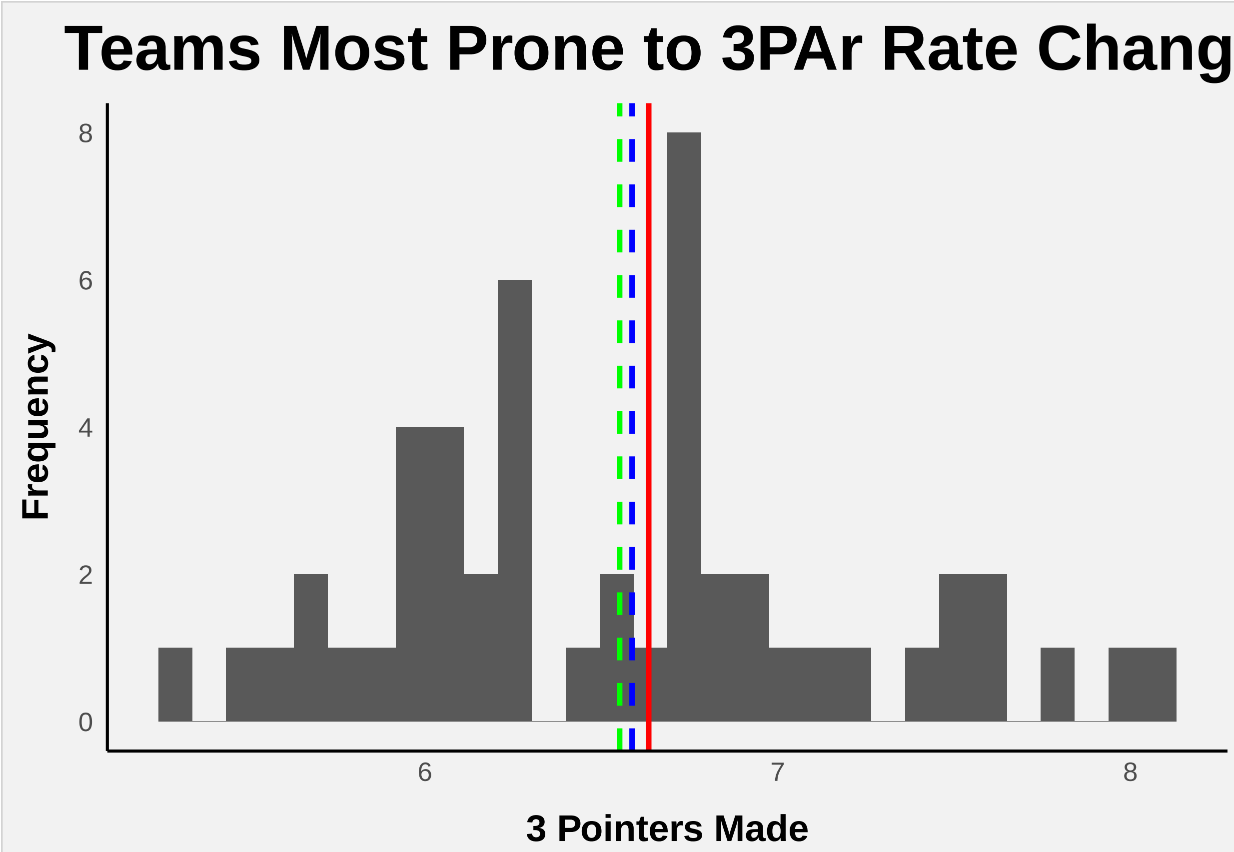

```r
# histogram teams most prone

ggplot(close_zero_most, aes(x = x)) +
  geom_histogram() +
  geom_vline(aes(xintercept=mean(total_wins$x)),
             color="red", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_most$x)),
             color="green",linetype="dashed", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_least$x)),
             color="blue",linetype="dashed", size=1) +
  theme_hodp() +
  labs(title="Teams Most Prone to 3PAr Rate Change") +
  xlab("Wins in 15 years") +
  ylab("Frequency")
```

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




```r
ggplot(close_zero_most, aes(x = x.1)) +
  geom_histogram() +
  geom_vline(aes(xintercept=mean(total_3s_made$x)),
             color="red", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_most$x.1)),
             color="green",linetype="dashed", size=1) +
  geom_vline(aes(xintercept=mean(close_zero_least$x.1)),
             color="blue",linetype="dashed", size=1) +
  theme_hodp() +
  labs(title="Teams Most Prone to 3PAr Rate Change") +
  theme(legend.position = c(0.8, 0.2)) +
  xlab("3 Pointers Made") +
  ylab("Frequency")
```

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



Comparing the Means with formal hypothesis testing

```r
# between the most and least affected teams - games
t.test(close_zero_least$x, close_zero_most$x)
```

```
##
##  Welch Two Sample t-test
##
## data:  close_zero_least$x and close_zero_most$x
## t = -0.33755, df = 97.98, p-value = 0.7364
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.650956  1.170956
## sample estimates:
## mean of x mean of y
##      9.90     10.14
```

```r
# between the msot and least affected - 3 pointers made per game
t.test(close_zero_least$x.1, close_zero_most$x.1)
```

```
