posterproject_ajn873.R

alyssapacleb

2020-01-04

Research Question:

Controlling for salary, does the linear trend predicting number of home runs hit every year change through time? Additionally, does batting hand significantly moderate the effect of year on home runs and/or does league moderate the effect of year on home runs?

```
library(SDSRegressionR)
library(tidyverse)
library(mosaic)
library(emmeans)

#Bring in data
final <- read_csv("data/final_dataset.csv")
names(final)
```

```
## [1] "yearID"   "playerID" "name"     "b_HR"     "salary"   "lgID"
## [7] "bats"
```

```
tally(~lgID, data=final)
```

```
## lgID
##   AL   NL
## 6192 6534
```

```
tally(~bats, data=final)
```

```
## bats
##    B    L    R
##  363 3489 8874
```

```
final <- final %>%
  mutate(lgID_f = factor(lgID, levels=c("AL", "NL")),
         bats_f = factor(bats, levels=c("B", "L", "R")))
tally(~lgID_f, data=final)
```

```
## lgID_f
##   AL   NL
## 6192 6534
```

```
tally(~bats_f, data=final)
```

```
## bats_f
##    B    L    R
##  363 3489 8874
```

```
final$salary2 <-  final$salary/1000
# Set up
breaks <- seq(1985,2016,1)
rmse <- rep(NA, length(breaks))

for(i in 1:length(breaks)){
```

```
  final2 <- final %>%
    mutate_at(vars(yearID), as.numeric)%>% #Initial catch all for numeric...
    mutate(year1 = yearID, #Simplet replication
           year2 = yearID - breaks[i], #Start second segment counting...
           year2 = case_when(year1 <= breaks[i]~0, #Make sure to start at zero BEFORE segment
                             TRUE~ year2),
           jump = case_when(yearID < breaks[i]~0, #Define the segment status...
                             yearID >= breaks[i]~1))
  mod <- lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID_f*y
  rmse[i]<- summary(mod)$sigma #Save the RMSE
}

potential_breakpoints_rmse = data.frame(br = breaks, rmse = rmse)
min(potential_breakpoints_rmse$rmse) # The lowest RMSE is 0.9473682 with breakpoint of 2009
```

```
## [1] 0.9473682
```

2. Assign the coding of the variables according to the change point indicated.

```
cutoff <- 2009
coded_final <- final %>%
  mutate_at(vars(yearID), as.numeric) %>%
  mutate(year1 = yearID,
         year2 = yearID - cutoff,
         year2 = case_when(year1 <= cutoff ~ 0, TRUE ~ year2),
         jump = case_when(yearID < cutoff ~ 0, yearID >= cutoff ~ 1))
```

Double check the mutations

```
plyr::count(coded_final, c("yearID", "year1", "year2", "jump"))
```

```
##    yearID year1 year2 jump freq
## 1    1985  1985     0    0  226
## 2    1986  1986     0    0  331
## 3    1987  1987     0    0  282
## 4    1988  1988     0    0  282
## 5    1989  1989     0    0  326
## 6    1990  1990     0    0  378
## 7    1991  1991     0    0  310
## 8    1992  1992     0    0  339
## 9    1993  1993     0    0  437
## 10   1994  1994     0    0  398
## 11   1995  1995     0    0  497
## 12   1996  1996     0    0  445
## 13   1997  1997     0    0  462
## 14   1998  1998     0    0  486
## 15   1999  1999     0    0  486
## 16   2000  2000     0    0  410
## 17   2001  2001     0    0  405
## 18   2002  2002     0    0  396
## 19   2003  2003     0    0  399
## 20   2004  2004     0    0  398
## 21   2005  2005     0    0  405
## 22   2006  2006     0    0  417
## 23   2007  2007     0    0  417
```

```
## 24   2008  2008     0    0  417
## 25   2009  2009     0    1  412
## 26   2010  2010     1    1  406
## 27   2011  2011     2    1  413
## 28   2012  2012     3    1  427
## 29   2013  2013     4    1  423
## 30   2014  2014     5    1  405
## 31   2015  2015     6    1  447
## 32   2016  2016     7    1  444
```

```
favstats(~b_HR, data = coded_final)
```

```
##  min Q1 median Q3 max      mean        sd     n missing
##    0  0      0  0  33 0.1224265 0.9532597 12726       0
```

```
favstats(~salary2, data = coded_final)
```

```
##  min  Q1   median   Q3   max     mean      sd     n missing
##   60 295 547.1215 2250 33000 1937.489 3172.81 12726       0
```

```
tally(~lgID_f, data=coded_final)
```

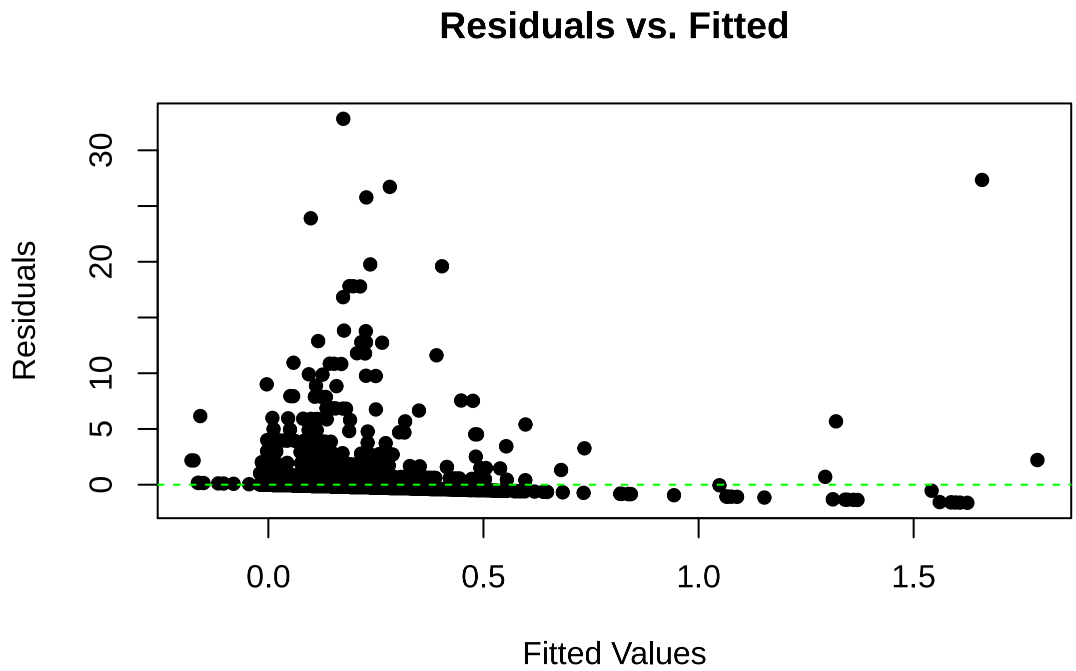
```
## lgID_f
##   AL   NL
## 6192 6534
```

```
tally(~bats_f, data=coded_final)
```

```
## bats_f
##    B    L    R
##  363 3489 8874
```

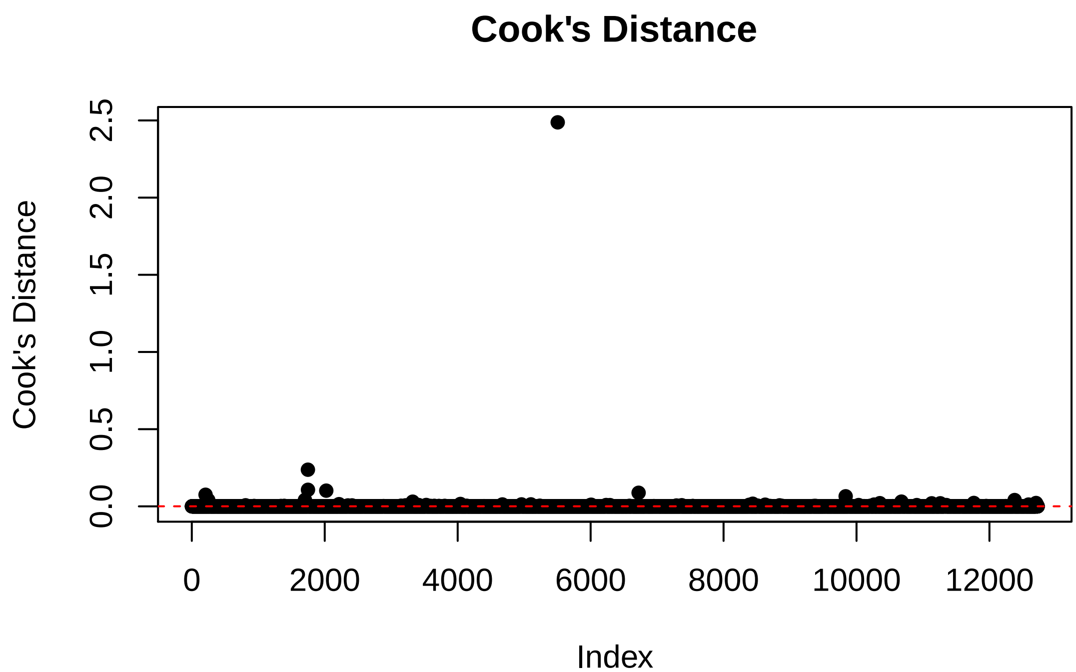
3. Look for outliers by running the full model.

```
init_mod = lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID_f
residFitted(init_mod)
```



