Moneyball - CUNY Data Science 621

Eric Hirsch

2/20/2021

1. Description of the Dataset

a. ASSIGNMENT: In this assignment we explore, analyze and model a data set containing approximately 2276 records, each representing a professional baseball team from the years 1871 to 2006 inclusive. Each record has the performance of the team for the given year, with all of the statistics adjusted to match the performance of a 162 game season.

We will build a multiple linear regression model on the training data to predict the number of wins for the team.

b. THE ISSUE OF HIDDEN GROUPINGS: An issue with the data is hidden groupings. Records may not be independent of each other, as team data in one year will be related to team data in the next year. We know that if some records were adjusted to match a longer season, there may be an "eras of baseball" effect as teams from earlier years behave differently from later ones. Finally, within the record, columns may not be independent. In particular, teams with high offensive stats (like hitting) may have lower defensive stats (like pitching), as the teams on limited budgets make strategic choices between the two. We will attempt to address some of these issues in this analysis.

1. Data Exploration

All of the columns in the dataset are numeric. We begin by examining their means, medians and distributions.

```
##      INDEX         TARGET_WINS       BATTING_H      BATTING_2B
##  Min.   :   1.0   Min.   :  0.00   Min.   : 891   Min.   : 69.0
##  1st Qu.: 630.8   1st Qu.: 71.00   1st Qu.:1383   1st Qu.:208.0
##  Median :1270.5   Median : 82.00   Median :1454   Median :238.0
##  Mean   :1268.5   Mean   : 80.79   Mean   :1469   Mean   :241.2
##  3rd Qu.:1915.5   3rd Qu.: 92.00   3rd Qu.:1537   3rd Qu.:273.0
##  Max.   :2535.0   Max.   :146.00   Max.   :2554   Max.   :458.0
##
##    BATTING_3B       BATTING_HR       BATTING_BB      BATTING_SO
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.0   Min.   :   0.0
##  1st Qu.: 34.00   1st Qu.: 42.00   1st Qu.:451.0   1st Qu.: 548.0
##  Median : 47.00   Median :102.00   Median :512.0   Median : 750.0
##  Mean   : 55.25   Mean   : 99.61   Mean   :501.6   Mean   : 735.6
##  3rd Qu.: 72.00   3rd Qu.:147.00   3rd Qu.:580.0   3rd Qu.: 930.0
##  Max.   :223.00   Max.   :264.00   Max.   :878.0   Max.   :1399.0
##                                                    NA's   :102
##    BASERUN_SB       BASERUN_CS      BATTING_HBP      PITCHING_H
##  Min.   :  0.0   Min.   :  0.0   Min.   :29.00   Min.   : 1137
```

```
##  1st Qu.: 66.0   1st Qu.: 38.0   1st Qu.:50.50   1st Qu.: 1419
##  Median :101.0   Median : 49.0   Median :58.00   Median : 1518
##  Mean   :124.8   Mean   : 52.8   Mean   :59.36   Mean   : 1779
##  3rd Qu.:156.0   3rd Qu.: 62.0   3rd Qu.:67.00   3rd Qu.: 1682
##  Max.   :697.0   Max.   :201.0   Max.   :95.00   Max.   :30132
##  NA's   :131     NA's   :772     NA's   :2085
##   PITCHING_HR     PITCHING_BB      PITCHING_SO        FIELDING_E
##  Min.   :  0.0   Min.   :   0.0   Min.   :    0.0   Min.   :  65.0
##  1st Qu.: 50.0   1st Qu.: 476.0   1st Qu.:  615.0   1st Qu.: 127.0
##  Median :107.0   Median : 536.5   Median :  813.5   Median : 159.0
##  Mean   :105.7   Mean   : 553.0   Mean   :  817.7   Mean   : 246.5
##  3rd Qu.:150.0   3rd Qu.: 611.0   3rd Qu.:  968.0   3rd Qu.: 249.2
##  Max.   :343.0   Max.   :3645.0   Max.   :19278.0   Max.   :1898.0
##                                   NA's   :102
##   FIELDING_DP
##  Min.   : 52.0
##  1st Qu.:131.0
##  Median :149.0
##  Mean   :146.4
##  3rd Qu.:164.0
##  Max.   :228.0
##  NA's   :286
```

We note that a number of columns have NAs. Batting_SO and Pitching_SO have the same number of NA's and may be related.