##
##  Welch Two Sample t-test
##
## data:  close_zero_least$x.1 and close_zero_most$x.1
## t = 0.27396, df = 97.588, p-value = 0.7847
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2216405  0.2926328
## sample estimates:
## mean of x mean of y
##  6.587455  6.551959
```

Conducting a similar method on the piecewise linear model

```
# more tests on the piece slopes
piece.slopes = coef(lmer9a)$School
piece.slopes$time = abs(piece.slopes$time)
piece.slopes = piece.slopes[order(piece.slopes$time),]

era.zeroslopes = piece.slopes[, 2]
era.oneslopes = era.zeroslopes + piece.slopes[, 5]
era.twoslopes = era.zeroslopes + piece.slopes[, 6]

piece_df <- data.frame(school = school_list, slope0 = era.zeroslopes, slope1 = era.oneslopes, slope2 = e
piece_df0 = piece_df[order(piece_df$slope0),]
piece_df1 = piece_df[order(piece_df$slope1),]
piece_df2 = piece_df[order(piece_df$slope2),]
```

Comparing the Means for piecewise with formal hypothesis testing

```
piece_df0_least <- piece_df0[1:50, ]
piece_df0_most <- piece_df0[(nrow(piece_df0) - 50):nrow(piece_df0), ]
piece_df1_least <- piece_df1[1:50, ]
piece_df1_most <- piece_df1[(nrow(piece_df1) - 50):nrow(piece_df1), ]
piece_df2_least <- piece_df2[1:50, ]
piece_df2_most <- piece_df2[(nrow(piece_df2) - 50):nrow(piece_df2), ]

# t tests for wins
mean(total_wins$x)
```

```
## [1] 10.25
```

```
t.test(piece_df0_least$x, piece_df0_most$x)
```

```
##
##  Welch Two Sample t-test
##
## data:  piece_df0_least$x and piece_df0_most$x
## t = -2.7837, df = 86.811, p-value = 0.006595
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4838802 -0.5812178
## sample estimates:
## mean of x mean of y
##   9.34000  11.37255
```

```r
t.test(piece_df1_least$x, piece_df1_most$x)
```

```
##
##  Welch Two Sample t-test
##
## data:  piece_df1_least$x and piece_df1_most$x
## t = -2.7837, df = 86.811, p-value = 0.006595
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4838802 -0.5812178