I think it's important to be candid here and say that this Residuals vs Fitted plot isn't wonderful. However, I'm going to keep doing this analysis because it's good enough for government work.

```
cooks_plot = cooksPlot(init_mod, key.variable="playerID", print.obs = T, save.cutoff = T)
```




```
cooks_plot
```

```
##      playerID b_HR year1 jump year2   salary2 bats_f lgID_f  Predicted_Y
## 1   swishni01   29  2009    1     0  5400.000      B     AL  1.659182990
## 2   davisch01   27  1993    0     0  2400.000      B     AL  0.282187080
## 3   davisch02   33  2012    1     3   488.000      L     AL  0.174092129
## 4   escobed01    6  2016    1     7  2150.000      B     AL -0.158460084
```

```
## 5    rossco01  24  2009    1     0  2225.000      R      NL  0.098432178
## 6     dunnad01  20  2014    1     5 15000.000      L      AL  0.403577592
## 7   lopezfe01   7  2010    1     1  1000.000      B      NL  1.319714240
## 8     hallbi03  18  2010    1     1  8525.000      R      AL  0.213147299
## 9   cuddymi01  20  2011    1     2 10500.000      R      AL  0.236761039
## 10  wallati01  26  1987    0     0   765.000      R      NL  0.227687614
## 11  oquenjo01   7  1988    0     0   275.000      B      NL  0.350009222
## 12  larocad01  12  2015    1     6 12000.000      L      AL  0.390811506
## 13  snidetr01  13  2014    1     5  1200.000      L      NL  0.226957712
## 14  zambrca01   6  2006    0     0  6500.000      B      NL  0.597710496
## 15  montemi01   8  2016    1     7 14000.000      L      NL  0.448070733
## 16  robincl01  10  2015    1     6   525.000      L      NL  0.249553528
## 17  relafde01   8  2001    0     0   475.000      B      NL  0.475496178
## 18  francje02  13  2015    1     6   950.000      R      NL  0.115673195
## 19  zambrca01   4  2009    1     0 18750.000      B      NL  1.787827939
## 20  murphda07  13  2013    1     4  5775.000      L      AL  0.264355058
## 21  frankni01   3  2015    1     6  1021.800      B      AL  0.081404463
## 22  sardilu01   2  2016    1     7   512.000      B      AL -0.178824695
## 23  rominan01   2  2016    1     7   900.000      B      AL -0.174000843
## 24  ramiral03  10  2015    1     6 10000.000      R      AL  0.226782101
## 25  zeileto01  18  2002    0     0  6833.333      R      NL  0.197093517
## 26  wilsogl01  14  1987    0     0   662.400      R      NL  0.226412029
## 27  zambrca01   4  2008    0     0 16000.000      B      NL  0.734743407
## 28  milesaa01   4  2008    0     0  1400.000      B      NL  0.553227343
## 29  gloadro01   6  2009    1     0  1900.000      L      NL  0.080721412
## 30  whitety01   8  2016    1     7   507.500      R      AL  0.107824920
## 31  spiezsc01   4  2007    0     0  2100.000      B      NL  0.552468596
## 32  finlest01  14  2001    0     0  5375.000      L      NL  0.175596948
## 33  macdomi01   0  2009    1     0  2650.000      B      AL  1.624993320
## 34  lewissc02   0  2009    1     0   404.000      B      AL  1.597069685
## 35    lawva01  12  1987    0     0   525.000      R      NL  0.224703788
## 36  whitema01   6  1998    0     0   200.000      B      AL  0.317876834
## 37  martida01  11  1990    0     0   410.000      L      NL  0.169420524
## 38  romerjc01   0  2009    1     0  4250.000      B      NL  1.607555137
## 39  perezto03   5  2002    0     0   475.000      B      NL  0.484957750
## 40  macdomi01   0  2009    1     0  2650.000      B      NL  1.587662965
## 41  rodriwa01   0  2009    1     0  2600.000      B      NL  1.587041335
## 42   bellde01  18  2000    0     0  5000.000      R      NL  0.188438981
## 43  riversa01   0  2009    1     0   475.000      B      NL  1.560622045
## 44  wallati01  13  1989    0     0   950.000      R      NL  0.215849073
## 45  bonilbo01   5  2001    0     0   900.000      B      NL  0.480780036
## 46  arencjp01  10  2014    1     5  1800.000      R      AL  0.125775382
## 47  raburry01   8  2015    1     6  2500.000      R      AL  0.133537548
## 48  hamilje01  12  1989    0     0   150.000      R      NL  0.205902988
## 49  kellydo01   7  2011    1     2   423.000      L      AL  0.148752275
## 50  greenni01   6  2009    1     0   550.000      R      AL  0.114937916
## 51  gladdda01  11  1988    0     0   360.000      R      AL  0.058300201
## 52  bethach01   6  2016    1     7   511.200      R      NL  0.115733199
## 53  oneilpa01   7  1987    0     0    65.000      L      NL  0.180281676
## 54  gaettga01  17  1997    0     0  2100.000      R      NL  0.173592280
## 55  bogusbr01   7  2012    1     3   483.000      L      NL  0.156067883
## 56  gomesjo01   7  2015    1     6  4000.000      R      NL  0.153592647
## 57  reynocr01   6  1986    0     0   416.667      L      NL  0.189703947
## 58  murphda07   5  2015    1     6  6000.000      L      AL  0.316215863
```

```
## 59  jonesga02    5  2015    1     6  5000.000       L      AL  0.303783256
## 60  shuckja01    4  2016    1     7   521.000       L      AL  0.272629343
## 61  foleyto02    7  1989    0     0   320.000       L      NL  0.173351723
## 62  neshepa01    0  2010    1     1   625.000       B      AL  1.345926277
## 63   huffda01    0  2010    1     1   410.700       B      AL  1.343261969
## 64  rodriwa01    0  2010    1     1  5000.000       B      NL  1.369444668
## 65  romerjc01    0  2010    1     1  4250.000       B      NL  1.360120213
## 66  milesaa01    0  2010    1     1  2700.000       B      NL  1.340849672
## 67  moorety01    6  2015    1     6   518.200       R      NL  0.110304796
## 68   mockga01    0  2010    1     1   411.000       B      NL  1.312391434
## 69  medlekr01    0  2010    1     1   407.500       B      NL  1.312347920
## 70  mathije01    8  2012    1     3  1500.000       R      AL  0.123926917
## 71  menecfr01    9  2004    0     0   400.000       R      AL -0.003963463
## 72  cansejo01   10  1993    0     0  4800.000       R      AL  0.093888173
## 73  hamptmi01    7  2001    0     0 10500.000       R      NL  0.249749033
## 74  rominan01    2  2014    1     5   504.000       B      AL  0.328857873
## 75  martean01    5  2010    1     1   413.400       R      AL  0.112298963
## 76  kunkeje01    8  1989    0     0    80.000       R      AL  0.050896510
## 77  gladdda01    8  1989    0     0   610.000       R      AL  0.057485792
## 78  zeileto01    9  2004    0     0  1000.000       R      NL  0.110431407
## 79  mcdonda02    6  2011    1     2   470.000       R      AL  0.112061990
## 80  johnsjo09    3  2009    1     0  1400.000       L      NL  0.074505108
## 81  jimenda01    3  2002    0     0   240.000       B      NL  0.482036088
## 82  gimench01    4  2016    1     7   975.000       R      AL  0.113637164
## 83  barneda01    4  2016    1     7  1050.000       R      AL  0.114569609
## 84  buterdr01    4  2016    1     7  1162.500       R      AL  0.115968277
## 85  bumgama01    5  2015    1     6  6750.000       R      NL  0.187782316
## 86  gracema01    7  2002    0     0  3000.000       L      NL  0.141019373
## 87  syndeno01    3  2016    1     7   535.375       L      NL  0.280670341
## 88  goinsry01    3  2016    1     7   520.200       L      AL  0.272619397
## 89  flahery01    3  2016    1     7  1500.000       L      AL  0.284800866
## 90  gaettga01   11  1998    0     0   170.000       R      NL  0.142528062
## 91  gaettga01   11  1998    0     0  1000.000       R      NL  0.152847125
## 92  overbly01    4  2014    1     5  1500.000       L      NL  0.230687494
## 93  reckean01    6  2013    1     4   490.840       R      NL  0.098933776
## 94  milesaa01    2  2007    0     0  1000.000       B      NL  0.538792729
## 95  rodriwa01    0  2011    1     2  7500.000       B      NL  1.153091262
## 96  gallayo01    4  2010    1     1   450.000       R      NL  0.081879732
## 97  venturo01    5  2004    0     0  1200.000       L      NL  0.108540412
## 98    lawva01    5  1986    0     0   450.000       R      NL  0.230840629
## 99  zambrca01    2  2007    0     0 12400.000       B      NL  0.680524449
## 100 hollade01    0  2011    1     2   431.810       B      AL  1.089633407
## 101 amarial01    3  2015    1     6  1150.000       L      NL  0.257323907
## 102 romerjc01    0  2011    1     2  1350.000       B      NL  1.076630728
## 103 figuene01    0  2011    1     2   900.000       B      NL  1.071036055
## 104 neshepa01    0  2011    1     2   625.000       B      NL  1.067617088
## 105 macdomi01    0  2011    1     2   500.000       B      NL  1.066063012
## 106 medlekr01    0  2011    1     2   429.500       B      NL  1.065186513
## 107 storedr01    0  2011    1     2   418.000       B      NL  1.065043538
## 108 janseke01    0  2011    1     2   416.000       B      NL  1.065018673
## 109 davisik02    3  2015    1     6  3800.000       L      AL  0.288864128
## 110 gaettga01    9  1999    0     0  2000.000       R      NL  0.158210446
## 111 owingmi01    3  2009    1     0   420.000       R      NL  0.075991322
## 112 maynebr01    6  2000    0     0  1750.000       L      NL  0.135578881
```

```
## 113 bumgama01    3  2016    1      7  9916.667       R      NL   0.232667675
## 114 mathejo02    5  2012    1      3   490.000       R      NL   0.093407900
## 115 nunezab01    2  2004    0      0   625.000       B      NL   0.505745785
## 116 pecotbi01    6  1991    0      0   307.500       R      AL   0.045879807
## 117 maldoma01    4  2014    1      5   502.000       R      NL   0.104587955
## 118 zambrca01    2  2003    0      0   340.000       B      NL   0.492740920
## 119 lyonsst01    4  1991    0      0   650.000       L      AL   0.019511473
## 120 bumgama01    4  2014    1      5  3750.000       R      NL   0.144969063
## 121 bogarti01    7  2000    0      0   700.000       R      NL   0.134978770
## 122 greinza01    2  2015    1      6 25000.000       R      NL   0.414677396
## 123 rosalad01    3  2015    1      6   900.000       R      AL   0.113645377
## 124 martida01    5  1995    0      0   650.000       L      AL   0.011897842
## 125 ceronri01    4  1987    0      0   250.000       R      AL   0.060855174
## 126 menecfr01    6  2000    0      0   201.000       R      AL   0.009252690
## 127 denorch01    3  2015    1      6  2600.000       R      NL   0.136186997
## 128 wainwad01    2  2016    1      7 19500.000       R      NL   0.351813489
## 129 espinal01    5  1991    0      0   650.000       R      AL   0.050137975
## 130 rodriwa01    0  2012    1      3 10500.000       B      NL   0.942954159
## 131 willido03    3  2006    0      0  4350.000       L      NL   0.137602857
## 132 francte01    3  1989    0      0   245.000       L      AL   0.018283083
## 133 zambrca01    2  2011    1      2 18875.000       B      NL   1.294512167
## 134 cangejo01    2  1995    0      0   182.500       B      NL   0.415090209
## 135 zambrca01    1  2010    1      1 18875.000       B      NL   1.541947091
## 136 mccarda01    4  2004    0      0   500.000       R      AL  -0.002720203
## 137 hollade01    0  2012    1      3  1000.000       B      AL   0.842806475
## 138 romerjc01    0  2012    1      3   750.000       B      AL   0.839698324
## 139  huffda01    0  2012    1      3   486.200       B      AL   0.836418602
## 140 pattotr01    0  2012    1      3   483.500       B      AL   0.836385034
## 141  gottji01    3  1985    0      0   170.000       R      NL   0.234428786
## 142 holadbr01    2  2016    1      7   519.000       R      AL   0.107967895
## 143 gallayo01    2  2009    1      0   414.000       R      NL   0.075916727
## 144 buterdr01    3  2014    1      5   700.000       R      NL   0.107049612
## 145 romerjc01    0  2012    1      3   750.000       B      NL   0.821736240
## 146 macdomi01    0  2012    1      3   650.000       B      NL   0.820492979
## 147  wolfra02    3  2004    0      0  4375.000       L      NL   0.148013940
## 148 storedr01    0  2012    1      3   498.750       B      NL   0.818612548
## 149 janseke01    0  2012    1      3   491.000       B      NL   0.818516195
## 150 medlekr01    0  2012    1      3   490.000       B      NL   0.818503762
## 151  shawbr01    0  2012    1      3   483.000       B      NL   0.818416734
## 152  lynnla01    0  2012    1      3   482.000       B      NL   0.818404301
## 153 harrelu01    0  2012    1      3   482.000       B      NL   0.818404301
## 154 castile01    0  2012    1      3   480.000       B      NL   0.818379436
## 155 defraju01    0  2012    1      3   480.000       B      NL   0.818379436
## 156 wainwad01    2  2009    1      0  2787.500       R      NL   0.105425520
## 157 arrieja01    2  2016    1      7 10700.000       R      NL   0.242406546
## 158 francma01    4  1999    0      0   250.000       L      NL   0.121980104
## 159 boggswa01    4  1997    0      0  2000.000       L      AL   0.024875046
## 160  woodtr01    3  2014    1      5  3900.000       R      NL   0.146833954
## 161   leecl02    2  2011    1      2 11000.000       L      NL   0.255833786
## 162 dempsri01    4  1991    0      0   150.000       R      AL   0.043921672
## 163 garcile02    1  2014    1      5   505.500       B      AL   0.328876522
## 164 salazlu01    3  1987    0      0    75.000       R      NL   0.219109115
## 165 schatda01    2  1985    0      0   375.000       L      NL   0.194236051
## 166  woodja02    3  2007    0      0   390.000       R      NL   0.081639657
```

```
## 167 rodriwa01    0  2013    1     4 13500.000       B      NL  0.732817057
## 168 schumsk01    2  2011    1     2  2750.000       L      NL  0.153264778
## 169 harrijo05    3  2013    1     4   503.000       R      NL  0.099084956
## 170 hamptmi01    3  2002    0     0  9503.543       R      NL  0.230291188
## 171  woodtr01    3  2013    1     4   527.500       R      NL  0.099389555
## 172  dukeza01    2  2011    1     2  3500.000       L      NL  0.162589233
## 173 marquja01    2  2008    0     0  6375.000       L      NL  0.152678619
## 174  iorgda01    2  1986    0     0   210.000       L      NL  0.187134538
## 175 schumsk01    2  2014    1     5  2000.000       L      NL  0.236903798
## 176  bluevi01    1  1986    0     0   450.000       B      NL  0.333261784
## 177 willido03    2  2007    0     0  6450.000       L      NL  0.158661199
## 178 seitzke01    4  1993    0     0   109.000       R      AL  0.035566814
## 179 morelmi01    2  2014    1     5  2650.000       L      AL  0.250034895
## 180 mosesjo01    1  1989    0     0   180.000       B      AL  0.204153501
## 181 seitzke01    4  1993    0     0   600.000       R      AL  0.041671224
## 182 kershcl01    0  2016    1     7 33000.000       L      NL  0.684290268
## 183  sosajo02    3  2006    0     0  2200.000       R      NL  0.111211962
## 184 donnech01    3  2001    0     0   300.000       L      NL  0.112501467
## 185 hamptmi01    2  2004    0     0 14625.000       R      NL  0.279825678
## 186 sheldsc01    4  2000    0     0   200.000       R      AL  0.009240257
## 187 robindo01    3  1989    0     0   900.000       R      NL  0.215227443
## 188 lakerti01    3  2004    0     0   450.000       R      AL -0.003341833
## 189 oquenjo01    1  1987    0     0   100.000       B      NL  0.338371943
## 190 arrieja01    2  2015    1     6  3630.000       R      NL  0.148992582
## 191 maiermi01    2  2012    1     3   865.000       L      AL  0.178779222
## 192 romerjc01    0  2008    0     0  3250.000       B      NL  0.576227667
## 193 schumsk01    2  2013    1     4  1500.000       L      NL  0.199699669
## 194 hamptmi01    2  2003    0     0 13625.000       R      NL  0.274462358
##     Cooks_Distance        F_per
## 1     2.4876008747 9.993816e-01
## 2     0.2374017164 5.507155e-04
## 3     0.1072292032 1.699645e-06
## 4     0.1016790483 1.128092e-06
## 5     0.0879407745 3.643658e-07
## 6     0.0752063019 1.061419e-07
## 7     0.0651429169 3.378076e-08
## 8     0.0411265646 8.126292e-10
## 9     0.0410819397 8.054374e-10
## 10    0.0406992549 7.460102e-10
## 11    0.0305383792 7.013753e-11
## 12    0.0302200374 6.431572e-11
## 13    0.0219201297 4.470581e-12
## 14    0.0215373946 3.860137e-12
## 15    0.0204742423 2.530680e-12
## 16    0.0197547430 1.877383e-12
## 17    0.0195560249 1.725389e-12
## 18    0.0169105898 5.117843e-13
## 19    0.0164790347 4.121761e-13
## 20    0.0153364692 2.257367e-13
## 21    0.0142143840 1.193524e-13
## 22    0.0127420158 4.764647e-14
## 23    0.0126805977 4.575064e-14
## 24    0.0122416809 3.402581e-14
## 25    0.0121285630 3.147116e-14
```

```
## 26    0.0116139145 2.185506e-14
## 27    0.0115437201 2.076848e-14
## 28    0.0114085458 1.880917e-14
## 29    0.0107353295 1.127372e-14
## 30    0.0102817346 7.837900e-15
## 31    0.0099371881 5.881848e-15
## 32    0.0096864921 4.742615e-15
## 33    0.0087787307 2.069054e-15
## 34    0.0084928951 1.565023e-15
## 35    0.0084807716 1.546278e-15
## 36    0.0083854022 1.405579e-15
## 37    0.0083530115 1.360427e-15
## 38    0.0082798357 1.263094e-15
## 39    0.0081153619 1.066379e-15
## 40    0.0080762965 1.023830e-15
## 41    0.0080701218 1.017243e-15
## 42    0.0080112755 9.563190e-16
## 43    0.0078176413 7.778975e-16
## 44    0.0075296343 5.666417e-16
## 45    0.0070471154 3.239083e-16
## 46    0.0068189938 2.453024e-16
## 47    0.0064916280 1.618703e-16
## 48    0.0063989910 1.433558e-16
## 49    0.0063902346 1.417064e-16
## 50    0.0059623411 7.887765e-17
## 51    0.0059577004 7.836010e-17
## 52    0.0053577602 3.193504e-17
## 53    0.0051159801 2.160748e-17
## 54    0.0048610366 1.401998e-17
## 55    0.0047054983 1.064667e-17
## 56    0.0045643505 8.227236e-18
## 57    0.0042518690 4.513304e-18
## 58    0.0041596634 3.748565e-18
## 59    0.0041457452 3.643667e-18
## 60    0.0041133018 3.409131e-18
## 61    0.0038514978 1.953176e-18
## 62    0.0038105394 1.784061e-18
## 63    0.0037967897 1.730258e-18
## 64    0.0037825995 1.676236e-18
## 65    0.0037289937 1.485331e-18
## 66    0.0036237172 1.165344e-18
## 67    0.0035862947 1.067227e-18
## 68    0.0034809455 8.289739e-19
## 69    0.0034807381 8.285555e-19
## 70    0.0033665997 6.246262e-19
## 71    0.0031949377 4.008579e-19
## 72    0.0031791999 3.844269e-19
## 73    0.0029683029 2.148470e-19
## 74    0.0029382704 1.971010e-19
## 75    0.0027526786 1.133560e-19
## 76    0.0027194224 1.022549e-19
## 77    0.0027095004 9.913422e-20
## 78    0.0026603967 8.489302e-20
## 79    0.0026564921 8.384225e-20
```

```
## 80     0.0026278731 7.648433e-20
## 81     0.0025259247 5.468216e-20
## 82     0.0024682016 4.494677e-20
## 83     0.0024633686 4.420577e-20
## 84     0.0024563137 4.314346e-20
## 85     0.0023779439 3.276991e-20
## 86     0.0023588124 3.059981e-20
## 87     0.0023245368 2.702699e-20
## 88     0.0022023295 1.709393e-20
## 89     0.0021536998 1.414416e-20
## 90     0.0021002585 1.142873e-20
## 91     0.0019921951 7.300589e-21
## 92     0.0018956575 4.789777e-21
## 93     0.0018348634 3.632362e-21
## 94     0.0017860141 2.888953e-21
## 95     0.0016669734 1.608815e-21
## 96     0.0016194083 1.258342e-21
## 97     0.0016018696 1.147252e-21
## 98     0.0015911806 1.083874e-21
## 99     0.0015744674 9.909566e-22
## 100    0.0015463984 8.506348e-22
## 101    0.0015441778 8.403222e-22
## 102    0.0014430278 4.728116e-22
## 103    0.0014294695 4.363998e-22
## 104    0.0014213506 4.158006e-22
## 105    0.0014177013 4.068245e-22
## 106    0.0014156543 4.018648e-22
## 107    0.0014153211 4.010627e-22
## 108    0.0014152632 4.009234e-22
## 109    0.0013787226 3.210845e-22
## 110    0.0013744307 3.126974e-22
## 111    0.0013479716 2.651269e-22
## 112    0.0013160893 2.163707e-22
## 113    0.0013016840 1.970688e-22
## 114    0.0012775712 1.681410e-22
## 115    0.0012011246 9.957531e-23
## 116    0.0011281771 5.849264e-23
## 117    0.0010537250 3.275838e-23
## 118    0.0010506710 3.196083e-23
## 119    0.0009695433 1.615252e-23
## 120    0.0009474944 1.328597e-23
## 121    0.0009237038 1.070530e-23
## 122    0.0009171127 1.007362e-23
## 123    0.0009038555 8.901797e-24
## 124    0.0009031783 8.845315e-24
## 125    0.0008905920 7.851446e-24
## 126    0.0008113813 3.559217e-24
## 127    0.0008044149 3.307851e-24
## 128    0.0007964674 3.040321e-24
## 129    0.0007773661 2.473844e-24
## 130    0.0007735764 2.373256e-24
## 131    0.0007530083 1.887700e-24
## 132    0.0007381984 1.594613e-24
## 133    0.0006946990 9.519185e-25
```

```
## 134   0.0006701989 7.016824e-25
## 135   0.0006383363 4.639119e-25
## 136   0.0006280598 4.041632e-25
## 137   0.0006195869 3.601186e-25
## 138   0.0006154829 3.403499e-25
## 139   0.0006112215 3.208426e-25
## 140   0.0006111782 3.206498e-25
## 141   0.0006079693 3.066258e-25
## 142   0.0005907648 2.402681e-25
## 143   0.0005837446 2.170648e-25
## 144   0.0005753077 1.918113e-25
## 145   0.0005678206 1.716099e-25
## 146   0.0005663096 1.677687e-25
## 147   0.0005653471 1.653615e-25
## 148   0.0005640424 1.621474e-25
## 149   0.0005639268 1.618653e-25
## 150   0.0005639119 1.618290e-25
## 151   0.0005638076 1.615747e-25
## 152   0.0005637927 1.615385e-25
## 153   0.0005637927 1.615385e-25
## 154   0.0005637629 1.614659e-25
## 155   0.0005637629 1.614659e-25
## 156   0.0005511834 1.332968e-25
## 157   0.0005435264 1.183596e-25
## 158   0.0005399625 1.119258e-25
## 159   0.0005236087 8.618904e-26
## 160   0.0005193013 8.034756e-26
## 161   0.0005035475 6.184489e-26
## 162   0.0004994549 5.770212e-26
## 163   0.0004738796 3.691786e-26
## 164   0.0004727933 3.620496e-26
## 165   0.0004675647 3.294074e-26
## 166   0.0004616718 2.957547e-26
## 167   0.0004515692 2.450692e-26
## 168   0.0004443724 2.137989e-26
## 169   0.0004430565 2.084790e-26
## 170   0.0004425356 2.064056e-26
## 171   0.0004422407 2.052398e-26
## 172   0.0004395331 1.948049e-26
## 173   0.0004338029 1.742514e-26
## 174   0.0004137776 1.166209e-26
## 175   0.0004106608 1.093642e-26
## 176   0.0004082086 1.039379e-26
## 177   0.0003821945 5.939899e-27
## 178   0.0003795959 5.605347e-27
## 179   0.0003776528 5.366173e-27
## 180   0.0003762348 5.197355e-27
## 181   0.0003730743 4.837839e-27
## 182   0.0003724152 4.765697e-27
## 183   0.0003718100 4.700288e-27
## 184   0.0003699010 4.499131e-27
## 185   0.0003658371 4.096012e-27
## 186   0.0003600885 3.580252e-27
## 187   0.0003570773 3.333680e-27
```

```
## 188   0.0003545175 3.135977e-27
## 189   0.0003494344 2.773820e-27
## 190   0.0003335068 1.866005e-27
## 191   0.0003262380 1.547354e-27
## 192   0.0003191507 1.283909e-27
## 193   0.0003175129 1.228987e-27
## 194   0.0003169836 1.211686e-27
```