We more closely examine the distribution of columns in the dataset (fig. 1):

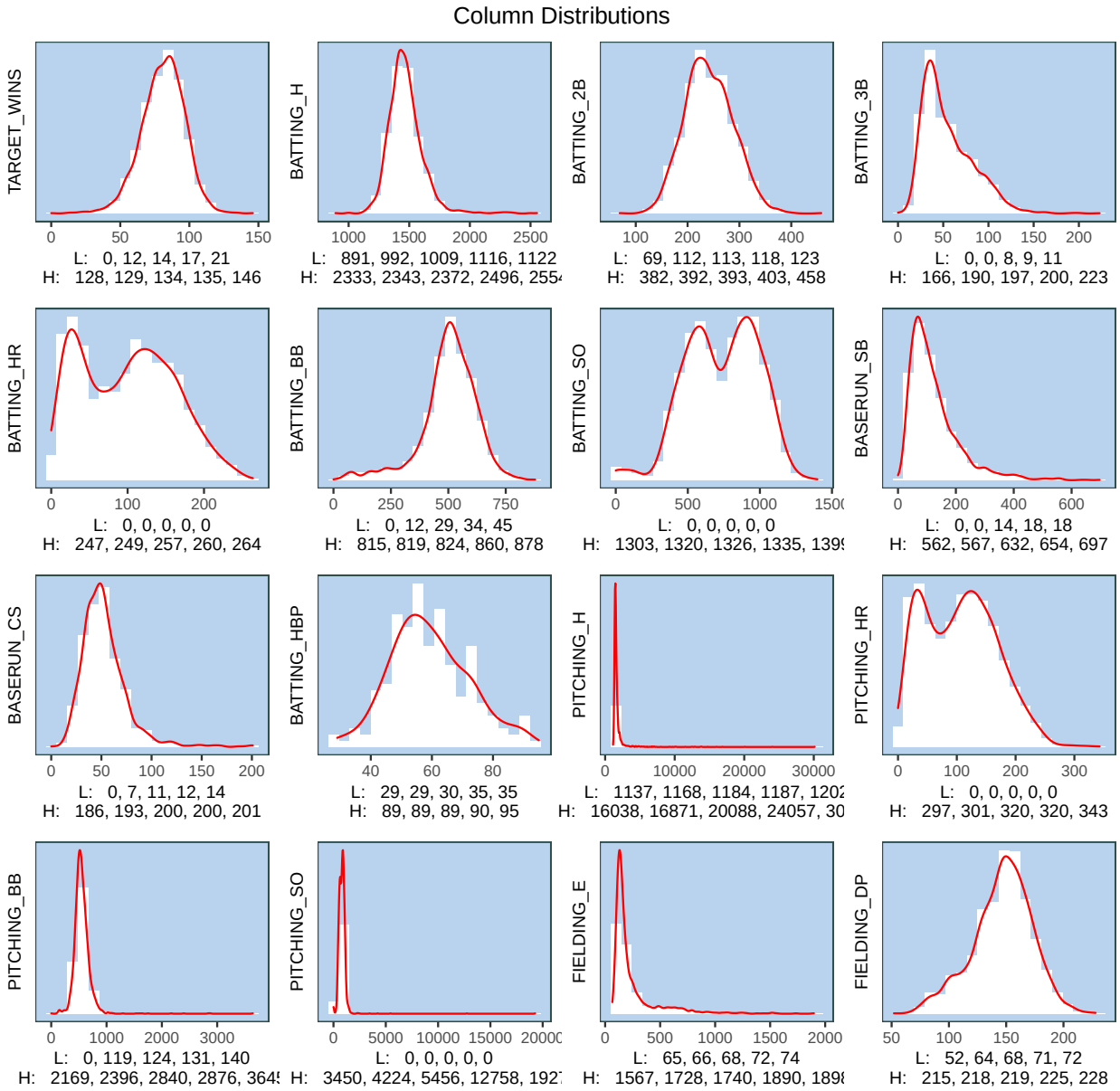


Fig. 1

Our dependent variable (Target Wins) appears to be normally distributed. However, a number of columns are severely skewed (Errors, Strikeouts, Pitching_H, etc.) A few columns (Batting SO, Pitching_HR and Batting_HR) have a bimodal distribution. This might point to some hidden groupings in the dataset.