## sample estimates:
## mean of x mean of y
##   9.34000  11.37255
```

```r
t.test(piece_df2_least$x, piece_df2_most$x)
```

```
##
##  Welch Two Sample t-test
##
## data:  piece_df2_least$x and piece_df2_most$x
## t = -2.7837, df = 86.811, p-value = 0.006595
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4838802 -0.5812178
## sample estimates:
## mean of x mean of y
##   9.34000  11.37255
```

```r
# t tests for 3 pointers made, p value 0.05
mean(total_3s_made$x)
```

```
## [1] 6.6343
```

```r
t.test(piece_df0_least$x.1, piece_df0_most$x.1)
```

```
##
##  Welch Two Sample t-test
##
## data:  piece_df0_least$x.1 and piece_df0_most$x.1
## t = 1.7417, df = 95.697, p-value = 0.08477
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03686786  0.56456499
## sample estimates:
## mean of x mean of y
##  6.794718  6.530870
```

```r
t.test(piece_df1_least$x.1, piece_df1_most$x.1)
```

```
##
##  Welch Two Sample t-test
##
## data:  piece_df1_least$x.1 and piece_df1_most$x.1
## t = 1.7417, df = 95.697, p-value = 0.08477
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03686786  0.56456499
```

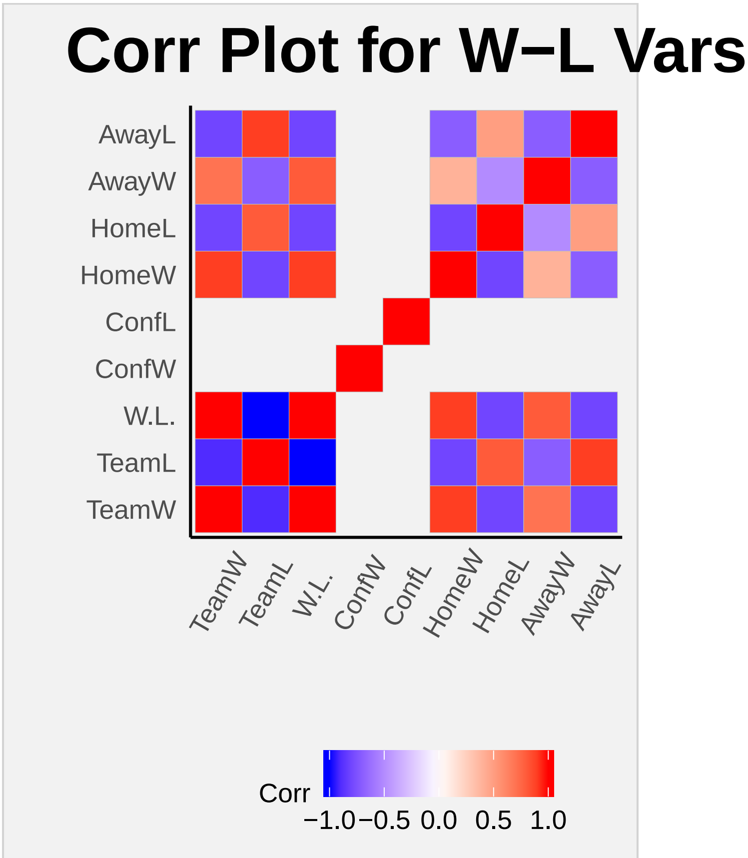
```
## sample estimates:
## mean of x mean of y
##  6.794718  6.530870
```

```
t.test(piece_df2_least$x.1, piece_df2_most$x.1)
```

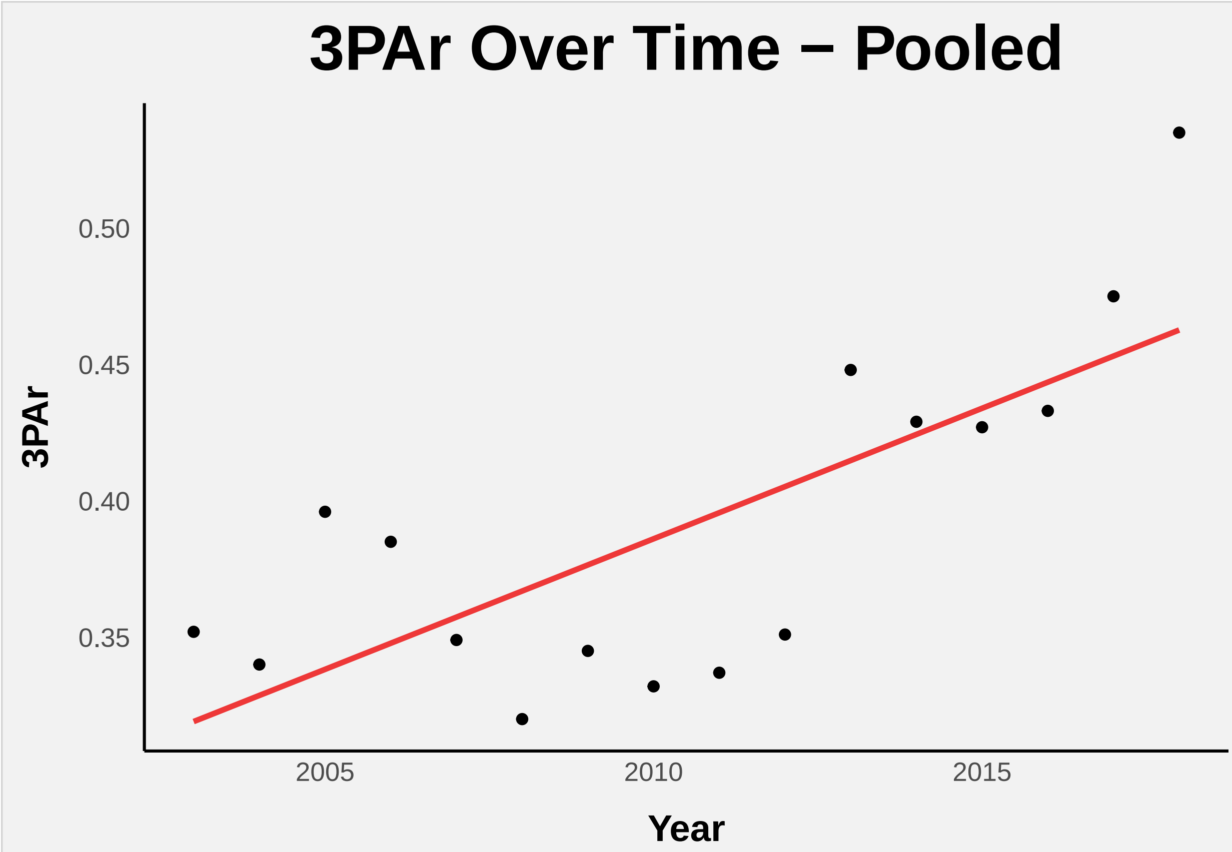
```
##
##  Welch Two Sample t-test
##
## data:  piece_df2_least$x.1 and piece_df2_most$x.1
## t = 1.7417, df = 95.697, p-value = 0.08477
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03686786  0.56456499
## sample estimates:
## mean of x mean of y
##  6.794718  6.530870
```

Comparisons between Schools

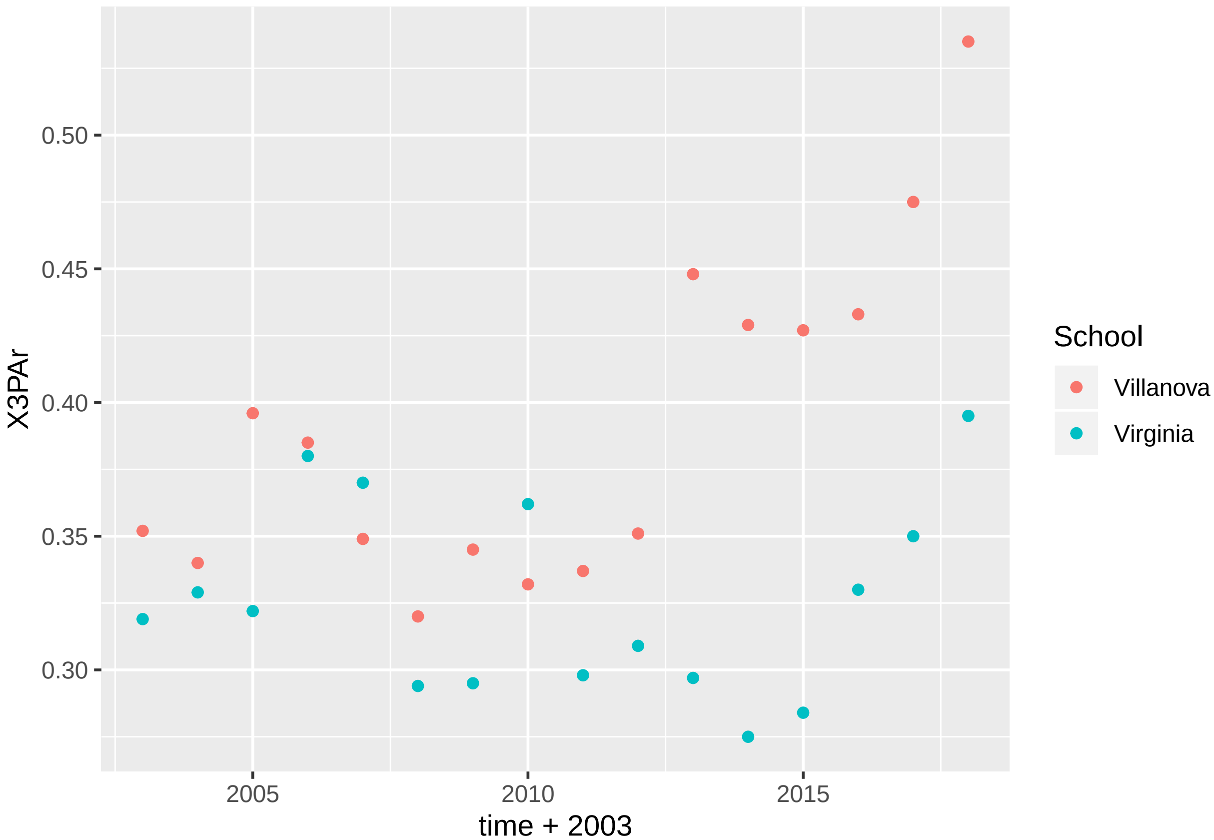
```
#eda correlation
data <- read.csv('data/full_data_raw.csv')
wl <- data %>% select(TeamW, TeamL, W.L., ConfW, ConfL, HomeW, HomeL, AwayW, AwayL)