There are two very distinctly different outliers. I'm going to remove them. Understandably, not much should change in a 12726 subject dataset.

```r
final %>%
  filter(playerID %in% c("davisch01", "swishni01"))
```

```
## # A tibble: 2 x 10
##   yearID playerID  name       b_HR salary lgID  bats  lgID_f bats_f salary2
##    <dbl> <chr>     <chr>     <dbl>  <dbl> <chr> <chr> <fct>  <fct>    <dbl>
## 1   1993 davisch01 Chili Da~    27 2.40e6 AL    B     AL     B         2400
## 2   2009 swishni01 Nick Swi~    29 5.40e6 AL    B     AL     B         5400
```

```r
good_final <- coded_final %>%
  filter(playerID %not_in% c("davisch01", "swishni01"))

mod2 = lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID_f*yea
summary(mod2)
```

```
##
## Call:
## lm(formula = b_HR ~ year1 + jump + year2 + salary2 + bats_f *
##     year1 + bats_f * jump + bats_f * year2 + lgID_f * year1 +
##     lgID_f * jump + lgID_f * year2, data = good_final)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.647 -0.163 -0.110 -0.028 32.842
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.223e+01  1.622e+01  -2.604  0.00923 **
## year1           2.125e-02  8.123e-03   2.616  0.00891 **
## jump           -1.866e-01  2.248e-01  -0.830  0.40646
## year2          -1.020e-02  4.542e-02  -0.225  0.82232
## salary2         1.147e-05  2.634e-06   4.356 1.33e-05 ***
## bats_fL         4.531e+01  1.681e+01   2.695  0.00704 **
## bats_fR         4.940e+01  1.630e+01   3.030  0.00245 **
## lgID_fNL        7.450e+00  5.564e+00   1.339  0.18060
## year1:bats_fL  -2.279e-02  8.418e-03  -2.707  0.00679 **
## year1:bats_fR  -2.483e-02  8.163e-03  -3.042  0.00236 **
## jump:bats_fL    2.629e-01  2.296e-01   1.145  0.25215
## jump:bats_fR    2.812e-01  2.240e-01   1.255  0.20938
## year2:bats_fL   4.377e-02  4.660e-02   0.939  0.34764
## year2:bats_fR   2.053e-02  4.550e-02   0.451  0.65182
## year1:lgID_fNL -3.663e-03  2.786e-03  -1.315  0.18859
## jump:lgID_fNL  -4.827e-02  6.763e-02  -0.714  0.47540
## year2:lgID_fNL -4.190e-03  1.354e-02  -0.309  0.75697
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8829 on 12707 degrees of freedom
## Multiple R-squared:  0.01014,    Adjusted R-squared:  0.008897
## F-statistic: 8.139 on 16 and 12707 DF,  p-value: < 2.2e-16
```

```
p1 <- summary(emmeans(mod2, "year1", at=list(year1=c(1984, cutoff), year2=0, jump=0)))
p2 <- summary(emmeans(mod2, "year1", at=list(year1=c(cutoff, 2017), year2=c(0, (2017-cutoff)), jump=1),
p2 <- p2 %>%
  slice(c(1,4))

mns1 <- summary(emmeans(mod2, "year1", at=list(year1 = seq(1984,cutoff,1)), year2 = 0, jump = 0, by="lg
mns2 <- summary(emmeans(mod2, "year1", at=list(year1 = seq(cutoff,2017,1)), year2 = c(0, (2017-cutoff))

mns3 <- summary(emmeans(mod2, "year1", at=list(year1 = seq(1984,cutoff,1)), year2 = 0, jump = 0, by="ba
mns4 <- summary(emmeans(mod2, "year1", at=list(year1 = seq(cutoff,2017,1)), year2 = c(0, (2017-cutoff))

below <- good_final %>%
  filter(yearID <= 2009)
favstats(~yearID, data = below)
```

```
##   min   Q1 median   Q3  max     mean       sd    n missing
##  1985 1992   1998 2004 2009 1997.743 6.848449 9759       0
```

```
ref_grid(mod2)
```

```
## 'emmGrid' object with variables:
##     year1 = 2001.3
##     jump = 0.26533
##     year2 = 0.94467
##     salary2 = 1937.2
##     bats_f = B, L, R
##     lgID_f = AL, NL
```

```
emmeans(mod2, "year1", at=list(year1 = c(0,1)), by="bats_f")
```

```
## bats_f = B:
##  year1 emmean    SE    df lower.CL upper.CL
##      0 -38.55 16.00 12707   -69.91    -7.20
##      1 -38.53 15.99 12707   -69.87    -7.19
##
## bats_f = L:
##  year1 emmean    SE    df lower.CL upper.CL
##      0   6.87  5.33 12707    -3.59    17.32
##      1   6.86  5.33 12707    -3.59    17.31
##
## bats_f = R:
##  year1 emmean    SE    df lower.CL upper.CL
##      0  10.94  3.40 12707     4.27    17.60
##      1  10.93  3.40 12707     4.27    17.59
##
## Results are averaged over the levels of: lgID_f
## Confidence level used: 0.95
```