Boxplots help us identify outliers (fig. 2):

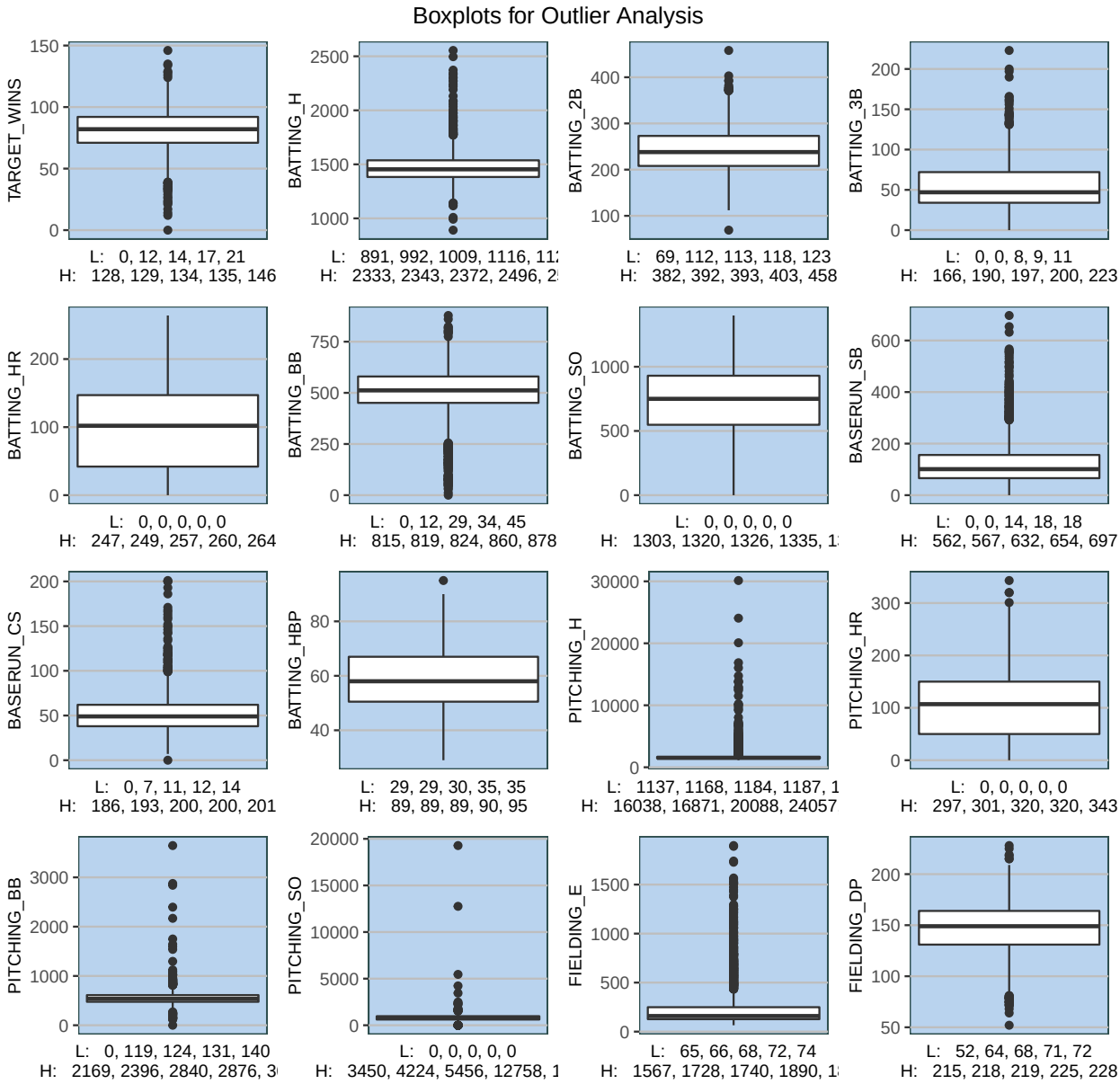


Fig. 2

There a number of outliers, both high and low. For example, there are many zeros, which may be implausible. In addition, many of the ranges appear extreme, such as giving up between 3,500 hits and 19,000 hits, or getting from 12 to over 800 walks.

We investigate correlations in the dataset, both between the dependent variable and the other variables (fig. 3), and between the dependent variables and each other (fig. 4).