cor <- round(cor(wl), 1)
p <- ggcorrplot(cor) +
  labs(title='Corr Plot for W-L Vars') +
  xlab('') + ylab('') +
  theme_hodp() +
  theme(axis.text.x=element_text(angle=60))
p
```




```r
#look at effects for individual teams
uva = df.tourney[df.tourney$School == 'Virginia',]
nova = df.tourney[df.tourney$School == 'Villanova',]
data_teams = rbind(uva, nova)
#interaction between school and time and time and era
data_teams$era <- as.factor((data_teams$year > 2006) + (data_teams$year > 2012))

#plot villanova and overall linear fit
p <- ggplot(nova, aes(x = time + 2003, y = X3PAr)) +
  geom_point() +
  stat_smooth(method = "lm", col = '#EE3838', se = F) +
  labs(title="3PAr Over Time - Pooled") +
  xlab("Year") +
  ylab("3PAr") +
  #ylim(c(0,0.6)) +
  theme_hodp()
p
```




```r
#plot villanova and uva by color
p <- ggplot(data_teams, aes(x = time + 2003, y = X3PAr, color = School)) +
  geom_point(aes(color = School))
p
```




```r
#interaction between time and school
lmer9 <- lmer(X3PAr ~ time*era + (1  + time|School), data=data_teams) # fails to converge
lm2 <- lm(X3PAr ~ time*School, data_teams)
summary(lm2)
```

```
##
## Call:
## lm(formula = X3PAr ~ time * School, data = data_teams)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.058660 -0.030860 -0.005394  0.032978  0.072346
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.319096   0.019212  16.610 4.97e-16 ***
## time                   0.009571   0.002182   4.386 0.000148 ***
## SchoolVirginia         0.006853   0.027169   0.252 0.802705
## time:SchoolVirginia -0.009622   0.003086  -3.118 0.004189 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04024 on 28 degrees of freedom
## Multiple R-squared:  0.5901, Adjusted R-squared:  0.5462
## F-statistic: 13.44 on 3 and 28 DF,  p-value: 1.287e-05
```

```r
names(df.clean)
```