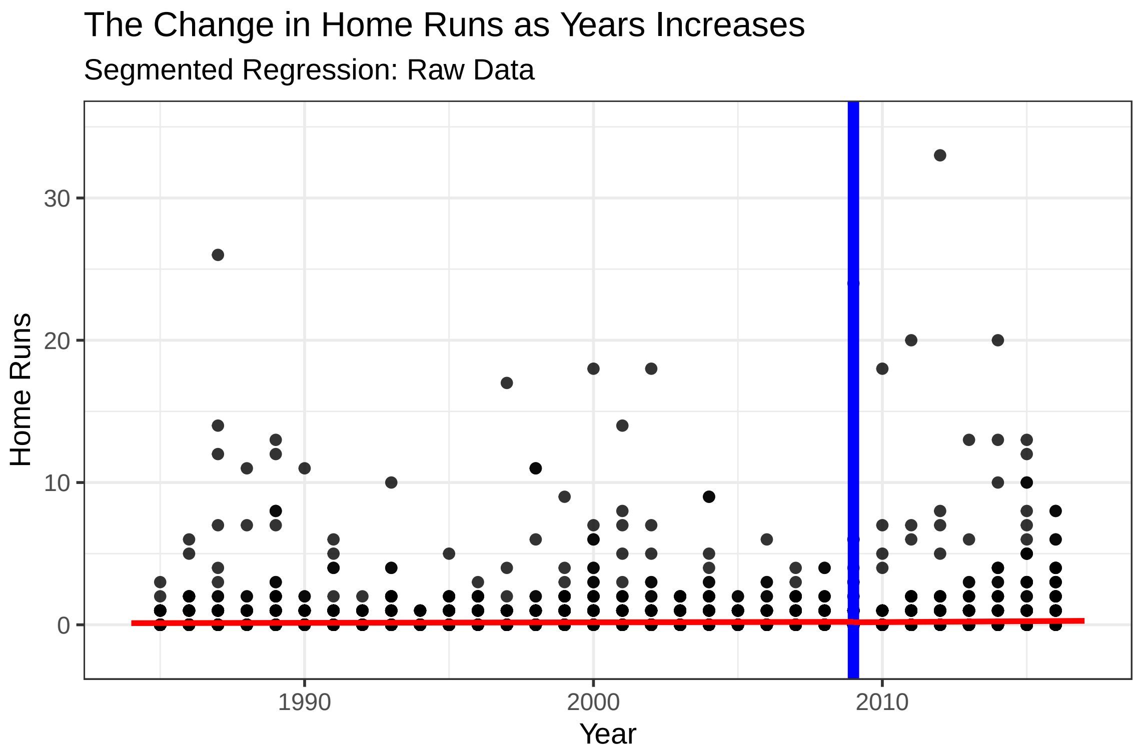
```
bat_means <- emmeans(mod2, "year1", at=list(year1 = c(0,1)), by="bats_f")
bat_means # only interaction significant.
```

```
## bats_f = B:
##  year1 emmean    SE    df lower.CL upper.CL
##      0 -38.55 16.00 12707   -69.91    -7.20
##      1 -38.53 15.99 12707   -69.87    -7.19
##
## bats_f = L:
##  year1 emmean    SE    df lower.CL upper.CL
##      0   6.87  5.33 12707    -3.59    17.32
##      1   6.86  5.33 12707    -3.59    17.31
##
## bats_f = R:
##  year1 emmean    SE    df lower.CL upper.CL
##      0  10.94  3.40 12707     4.27    17.60
##      1  10.93  3.40 12707     4.27    17.59
##
## Results are averaged over the levels of: lgID_f
## Confidence level used: 0.95
```

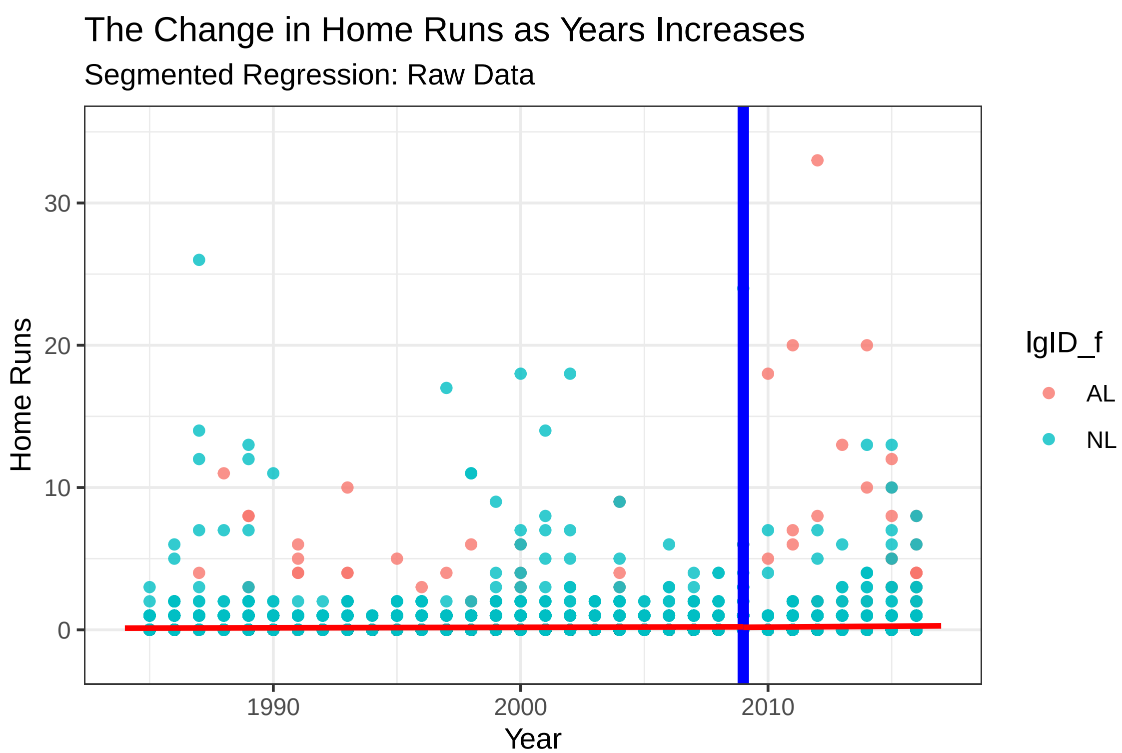
```
#Test of Simple Slopes
pairs(bat_means, reverse=TRUE) # SIMPLE SLOPES
```

```
## bats_f = B:
##  contrast estimate      SE    df t.ratio p.value
##  1 - 0     0.01942 0.00799 12707   2.429  0.0151
##
## bats_f = L:
##  contrast estimate      SE    df t.ratio p.value
##  1 - 0    -0.00337 0.00266 12707 -1.265  0.2058
##
## bats_f = R:
##  contrast estimate      SE    df t.ratio p.value
##  1 - 0    -0.00541 0.00170 12707 -3.185  0.0015
##
## Results are averaged over the levels of: lgID_f
```