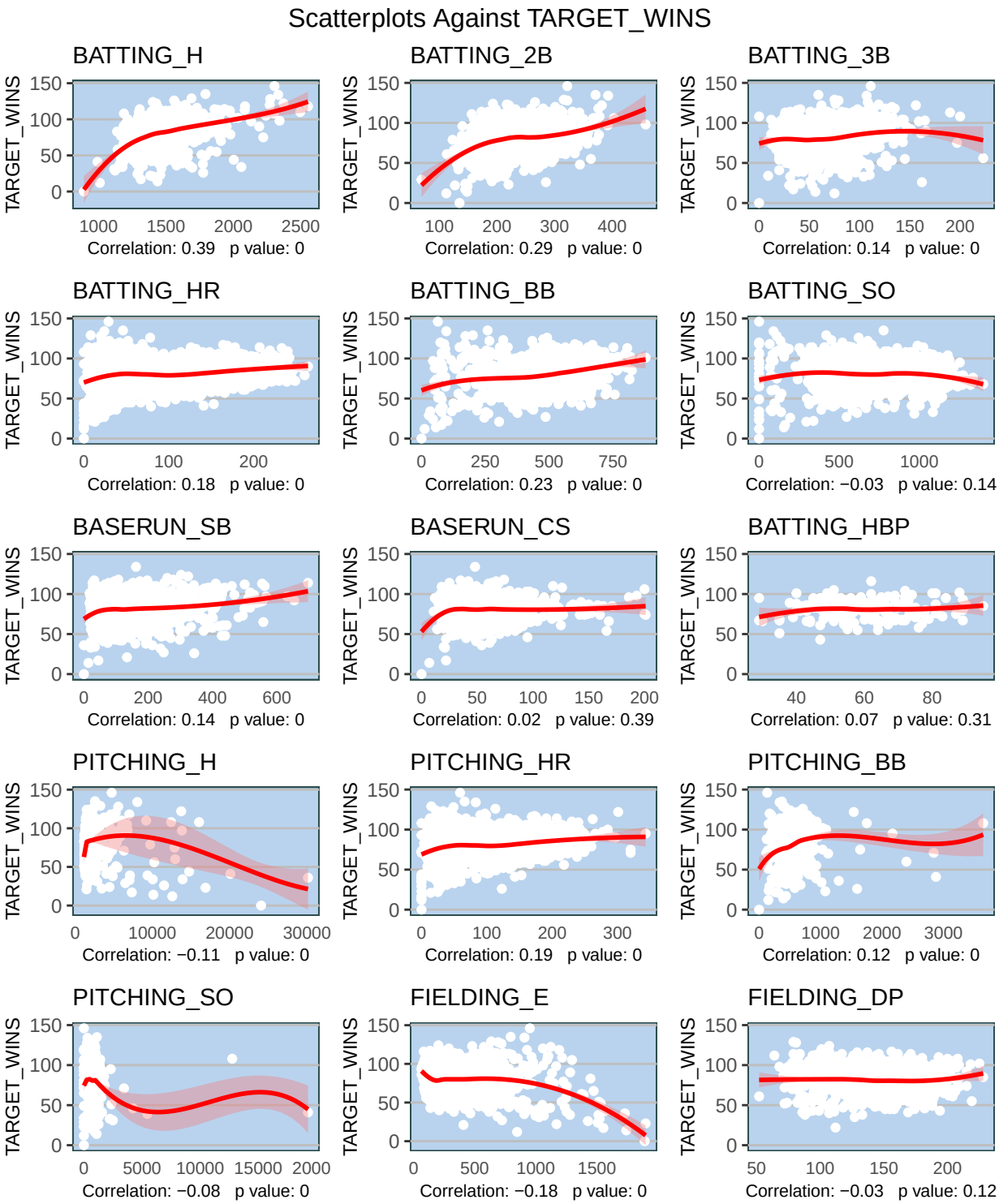


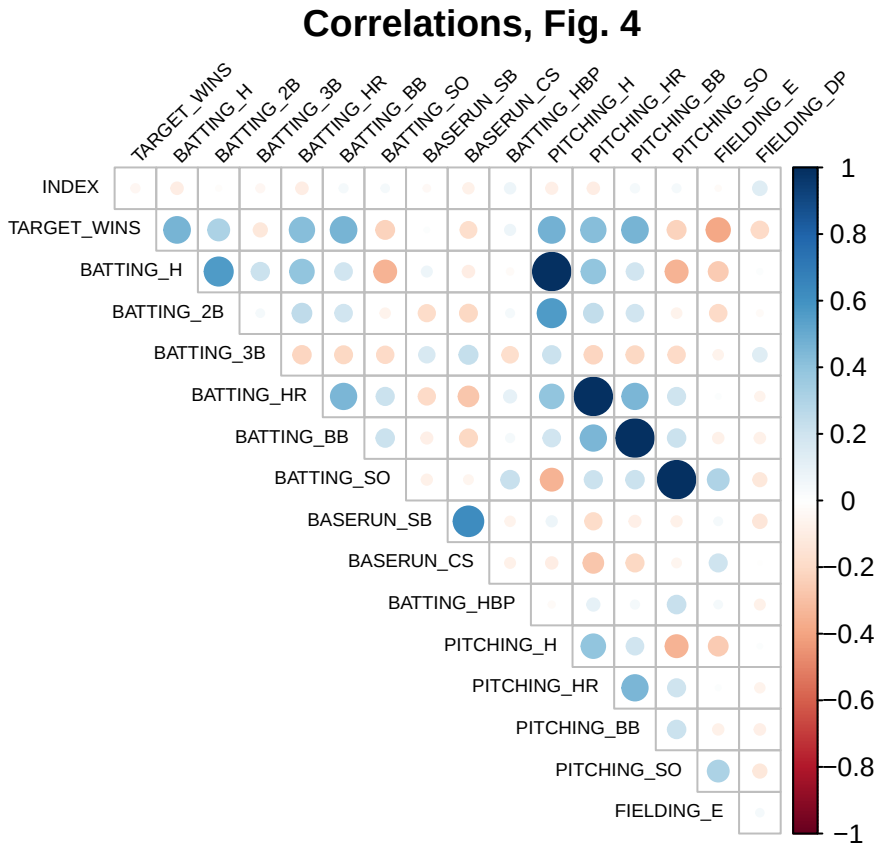
Fig.3

Here we see a number of puzzles, mainly among the pitching correlations. Hits should show a much stronger negative correlation, and in fact appear positive for a portion. Making double plays is surprisingly neutral, as are strikeouts. Pitching_HR is also positive when we would expect negative.

We do need to acknowledge here the possibility of strategy groupings (defense and offense) which may contribute to these anomalies. In other words, a team with poor pitching may have strong hitting, which

then wins games.

We can look for evidence of this possibility by examining multicollinearity:



Indeed, the pitching categories are strongly correlated with their hitting counterparts. All four of the pitching categories follow this pattern.

2. Data Preparation

We begin by devising a strategy for the NAs. We can eliminate the Batting_HBP and Baserun_CS columns because they have too many NA's. We also create flags for the other columns with significant NA's.

We are particularly interested in the SO columns because they do not appear random, and investigation establishes that they have complete overlap with each other. Fielding_DP and BR_SB aslo have some overlap. These may relate to eras of baseball when certain statistics were not collected. (see Fig. 5)