```
##  [1] "School"     "G"          "TeamW"      "TeamL"      "W.L."
##  [6] "SRS"        "SOS"        "Tm."        "Opp."       "FTr"
## [11] "X3PAr"      "TS."        "TRB."       "AST."       "BLK."
## [16] "eFG."       "TOV."       "FT.FGA"     "FG"         "FGA"
## [21] "FG."        "X3P"        "X3PA"       "X3P."       "FT"
## [26] "FTA"        "FT."        "ORB"        "TRB"        "AST"
## [31] "STL"        "BLK"        "TOV"        "PF"         "year"
## [36] "time"       "same.coach"
```

```r
lm3 <- lm(X3PAr ~ time*(School + era), data_teams)
lmer10 <- lmer(X3PAr ~ time*era + (1  + time|School), data=data_teams) # fails to converge
```

```r
# Compare mixed model to lm
summary(lm3)
```

```
##
## Call:
## lm(formula = X3PAr ~ time * (School + era), data = data_teams)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04196 -0.01550 -0.00673  0.02005  0.06335
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.324624   0.021217  15.300 7.05e-14 ***
## time                  0.021361   0.009922   2.153 0.041598 *
## SchoolVirginia        0.006853   0.021031   0.326 0.747367
## era1                  0.018645   0.039896   0.467 0.644467
## era2                 -0.177502   0.068934  -2.575 0.016618 *
## time:SchoolVirginia -0.009622   0.002389  -4.028 0.000491 ***
## time:era1            -0.019093   0.011169  -1.709 0.100269
## time:era2             0.002593   0.011169   0.232 0.818392
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03115 on 24 degrees of freedom
## Multiple R-squared:  0.7895, Adjusted R-squared:  0.7281
## F-statistic: 12.86 on 7 and 24 DF,  p-value: 9.297e-07
```

```r
summary(lmer10)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: X3PAr ~ time * era + (1 + time | School)
##    Data: data_teams
##
## REML criterion at convergence: -87.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4080 -0.5117 -0.2005  0.6737  2.0573
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
```

```
##  School   (Intercept) 2.303e-05 0.004799
##           time        4.539e-05 0.006737 -1.00
##  Residual             9.315e-04 0.030520
## Number of obs: 32, groups:  School, 2
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  0.328050   0.018372  17.856