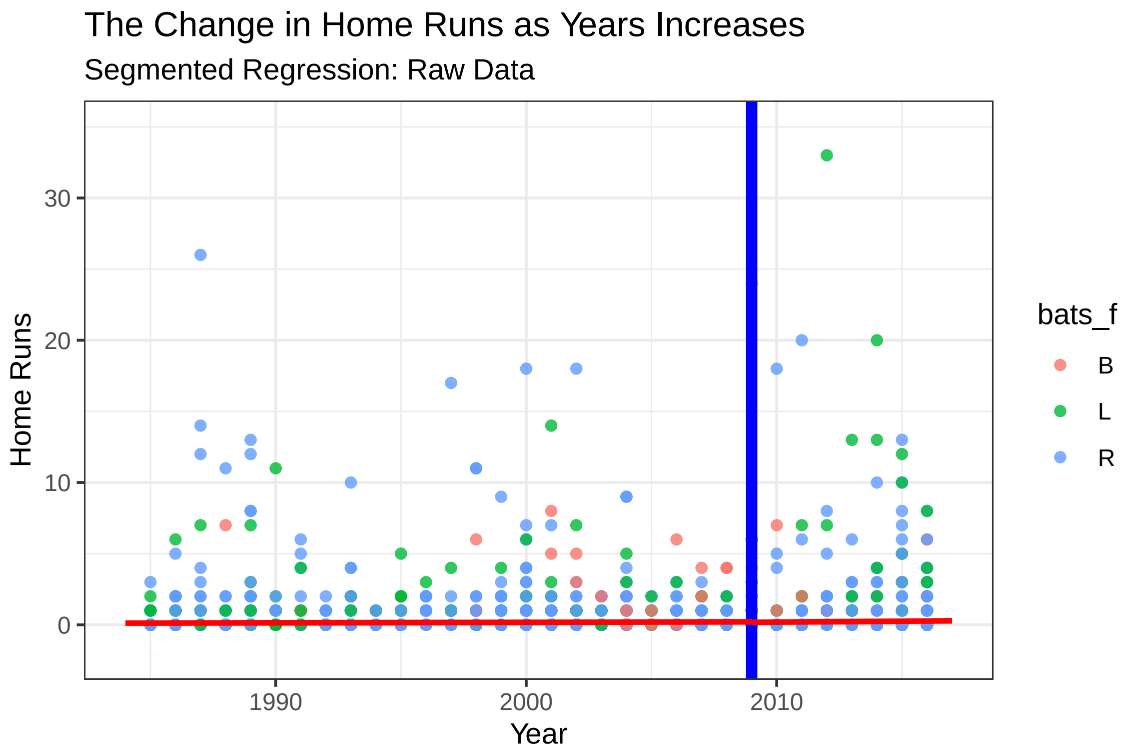
```
g <- gf_point(b_HR ~ yearID, data = good_final, alpha=0.8) %>%
  gf_theme(theme_bw())
g %>%
  gf_labs(title="The Change in Home Runs as Years Increases", subtitle="Segmented Regression: Raw Data"
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```



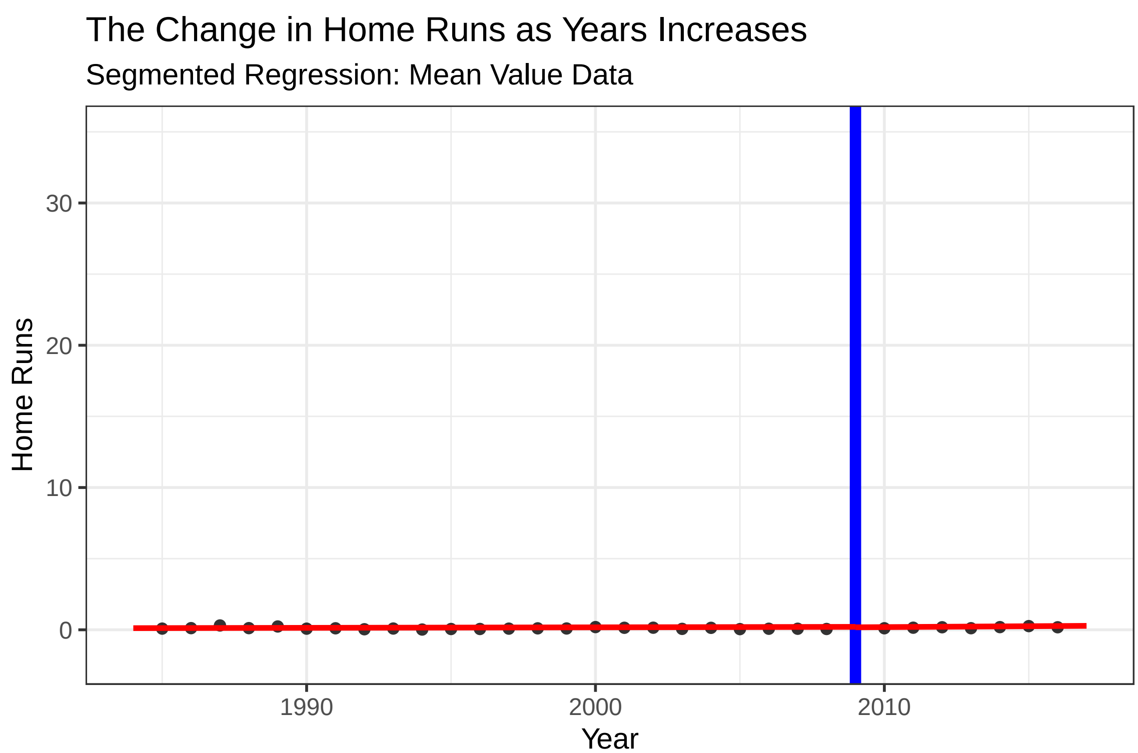

```r
g2 <- gf_point(b_HR ~ yearID, data = good_final, alpha=0.8, color = ~lgID_f) %>%
  gf_theme(theme_bw())
g2 %>%
  gf_labs(title="The Change in Home Runs as Years Increases", subtitle="Segmented Regression: Raw Data"
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```




```
  #gf_line(emmean ~ year1, data = mns1, color =~lgID_f, size = 1) %>%
  #gf_line(emmean ~ year1, data = mns2, color =~lgID_f, size = 1) %>%


g3 <- gf_point(b_HR ~ yearID, data = good_final, alpha=0.8, color = ~bats_f) %>%
  gf_theme(theme_bw())
g3 %>%
  gf_labs(title="The Change in Home Runs as Years Increases", subtitle="Segmented Regression: Raw Data"
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```




```
  #gf_line(emmean ~ year1, data = mns3, color =~bats_f, size = 1) %>%
  #gf_line(emmean ~ year1, data = mns4, color =~bats_f, size = 1) %>%


mns_final <- good_final %>%
  group_by(yearID) %>%
  summarise(mean_HR = mean(b_HR, na.rm=TRUE))
g_mns <- gf_point(mean_HR ~ yearID, data = mns_final, alpha=0.8) %>%
  gf_theme(theme_bw()) %>%
  gf_labs(title="The Change in Home Runs as Years Increases", subtitle="Segmented Regression: Mean Valu
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1, color="red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2, color="red", size = 1) + xlim(1984,2017) + ylim(-2,35)
g_mns
```




```r
# Coding for the second segment
good_final2 <- good_final %>%
  mutate(year1_part2 = case_when(year1 >= cutoff ~ cutoff,
                                 TRUE ~ year1))

mod4 = lm(b_HR ~ year1_part2 + jump + year2 + salary2 + bats_f*year1_part2 + bats_f*jump + bats_f*year2
summary(mod4)
```

```
##
## Call:
## lm(formula = b_HR ~ year1_part2 + jump + year2 + salary2 + bats_f *
##     year1_part2 + bats_f * jump + bats_f * year2 + lgID_f * year1_part2 +
##     lgID_f * jump + lgID_f * year2, data = good_final2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.647 -0.163 -0.110 -0.028 32.842
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -4.223e+01  1.622e+01  -2.604  0.00923 **
## year1_part2            2.125e-02  8.123e-03   2.616  0.00891 **
## jump                  -1.866e-01  2.248e-01  -0.830  0.40646
## year2                  1.105e-02  4.469e-02   0.247  0.80469
## salary2                1.147e-05  2.634e-06   4.356 1.33e-05 ***
## bats_fL                4.531e+01  1.681e+01   2.695  0.00704 **
## bats_fR                4.940e+01  1.630e+01   3.030  0.00245 **
## lgID_fNL               7.450e+00  5.564e+00   1.339  0.18060
## year1_part2:bats_fL   -2.279e-02  8.418e-03  -2.707  0.00679 **
## year1_part2:bats_fR   -2.483e-02  8.163e-03  -3.042  0.00236 **
```

```
## jump:bats_fL          2.629e-01  2.296e-01   1.145  0.25215
## jump:bats_fR          2.812e-01  2.240e-01   1.255  0.20938
## year2:bats_fL         2.098e-02  4.584e-02   0.458  0.64719
## year2:bats_fR        -4.300e-03  4.476e-02  -0.096  0.92347
## year1_part2:lgID_fNL -3.663e-03  2.786e-03  -1.315  0.18859
## jump:lgID_fNL        -4.827e-02  6.763e-02  -0.714  0.47540
## year2:lgID_fNL       -7.853e-03  1.325e-02  -0.593  0.55340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8829 on 12707 degrees of freedom
## Multiple R-squared:  0.01014,    Adjusted R-squared:  0.008897
## F-statistic: 8.139 on 16 and 12707 DF,  p-value: < 2.2e-16
```

WHAT IF WE REMOVED ALL ZERO VALUES

```
names(final)
```

```
##  [1] "yearID"   "playerID" "name"     "b_HR"     "salary"   "lgID"
##  [7] "bats"     "lgID_f"   "bats_f"   "salary2"
```

```
final_no_zeros = final[final$b_HR != 0,]

# Set up
breaksnz <- seq(1985,2016,1)
rmsenz <- rep(NA, length(breaksnz))

for(i in 1:length(breaksnz)){
  final2_nz <- final_no_zeros %>%
    mutate_at(vars(yearID), as.numeric)%>% #Initial catch all for numeric...
    mutate(year1 = yearID, #Simplet replication
           year2 = yearID - breaksnz[i], #Start second segment counting...
           year2 = case_when(year1 <= breaksnz[i]~0, #Make sure to start at zero BEFORE segment
                             TRUE~ year2),
           jump = case_when(yearID < breaksnz[i]~0, #Define the segment status...
                            yearID >= breaksnz[i]~1))
  mod <- lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID_f*y
  rmsenz[i]<- summary(mod)$sigma #Save the RMSE
}

potential_breakpoints_rmsenz = data.frame(br = breaksnz, rmse = rmsenz)
min(potential_breakpoints_rmsenz$rmse)
```

```
## [1] 3.206027
```

```
min( rmsenz[rmsenz!=min(rmsenz)] )
```

```
## [1] 3.224517
```

2. Assign the coding of the variables according to the change point indicated.