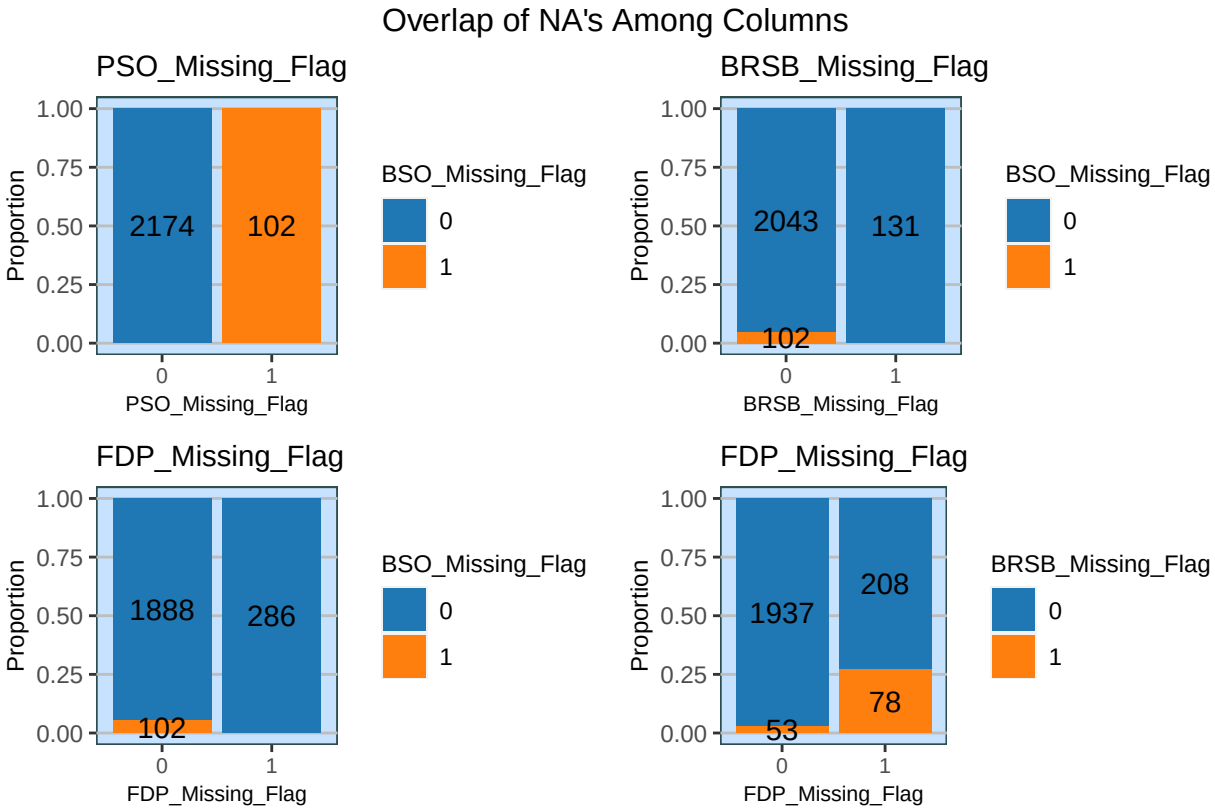


Fig. 5

We eliminate the pitching SO column because it is redundant. While not MCAR (missing completely at random), if the Batting_SO column is MAR (missing at random), we may be bale to eliminate these rows, as there are not so many (5% of the total).

One way to investigate the randomness of this missing cohort is to look for interactions between the cohort and other dataset columns. In fact, we see that there are a number of columns with strong, even extreme interactions (see fig. 6).

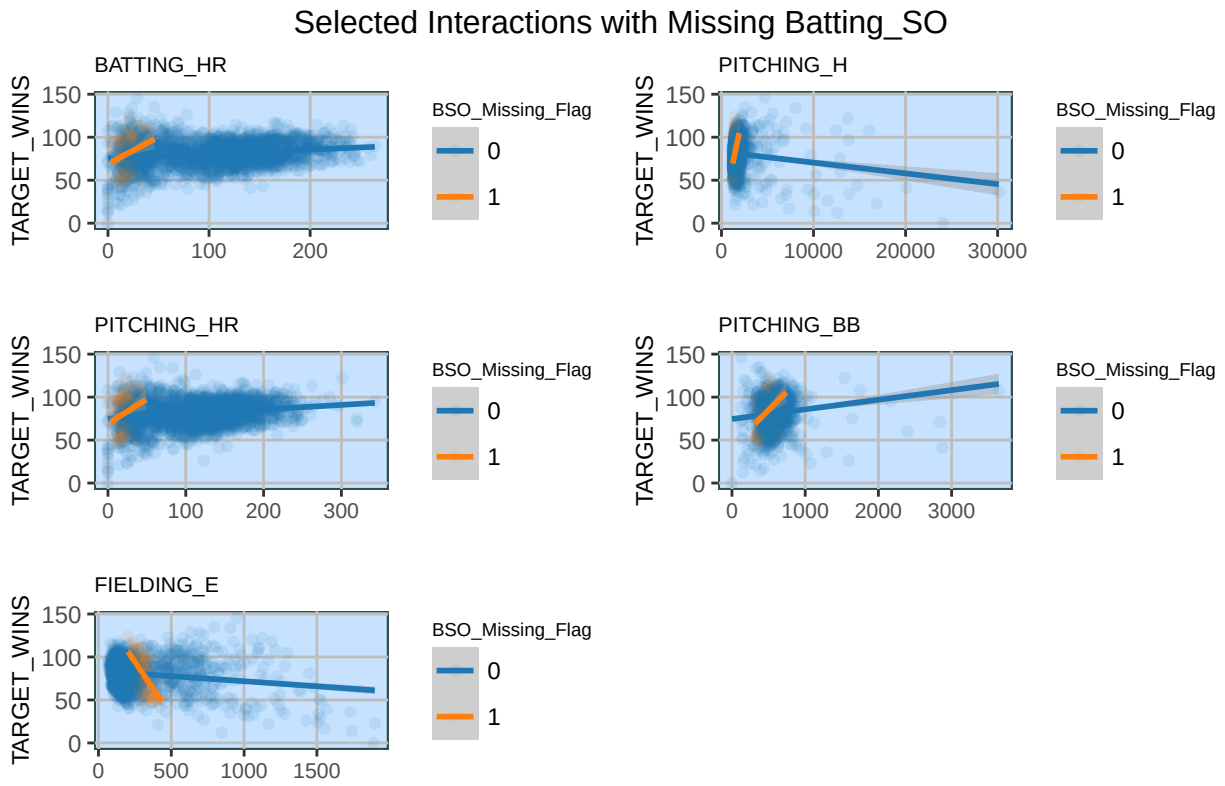


Fig. 6

It is possible this cohort represents a different baseball era when such statistics were not collected. In any case, we cannot eliminate these rows without losing critical data, so we employ the following strategy: 1) create the rows and impute a value, 2) retain a "missing" flag to keep track of the cohort, and 2) add interaction terms where appropriate.