## time         0.016550   0.010763   1.538
## era1         0.018645   0.039090   0.477
## era2        -0.177502   0.067543  -2.628
## time:era1   -0.019093   0.010944  -1.745
## time:era2    0.002593   0.010944   0.237
##
## Correlation of Fixed Effects:
##           (Intr) time   era1   era2   tim:r1
## time      -0.788
## era1      -0.454  0.332
## era2      -0.263  0.192  0.123
## time:era1  0.695 -0.791 -0.731 -0.189
## time:era2  0.695 -0.791 -0.327 -0.639  0.778
```

```r
AIC(lm3)
```

```
## [1] -122.4083
```

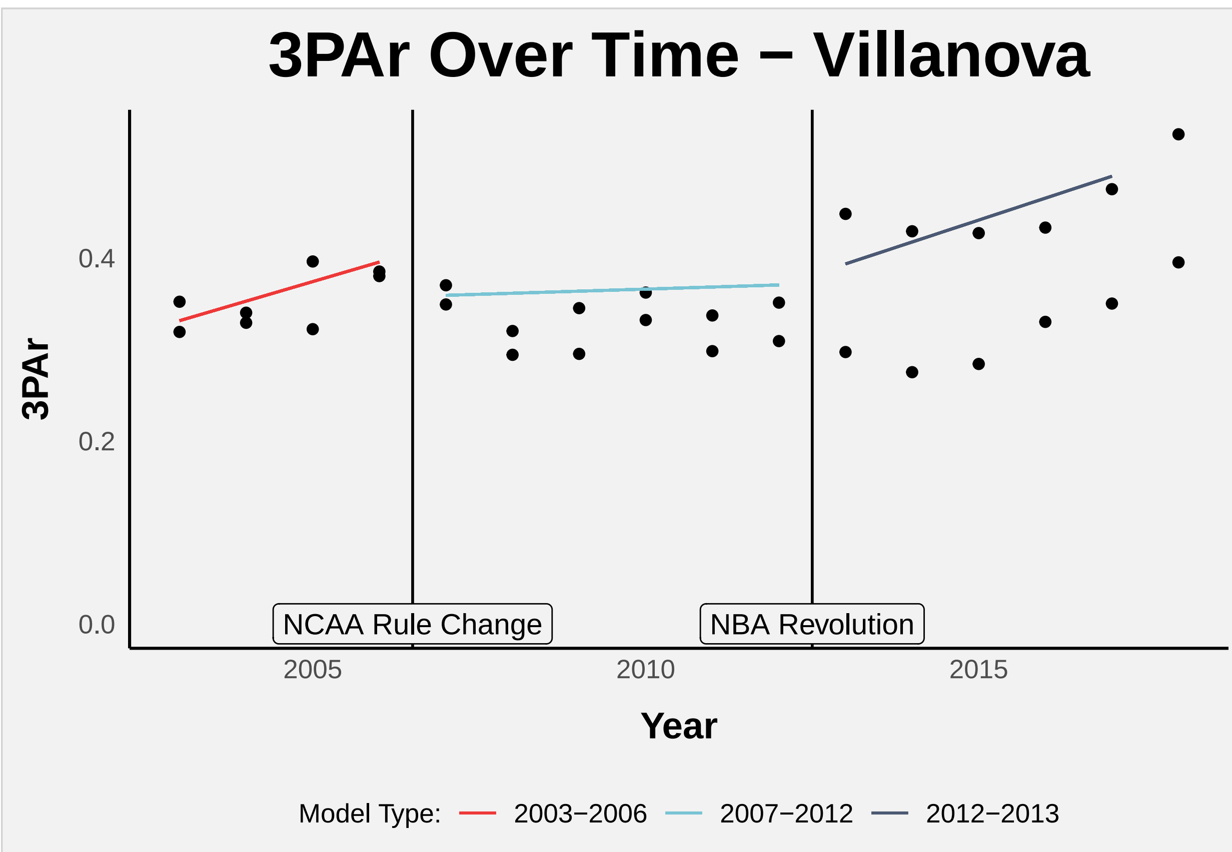
```r
AIC(lmer10)
```

```
## [1] -67.22047
```

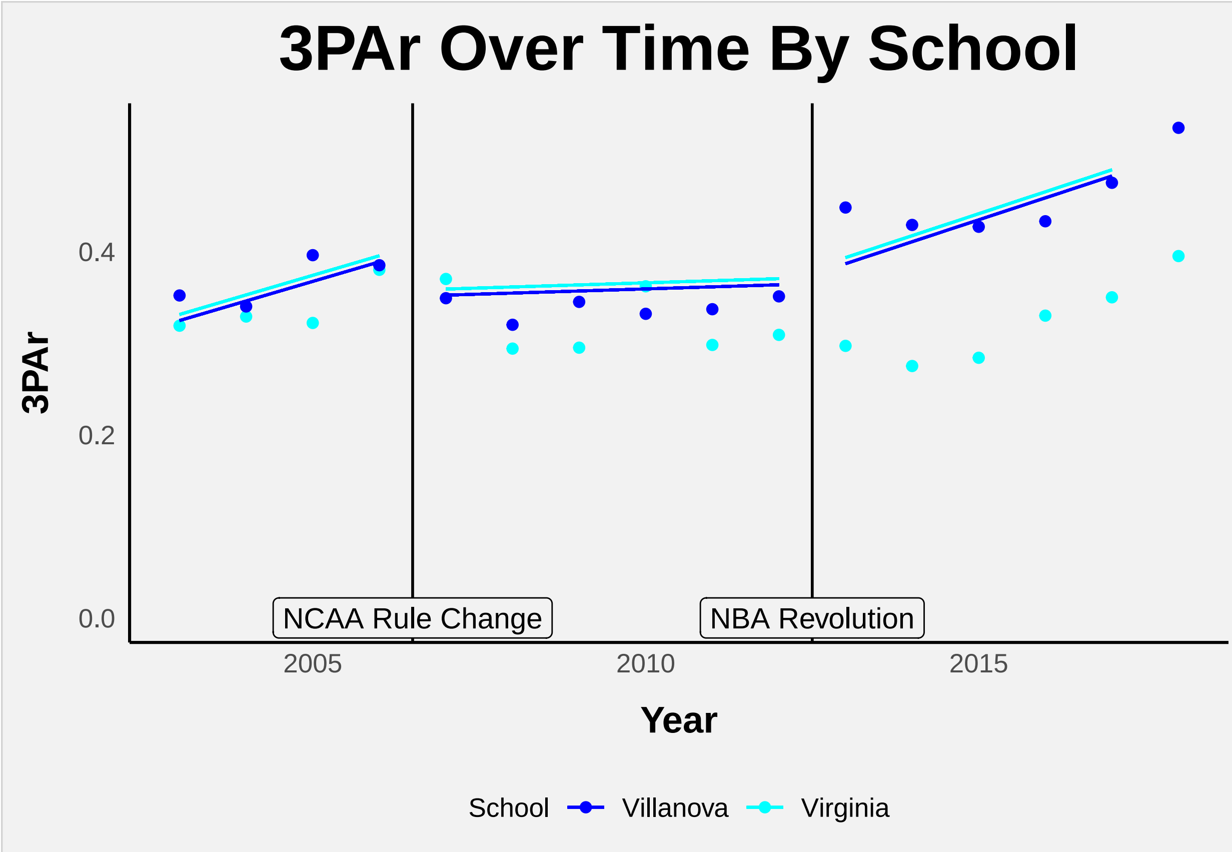
```r
# lm outperforms mixed model

intercept.0 <- summary(lm3)$coef[1,1]
slope.0 <- summary(lm3)$coef[2,1]
intercept.virginia <- summary(lm3)$coef[3,1]
intercept.1 <- summary(lm3)$coef[4,1]
intercept.2 <- summary(lm3)$coef[5,1]
slope.virginia <- summary(lm3)$coef[6,1]
slope.1 <- summary(lm3)$coef[7,1]
slope.2 <- summary(lm3)$coef[8,1]
```

```r
#plot lm3
year <- 2003:2017

fnnova <-  function(year, era1, era2) {
  return(intercept.0 + (year - 2003) * slope.0 +
           intercept.1*era1 + slope.1 * (year - 2003) * era1 +
           intercept.2*era2 + slope.2 * (year - 2003) * era2)
}

fnuva <- function(year, era1, era2) {
  return(intercept.0 + intercept.virginia + (year - 2003 + slope.virginia) * slope.0 +
           intercept.1*era1 + slope.1 * (year - 2003) * era1 +
           intercept.2*era2 + slope.2 * (year - 2003) * era2)
}
```

```
#plot interaction between school and time and time and era for just uva
p <- ggplot(data_teams, aes(x = time + 2003, y = X3PAr)) +
  geom_point() +
  geom_segment(aes(x = 2003, y = fnuva(2003,0,0), xend = 2006, yend = fnuva(2006,0,0),
                   colour = '#EE3838'),
               data = data_teams) +
  geom_segment(aes(x = 2007, y = fnuva(2007,1,0), xend = 2012, yend = fnuva(2012,1,0),
                   colour = '#78C4D4'),
               data = data_teams) +
  geom_segment(aes(x = 2013, y = fnuva(2013,0,1), xend = 2017, yend = fnuva(2017,0,1),
                   colour = '#4B5973'),
               data = data_teams) +
  geom_vline(xintercept = 2012.5) +
  geom_vline(xintercept = 2006.5) +
  scale_colour_identity(name="Model Type:",
                        breaks = c('#EE3838','#78C4D4','#4B5973'),
                        labels = c("2003-2006", "2007-2012", "2012-2013"),
                        guide = "legend") +
  annotate(geom="label", x = 2012.5, y = 0, label = "NBA Revolution", fill ="#F2F2F2", color = "black")