```
cutoffnz <- 2009
coded_finalnz <- final_no_zeros %>%
  mutate_at(vars(yearID), as.numeric) %>%
```

```r
  mutate(year1 = yearID,
         year2 = yearID - cutoffnz,
         year2 = case_when(year1 <= cutoffnz ~ 0, TRUE ~ year2),
         jump = case_when(yearID < cutoffnz ~ 0, yearID >= cutoffnz ~ 1))
```

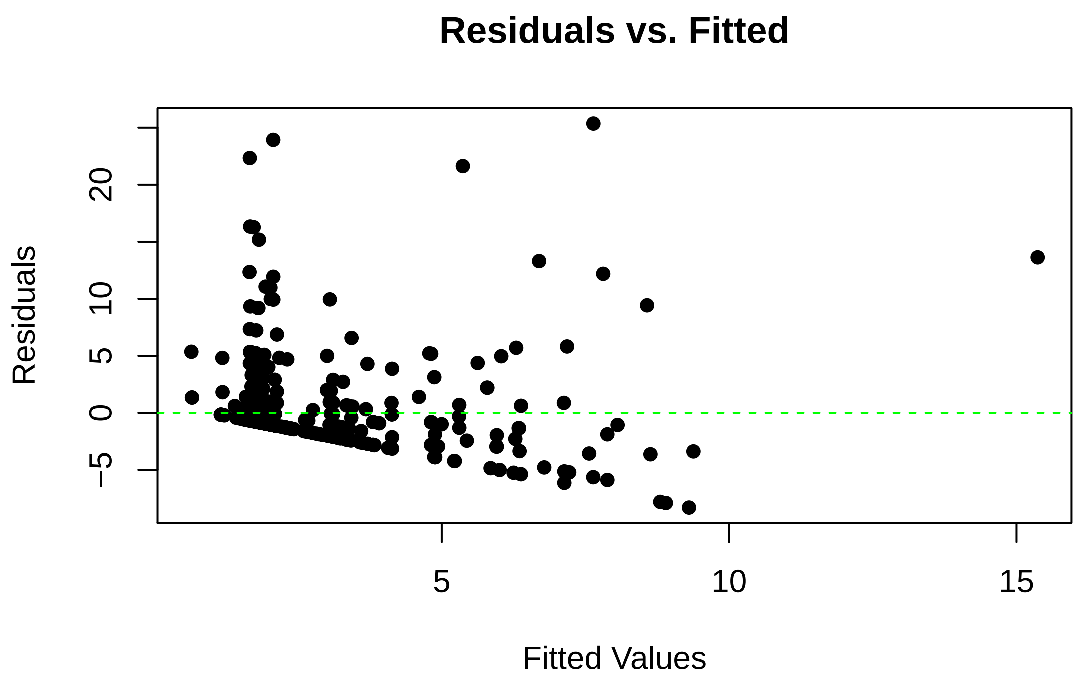
Double check the mutations

```r
plyr::count(coded_finalnz, c("yearID", "year1", "year2", "jump"))
```

```
##    yearID year1 year2 jump freq
## 1    1985  1985     0    0   16
## 2    1986  1986     0    0   24
## 3    1987  1987     0    0   24
## 4    1988  1988     0    0   15
## 5    1989  1989     0    0   24
## 6    1990  1990     0    0   18
## 7    1991  1991     0    0   17
## 8    1992  1992     0    0   12
## 9    1993  1993     0    0   19
## 10   1994  1994     0    0    7
## 11   1995  1995     0    0   19
## 12   1996  1996     0    0   18
## 13   1997  1997     0    0   18
## 14   1998  1998     0    0   23
## 15   1999  1999     0    0   26
## 16   2000  2000     0    0   31
## 17   2001  2001     0    0   22
## 18   2002  2002     0    0   25
## 19   2003  2003     0    0   20
## 20   2004  2004     0    0   23
## 21   2005  2005     0    0   17
## 22   2006  2006     0    0   20
## 23   2007  2007     0    0   21
## 24   2008  2008     0    0   15
## 25   2009  2009     0    1   23
## 26   2010  2010     1    1   14
## 27   2011  2011     2    1   26
## 28   2012  2012     3    1   23
## 29   2013  2013     4    1   23
## 30   2014  2014     5    1   19
## 31   2015  2015     6    1   33
## 32   2016  2016     7    1   30
```

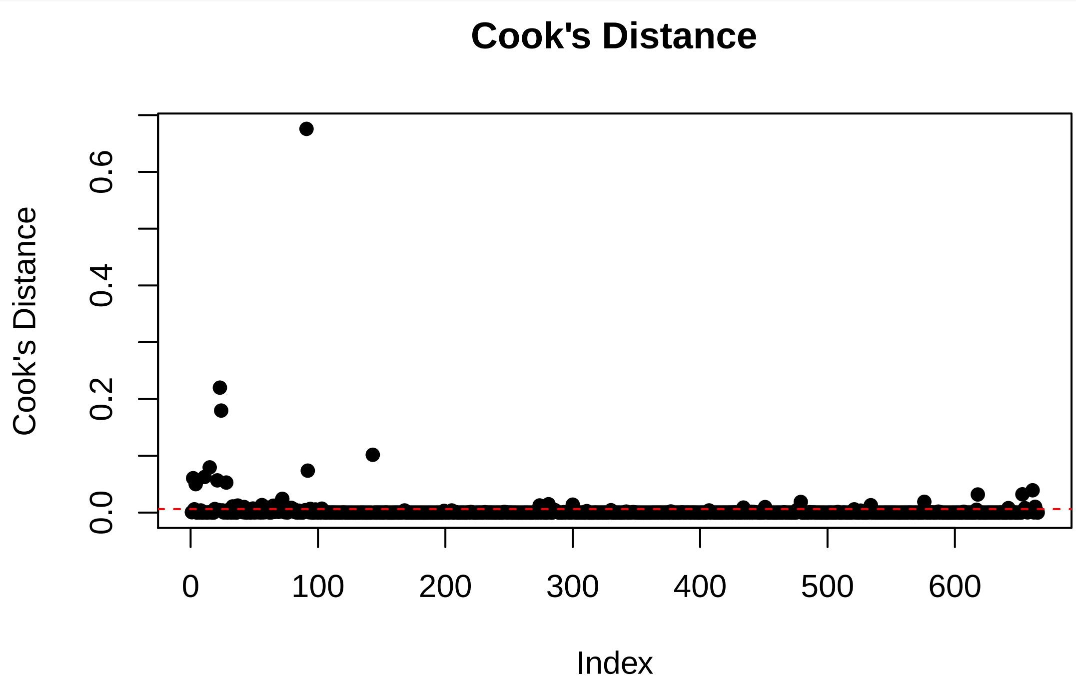
3. Look for outliers by running the full model.

```r
init_modnz= lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID
residFitted(init_modnz)
```



I think it's important to note that this Residuals vs Fitted plot isn't wonderful.

```
cooks_plotnz = cooksPlot(init_modnz, key.variable="playerID", print.obs = T, save.cutoff = T)
```




```
cooks_plotnz
```

```
##      playerID b_HR year1 jump year2   salary2 bats_f lgID_f Predicted_Y
## 1   swishni01   29  2009    1     0  5400.000      B     AL  15.3679015
## 2   davisch01   27  1993    0     0  2400.000      B     AL   5.3674170
## 3   davisch02   33  2012    1     3   488.000      L     AL   7.6390985
## 4    rossco01   24  2009    1     0  2225.000      R     NL   1.6605219
## 5   buehrma01    1  2009    1     0 14000.000      L     AL   8.8017691
```

```
## 6  vanevjo01    1  2009    1     0   400.000       L      AL   8.9002260
## 7  beckejo02    1  2009    1     0 11166.666       R      AL   9.3024200
## 8   dunnad01   20  2014    1     5 15000.000       L      AL   6.6937123
## 9  cuddymi01   20  2011    1     2 10500.000       R      AL   7.8091847
## 10 escobed01    6  2016    1     7  2150.000       B      AL   0.6440099
## 11  hallbi03   18  2010    1     1  8525.000       R      AL   8.5725134
## 12 zambrca01    1  2010    1     1 18875.000       B      NL   6.2502397
## 13 zambrca01    4  2009    1     0 18750.000       B      NL   7.5646235
## 14 wallati01   26  1987    0     0   765.000       R      NL   2.0683537
## 15 mosesjo01    1  1989    0     0   180.000       B      AL   6.0073365
## 16 snidetr01   13  2014    1     5  1200.000       L      NL   3.0534628
## 17 mosesjo01    1  1990    0     0   395.000       B      AL   5.8498180
## 18 oquenjo01    7  1988    0     0   275.000       B      NL   2.1743234
## 19 francje02   13  2015    1     6   950.000       R      NL   1.9353394
## 20 gloadro01    6  2009    1     0  1900.000       L      NL   1.1827357
## 21 larocad01   12  2015    1     6 12000.000       L      AL   6.2952673
## 22 robincl01   10  2015    1     6   525.000       L      NL   3.4314813
## 23 finlest01   14  2001    0     0  5375.000       L      NL   1.6563724
## 24 gladdda01   11  1988    0     0   360.000       R      AL   6.0356954
## 25 mccoymi01    2  2011    1     2   422.300       R      AL   7.8821419
## 26 menecfr01    9  2004    0     0   400.000       R      AL   2.1328543
## 27 garcile02    1  2014    1     5   505.500       B      AL   4.8694637
## 28 zeileto01   18  2002    0     0  6833.333       R      NL   1.6662584
## 29 norrida01    1  2015    1     6   508.700       L      AL   6.3784583
## 30 greenni01    6  2009    1     0   550.000       R      AL   9.3792791
## 31 montemi01    8  2016    1     7 14000.000       L      NL   3.7070612
## 32 maiermi01    2  2012    1     3   865.000       L      AL   7.6363692
## 33 martida01   11  1990    0     0   410.000       L      NL   1.6693879
## 34 gentrcr01    1  2012    1     3   484.300       R      AL   7.1326623
## 35 ramiral03   10  2015    1     6 10000.000       R      AL   4.8166811
## 36 wilsogl01   14  1987    0     0   662.400       R      NL   2.0690965
## 37 lyonsst01    1  1990    0     0   525.000       L      AL   5.2167789
## 38 zambrca01    2  2011    1     2 18875.000       B      NL   4.9367607
## 39 murphda07   13  2013    1     4  5775.000       L      AL   7.1806599
## 40 martean01    5  2010    1     1   413.400       R      AL   8.6312372
## 41 howarda02    1  1994    0     0   220.000       B      AL   5.2272371
## 42  bellde01   18  2000    0     0  5000.000       R      NL   1.7272859
## 43 whitety01    8  2016    1     7   507.500       R      AL   4.1363710
## 44 cansejo01   10  1993    0     0  4800.000       R      AL   4.7840048
##    Cooks_Distance        F_per
## 1     0.675762604 1.716293e-01
## 2     0.220032944 3.474349e-04
## 3     0.179618049 8.421161e-05
## 4     0.101771467 1.223754e-06
## 5     0.079595711 1.798099e-07
## 6     0.073868895 9.963581e-08
## 7     0.062740240 2.710572e-08
## 8     0.060567921 2.043113e-08
## 9     0.056735134 1.207472e-08
## 10    0.052876955 6.837467e-09
## 11    0.050083440 4.404983e-09
## 12    0.039289929 6.086039e-10
## 13    0.032169336 1.175548e-10
## 14    0.031882372 1.091772e-10
```