Before we address imputation, we want to work with the implausible zeros in the dataet. In particular, we note that the 0s in Pitching_SO and Batting_SO are a complete overlap, and we can see from the histograms that the jump between 0 and the next lowest value is not smooth, and so we will treat them as NA's. We do the same with batting and pitching HR, since there is also a jump up after zero which suggests it is being used as an indicator of missing value.

Just so we have some reasonable criteria for imputation strategy, we compare the r-squared of three regressions - with NA's imputed as means, with NA's imputed as medians, and with NA rows eliminated altogether.

```
## [1] "type:" "mean"
## [1] "r2mean:" "0.4031"
## [1] "r2median:" "0.403"
## [1] "r2omit" "0.4019"
```

The mean and median have the same r-squared, while the elimination of the rows has a smaller r-squared. We therefore choose to impute the mean.

Not surprisingly, the evaluation dataset shows the same results:

```
## [1] "type:" "mean"
## [1] "r2mean:" "0.4031"
```

```
## [1] "r2median:" "0.403"
## [1] "r2omit" "0.4019"
```

Although outliers and possible bad data appear in a number of places, without domain knowledge we are reluctant to eliminate any other outliers or influential points at this point without good reason. We don't know if extreme numbers are necessarily implausible. Therefore the outliers will remain.

3. Data Modeling

1. We create a flag for hits under 1500

As previously noted, Pitching_H is surprisingly weak in it's relationship to wins, and in fact appears positive for a large portion of its distribution. We examine more closely the relationship between pitching hits and wins, paying particular attention to the portion of the relationship where hits are below 3,000 (fig. 7).

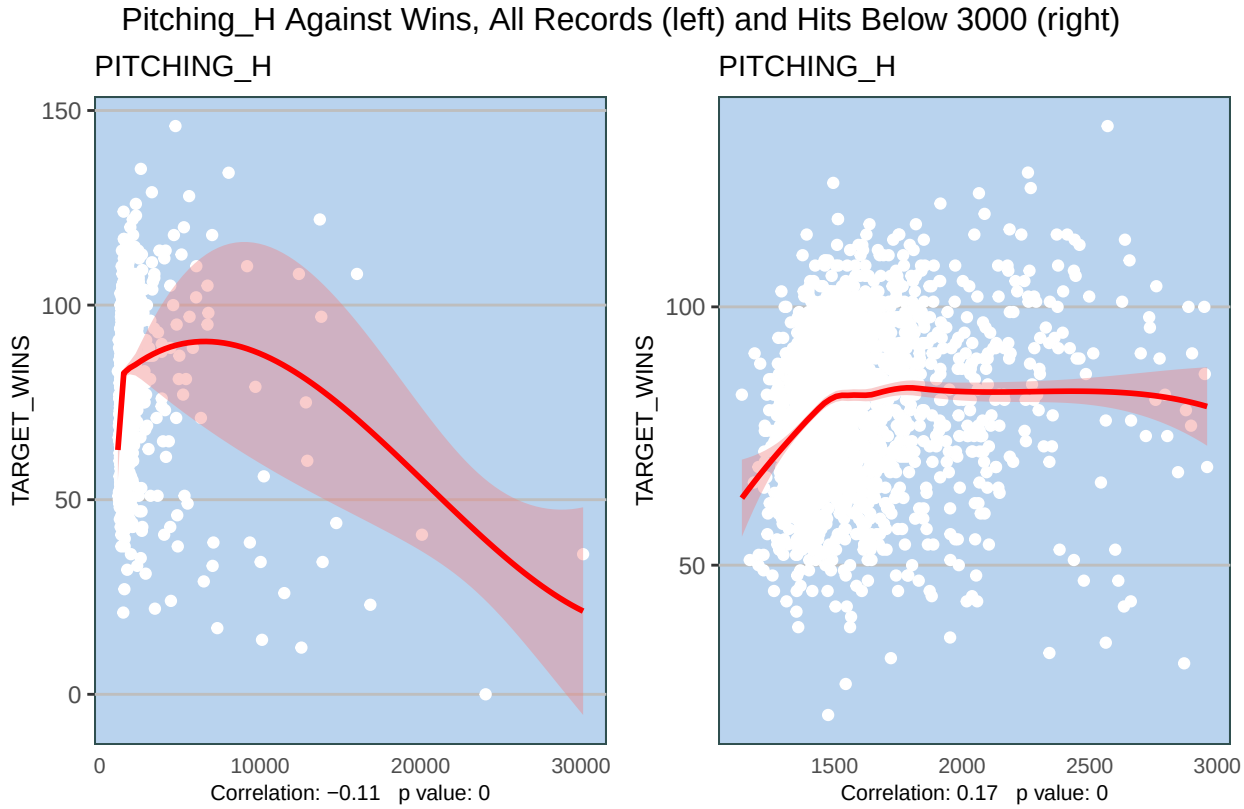


Fig.7

We can see here the positive correlation between pitching_h and wins. While we can't explain the phenomenon, we can account for it statistically by adding a binary flag for records with hits under 1500.