  annotate(geom="label", x = 2006.5, y = 0, label = "NCAA Rule Change", fill ="#F2F2F2", color = "black
  labs(title="3PAr Over Time - Villanova") +
  xlab("Year") +
  ylab("3PAr") +
  theme_hodp()
p
```



```r
#plot interaction between school and time and time and era for both schools
p <- ggplot(data_teams, aes(x = time + 2003, y = X3PAr, color = School)) +
  geom_point(aes(color = School)) +
  scale_color_manual(values =  c("#0000FF", "#00FFFF", "#0000FF", "#00FFFF"))+
  #geom_point(color = c('#EE3838', '#78C4D4')) +
  geom_segment(aes(x = 2003, y = fnnova(2003,0,0), xend = 2006, yend = fnnova(2006,0,0),
                   colour = 'Villanova'),
               data = data_teams) +
  geom_segment(aes(x = 2007, y = fnnova(2007,1,0), xend = 2012, yend = fnnova(2012,1,0),
                   colour = 'Villanova'),
               data = data_teams) +
  geom_segment(aes(x = 2013, y = fnnova(2013,0,1), xend = 2017, yend = fnnova(2017,0,1),
                   colour = 'Villanova'),
               data = data_teams) +

  geom_segment(aes(x = 2003, y = fnuva(2003,0,0), xend = 2006, yend = fnuva(2006,0,0),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_segment(aes(x = 2007, y = fnuva(2007,1,0), xend = 2012, yend = fnuva(2012,1,0),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_segment(aes(x = 2013, y = fnuva(2013,0,1), xend = 2017, yend = fnuva(2017,0,1),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_vline(xintercept = 2012.5) +
  geom_vline(xintercept = 2006.5) +
  annotate(geom="label", x = 2012.5, y = 0, label = "NBA Revolution", fill ="#F2F2F2", color = "black")
  annotate(geom="label", x = 2006.5, y = 0, label = "NCAA Rule Change", fill ="#F2F2F2", color = "black
  labs(title="3PAr Over Time By School") +
  xlab("Year") +
  ylab("3PAr") +
  theme_hodp()
p
```




```
#interaction between school and time and time and era and school and era
data_teams$era <- as.factor((data_teams$year > 2006) + (data_teams$year > 2012))
lm4 <- lm(X3PAr ~ time*(School + era) +School*era , data_teams)
summary(lm4)
```

```
##
## Call:
## lm(formula = X3PAr ~ time * (School + era) + School * era, data = data_teams)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.034627 -0.019078 -0.000164  0.017521  0.042160
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.3427594  0.0184098  18.618 5.92e-15 ***