```
## 15    0.024216271 1.118836e-11
## 16    0.019027619 1.500262e-12
## 17    0.018785471 1.348050e-12
## 18    0.018744781 1.323851e-12
## 19    0.014734823 1.765446e-13
## 20    0.013972976 1.130984e-13
## 21    0.013231375 7.155713e-14
## 22    0.012977386 6.080812e-14
## 23    0.012595289 4.730779e-14
## 24    0.012398146 4.143462e-14
## 25    0.012111259 3.403383e-14
## 26    0.011821795 2.777096e-14
## 27    0.010850653 1.350344e-14
## 28    0.010167136 7.808575e-15
## 29    0.009908106 6.283203e-15
## 30    0.009866387 6.063817e-15
## 31    0.009691749 5.216885e-15
## 32    0.008827146 2.373523e-15
## 33    0.008825772 2.370410e-15
## 34    0.008742289 2.187872e-15
## 35    0.008419101 1.592433e-15
## 36    0.007916383 9.472893e-16
## 37    0.007824158 8.580979e-16
## 38    0.007741267 7.843328e-16
## 39    0.007291294 4.731089e-16
## 40    0.007272326 4.628171e-16
## 41    0.007138419 3.956007e-16
## 42    0.006891230 2.937644e-16
## 43    0.006490037 1.769851e-16
## 44    0.006295189 1.367951e-16
```

```
c_outliersnz = cooks_plotnz %>%
  filter (Cooks_Distance > 0.05) %>%
  pull(playerID)
c_outliersnz
```

```
##  [1] "swishni01" "davisch01" "davisch02" "rossco01"  "buehrma01"
##  [6] "vanevjo01" "beckejo02" "dunnad01"  "cuddymi01" "escobed01"
## [11] "hallbi03"
```

```
final_no_zeros %>%
  filter(playerID %in% c(c_outliersnz))
```

```
## # A tibble: 13 x 10
##    yearID playerID  name      b_HR salary lgID  bats  lgID_f bats_f salary2
##     <dbl> <chr>     <chr>    <dbl>  <dbl> <chr> <chr> <fct>  <fct>    <dbl>
##  1   2014 dunnad01  Adam Du~    20 1.50e7 AL    L     AL     L       15000
##  2   2010 hallbi03  Bill Ha~    18 8.52e6 AL    R     AL     R        8525
##  3   2006 beckejo02 Josh Be~     1 4.32e6 AL    R     AL     R        4325
##  4   2009 beckejo02 Josh Be~     1 1.12e7 AL    R     AL     R       11167.
##  5   2009 buehrma01 Mark Bu~     1 1.40e7 AL    L     AL     L       14000
##  6   2011 cuddymi01 Michael~    20 1.05e7 AL    R     AL     R       10500
##  7   1993 davisch01 Chili D~    27 2.40e6 AL    B     AL     B        2400
##  8   2012 davisch02 Chris D~    33 4.88e5 AL    L     AL     L         488
##  9   2016 escobed01 Eduardo~     6 2.15e6 AL    B     AL     B        2150
```

```
## 10   2009 swishni01 Nick Sw~    29 5.40e6 AL    B     AL     B         5400
## 11   2009 vanevjo01 Jonatha~     1 4.00e5 AL    L     AL     L          400
## 12   2009 rossco01  Cody Ro~    24 2.22e6 NL    R     NL     R         2225
## 13   2005 beckejo02 Josh Be~     1 2.40e6 NL    R     NL     R         2400
```

```
good_finalnz <- coded_finalnz %>%
  filter(playerID %not_in% c(c_outliersnz))

mod2nz = lm(b_HR ~ year1 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lgID_f*y
mod3nz = lm(b_HR ~ year1 + jump + year2 + salary2, data = good_finalnz)
summary(mod2nz)
```

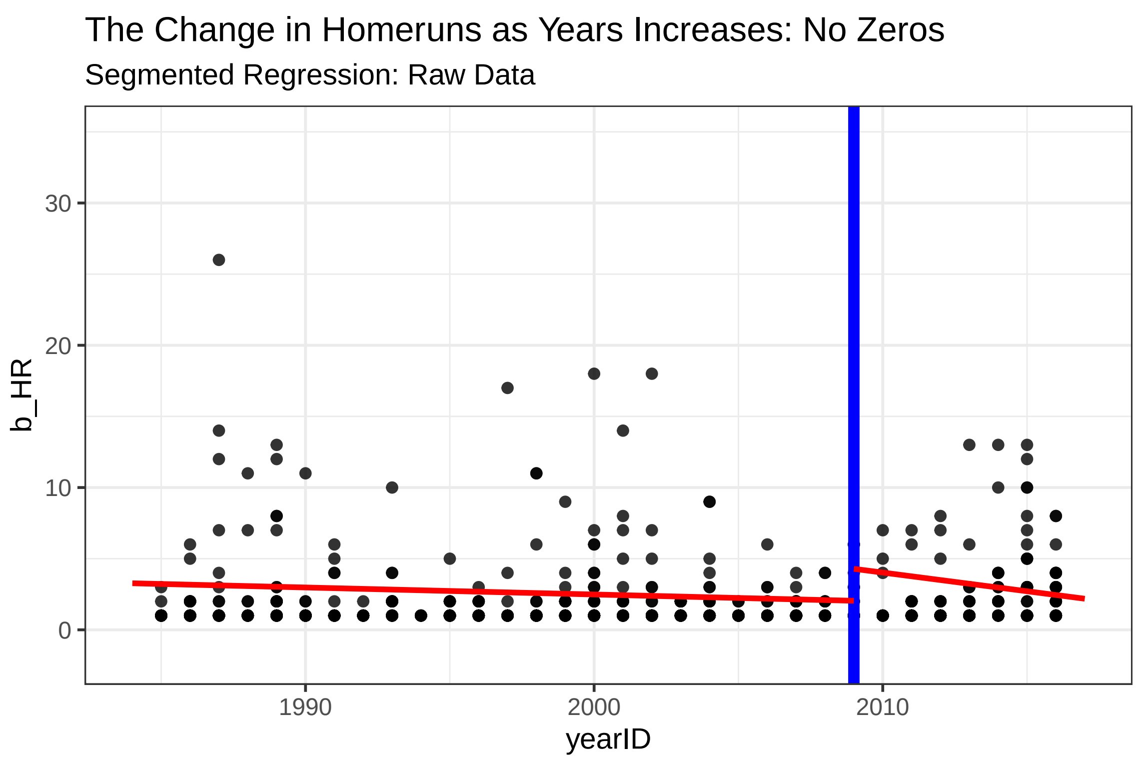
```
##
## Call:
## lm(formula = b_HR ~ year1 + jump + year2 + salary2 + bats_f *
##     year1 + bats_f * jump + bats_f * year2 + lgID_f * year1 +
##     lgID_f * jump + lgID_f * year2, data = good_finalnz)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8394 -0.9458 -0.7114  0.0894 23.8780
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     7.688e+01  1.713e+02   0.449  0.65380
## year1          -3.670e-02  8.578e-02  -0.428  0.66894
## jump            4.845e+00  2.171e+00   2.232  0.02598 *
## year2          -9.547e-01  3.852e-01  -2.478  0.01345 *
## salary2        -1.401e-05  2.738e-05  -0.512  0.60905
## bats_fL         2.737e+02  1.519e+02   1.802  0.07201 .
## bats_fR         3.246e+02  1.409e+02   2.305  0.02151 *
## lgID_fNL       -3.510e+02  1.200e+02  -2.924  0.00358 **
## year1:bats_fL  -1.372e-01  7.602e-02  -1.805  0.07155 .
## year1:bats_fR  -1.626e-01  7.049e-02  -2.307  0.02136 *
## jump:bats_fL   -1.622e-01  1.914e+00  -0.085  0.93249
## jump:bats_fR   -7.236e-01  1.761e+00  -0.411  0.68131
## year2:bats_fL   9.534e-01  3.781e-01   2.521  0.01194 *
## year2:bats_fR   9.666e-01  3.558e-01   2.717  0.00678 **
## year1:lgID_fNL  1.750e-01  6.009e-02   2.912  0.00372 **
## jump:lgID_fNL  -4.609e+00  1.458e+00  -3.162  0.00164 **
## year2:lgID_fNL  2.028e-01  2.410e-01   0.841  0.40048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.512 on 635 degrees of freedom
## Multiple R-squared:  0.0994, Adjusted R-squared:  0.07671
## F-statistic:  4.38 on 16 and 635 DF,  p-value: 3.065e-08
```