## time                 0.0169937  0.0087054   1.952   0.0638 .
## SchoolVirginia      -0.0294187  0.0203126  -1.448   0.1616
## era1                 0.0098848  0.0381567   0.259   0.7980
## era2                -0.1297586  0.0667713  -1.943   0.0649 .
## time:SchoolVirginia -0.0008875  0.0058036  -0.153   0.8799
## time:era1           -0.0190929  0.0093065  -2.052   0.0523 .
## time:era2            0.0025929  0.0093065   0.279   0.7831
## SchoolVirginia:era1  0.0175208  0.0374620   0.468   0.6446
## SchoolVirginia:era2 -0.0954875  0.0680944  -1.402   0.1748
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02595 on 22 degrees of freedom
## Multiple R-squared:  0.866,  Adjusted R-squared:  0.8112
## F-statistic:  15.8 on 9 and 22 DF,  p-value: 1.14e-07
```

```
intercept.0 <- summary(lm4)$coef[1,1]
slope.0 <- summary(lm4)$coef[2,1]
intercept.virginia <- summary(lm4)$coef[3,1]
intercept.1 <- summary(lm4)$coef[4,1]
intercept.2 <- summary(lm4)$coef[5,1]
slope.virginia <- summary(lm4)$coef[6,1]
slope.1 <- summary(lm4)$coef[7,1]
slope.2 <- summary(lm4)$coef[8,1]
intercept.1.virginia <- summary(lm4)$coef[9,1]
intercept.2.virginia <- summary(lm4)$coef[10,1]


year <- 2003:2017
fnnova <-  function(year, era1, era2) {
  return(intercept.0 + (year - 2003) * slope.0 +
           intercept.1*era1 + slope.1 * (year - 2003) * era1 +
           intercept.2*era2 + slope.2 * (year - 2003) * era2)
}
fnuva <- function(year, era1, era2) {
  return(intercept.0 + intercept.virginia + (year - 2003 + slope.virginia) * slope.0 +
           (intercept.1 + intercept.1.virginia) *era1 + slope.1 * (year - 2003) * era1 +
           (intercept.2 + intercept.2.virginia) *era2 + slope.2 * (year - 2003) * era2)
}

p <- ggplot(data_teams, aes(x = time + 2003, y = X3PAr, color = School)) +
  geom_point(aes(color = School)) +
  scale_color_manual(values =  c("#EE3838", "#0000FF", "#EE3838", "#0000FF"))+
  #geom_point(color = c('#EE3838', '#78C4D4')) +
  geom_segment(aes(x = 2003, y = fnnova(2003,0,0), xend = 2006, yend = fnnova(2006,0,0),
                   colour = 'Villanova'),
               data = data_teams) +
  geom_segment(aes(x = 2007, y = fnnova(2007,1,0), xend = 2012, yend = fnnova(2012,1,0),
                   colour = 'Villanova'),
               data = data_teams) +
  geom_segment(aes(x = 2013, y = fnnova(2013,0,1), xend = 2017, yend = fnnova(2017,0,1),
                   colour = 'Villanova'),
               data = data_teams) +

  geom_segment(aes(x = 2003, y = fnuva(2003,0,0), xend = 2006, yend = fnuva(2006,0,0),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_segment(aes(x = 2007, y = fnuva(2007,1,0), xend = 2012, yend = fnuva(2012,1,0),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_segment(aes(x = 2013, y = fnuva(2013,0,1), xend = 2017, yend = fnuva(2017,0,1),
                   colour = 'Virginia'),
               data = data_teams) +
  geom_vline(xintercept = 2012.5) +
```

```
  geom_vline(xintercept = 2006.5) +
  annotate(geom="label", x = 2012.5, y = 0, label = "NBA Revolution", fill ="#F2F2F2", color = "black")
  annotate(geom="label", x = 2006.5, y = 0, label = "NCAA Rule Change", fill ="#F2F2F2", color = "black
  labs(title="3PAr Over Time By School") +
  xlab("Year") +
  ylab("3PAr") +
  theme_hodp()
p
```