```
summary(mod3nz)
```

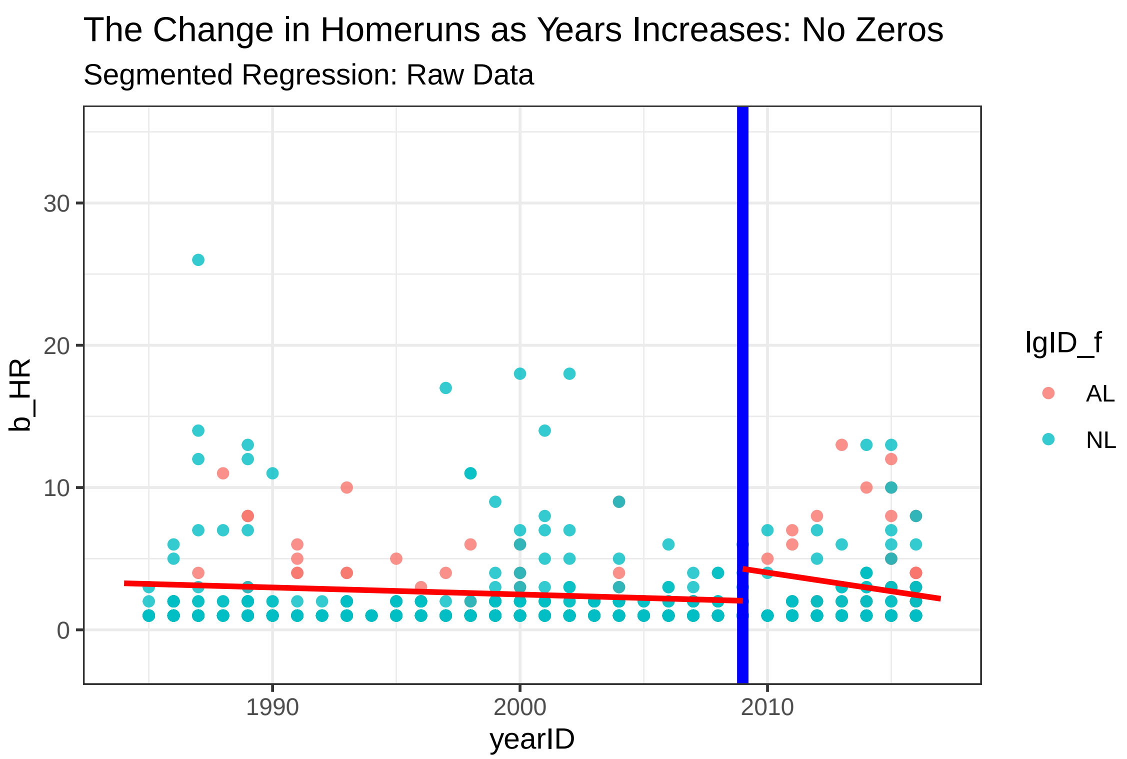
```
##
## Call:
## lm(formula = b_HR ~ year1 + jump + year2 + salary2, data = good_finalnz)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.1456 -1.1221 -0.9071 -0.0327 23.8166
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.292e+01  3.639e+01   0.905   0.3660
## year1       -1.546e-02  1.824e-02  -0.848   0.3969
## jump        -5.397e-02  4.572e-01  -0.118   0.9061
## year2        2.081e-01  8.719e-02   2.386   0.0173 *
## salary2     -3.114e-05  2.703e-05  -1.152   0.2498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.599 on 647 degrees of freedom
## Multiple R-squared:  0.01817,    Adjusted R-squared:  0.0121
## F-statistic: 2.994 on 4 and 647 DF,  p-value: 0.01824
```

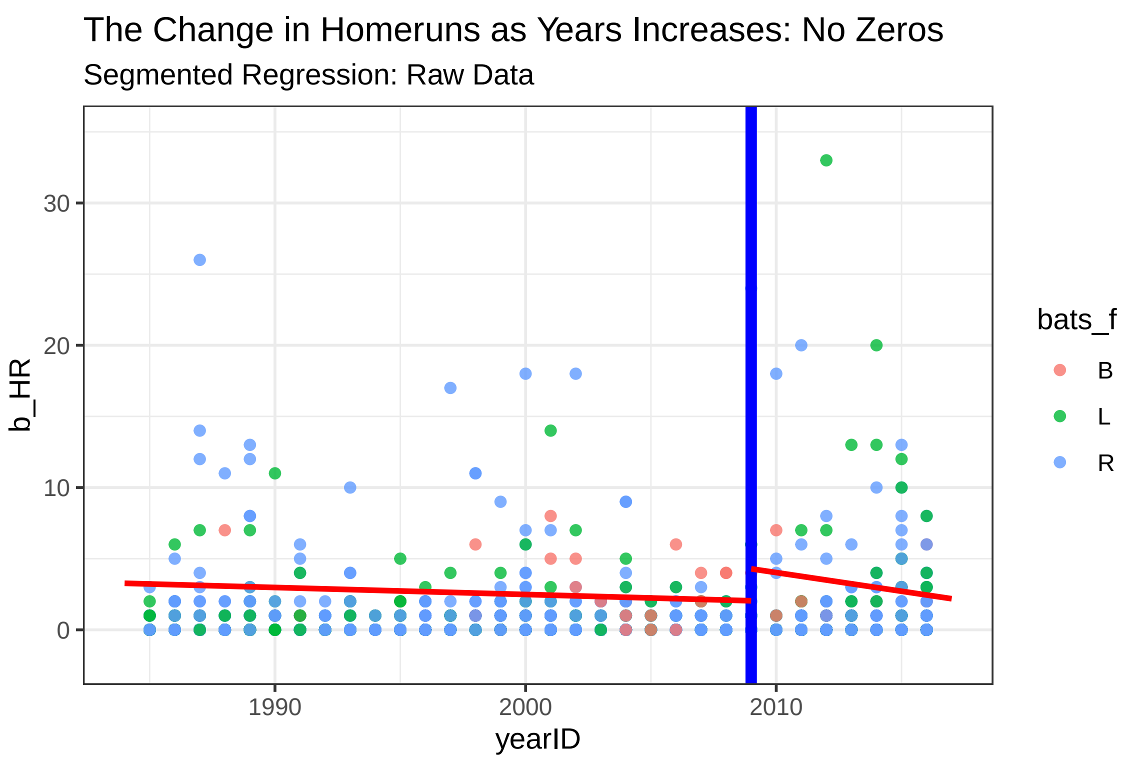
```r
p1nz <- summary(emmeans(mod2nz, "year1", at=list(year1=c(1984, cutoffnz), year2=0, jump=0)))
p2nz <- summary(emmeans(mod2nz, "year1", at=list(year1=c(cutoffnz, 2017), year2=c(0, (2017-cutoffnz)), j
p2nz <- p2nz %>%
  slice(c(1,4))

mns1nz <- summary(emmeans(mod2nz, "year1", at=list(year1 = seq(1984,cutoffnz,1)), year2 = 0, jump = 0, l
mns2nz <- summary(emmeans(mod2nz, "year1", at=list(year1 = seq(cutoffnz,2017,1)), year2 = c(0, (2017-cu

mns3nz <- summary(emmeans(mod2nz, "year1", at=list(year1 = seq(1984,cutoffnz,1)), year2 = 0, jump = 0, l
mns4nz <- summary(emmeans(mod2nz, "year1", at=list(year1 = seq(cutoffnz,2017,1)), year2 = c(0, (2017-cu

gnz <- gf_point(b_HR ~ yearID, data = good_finalnz, alpha=0.8) %>%
  gf_theme(theme_bw())
gnz %>%
  gf_labs(title="The Change in Homeruns as Years Increases: No Zeros", subtitle="Segmented Regression: l
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1nz, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2nz, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```



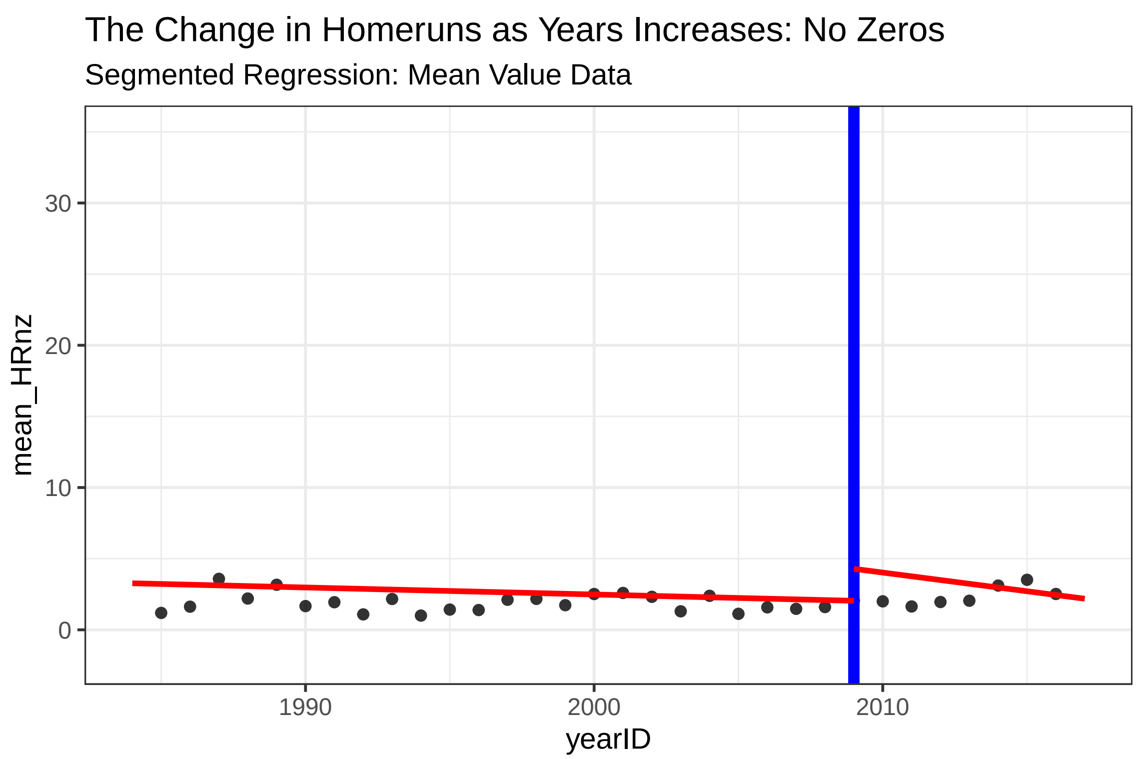

```
gnz2 <- gf_point(b_HR ~ yearID, data = good_finalnz, alpha=0.8, color = ~lgID_f) %>%
  gf_theme(theme_bw())
gnz2 %>%
  gf_labs(title="The Change in Homeruns as Years Increases: No Zeros", subtitle="Segmented Regression: R
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1nz, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2nz, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```




```
g3nz <- gf_point(b_HR ~ yearID, data = good_final, alpha=0.8, color = ~bats_f) %>%
  gf_theme(theme_bw())
g3 %>%
  gf_labs(title="The Change in Homeruns as Years Increases: No Zeros", subtitle="Segmented Regression: 
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1nz, color = "red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2nz, color = "red", size = 1) + xlim(1984,2017) + ylim(-2,35)
```




```
mns_finalnz <- good_finalnz %>%
  group_by(yearID) %>%
  summarise(mean_HRnz = mean(b_HR, na.rm=TRUE))
g_mnsnz <- gf_point(mean_HRnz ~ yearID, data = mns_finalnz, alpha=0.8) %>%
  gf_theme(theme_bw()) %>%
  gf_labs(title="The Change in Homeruns as Years Increases: No Zeros", subtitle="Segmented Regression: R
  gf_vline(xintercept = ~2009, color="blue", size = 2) %>%
  gf_line(emmean ~ year1, data = p1nz, color="red", size = 1) %>%
  gf_line(emmean ~ year1, data = p2nz, color="red", size = 1) + xlim(1984,2017) + ylim(-2,35)
g_mnsnz
```




```
# Coding for the second segment
good_final2nz <- good_finalnz %>%
  mutate(year1_part2 = case_when(year1 >= cutoff ~ cutoff,
                                 TRUE ~ year1))

mod4nz = lm(b_HR ~ year1_part2 + jump + year2 + salary2 + bats_f*year1 + bats_f*jump + bats_f*year2 + lg
summary(mod4nz)
```

```
##
## Call:
## lm(formula = b_HR ~ year1_part2 + jump + year2 + salary2 + bats_f *
##     year1 + bats_f * jump + bats_f * year2 + lgID_f * year1 +
##     lgID_f * jump + lgID_f * year2, data = good_final2nz)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8394 -0.9458 -0.7114  0.0894 23.8780
##
## Coefficients: (1 not defined because of singularities)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     7.688e+01  1.713e+02   0.449  0.65380
## year1_part2    -3.670e-02  8.578e-02  -0.428  0.66894
## jump            4.845e+00  2.171e+00   2.232  0.02598 *
## year2          -9.914e-01  3.749e-01  -2.644  0.00839 **
## salary2        -1.401e-05  2.738e-05  -0.512  0.60905
## bats_fL         2.737e+02  1.519e+02   1.802  0.07201 .
## bats_fR         3.246e+02  1.409e+02   2.305  0.02151 *
## year1                  NA         NA      NA       NA
## lgID_fNL       -3.510e+02  1.200e+02  -2.924  0.00358 **
## bats_fL:year1  -1.372e-01  7.602e-02  -1.805  0.07155 .
```

```
## bats_fR:year1  -1.626e-01  7.049e-02  -2.307  0.02136 *
## jump:bats_fL   -1.622e-01  1.914e+00  -0.085  0.93249
## jump:bats_fR   -7.236e-01  1.761e+00  -0.411  0.68131
## year2:bats_fL   9.534e-01  3.781e-01   2.521  0.01194 *
## year2:bats_fR   9.666e-01  3.558e-01   2.717  0.00678 **
## year1:lgID_fNL  1.750e-01  6.009e-02   2.912  0.00372 **
## jump:lgID_fNL  -4.609e+00  1.458e+00  -3.162  0.00164 **
## year2:lgID_fNL  2.028e-01  2.410e-01   0.841  0.40048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.512 on 635 degrees of freedom
## Multiple R-squared:  0.0994, Adjusted R-squared:  0.07671
## F-statistic:  4.38 on 16 and 635 DF,  p-value: 3.065e-08
```