2. We create an interaction between Fielding_DP and hits.

The Fielding_DP correlation with Target Wins is surprising, since making double plays should help a team win. On the other hand, a team that makes double plays is also a team that gives up hits.

We therefore create an interaction term for Fielding_DP and Pitching_H.

3. We drop PITCHING_HR because it is an implausibly close match with HITTING_HR.

Like many pitching columns, Pitching_HR is unexpectedly positively correlated with wins. However, what makes this column truly implausible is how close a match it is with BATTING_HR. The scatterplot below

(Fig. 8) shows that the vast majority of the figures for pitching HR are exactly the same or within 2 or 3 of Batting HR. We therefore drop it since this makes no sense.

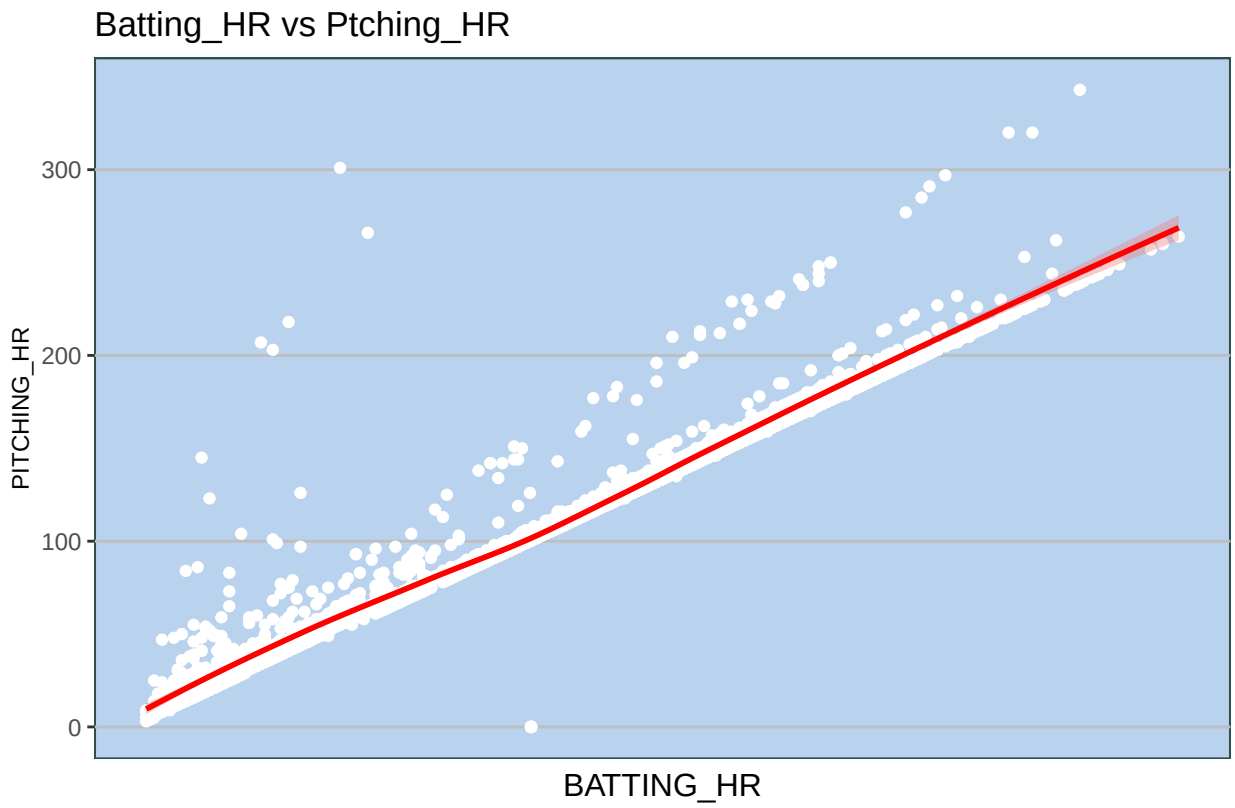


Fig. 8

4. We create a flag to account for the bimodal distribution of Batting HR.

Batting HR has a bimodal distribution (see Fig. 9). We don't explain this, but speculate that it may be related to different eras of baseball. Therefore, we create a flag to separate records with less than 80 HR form those with more.

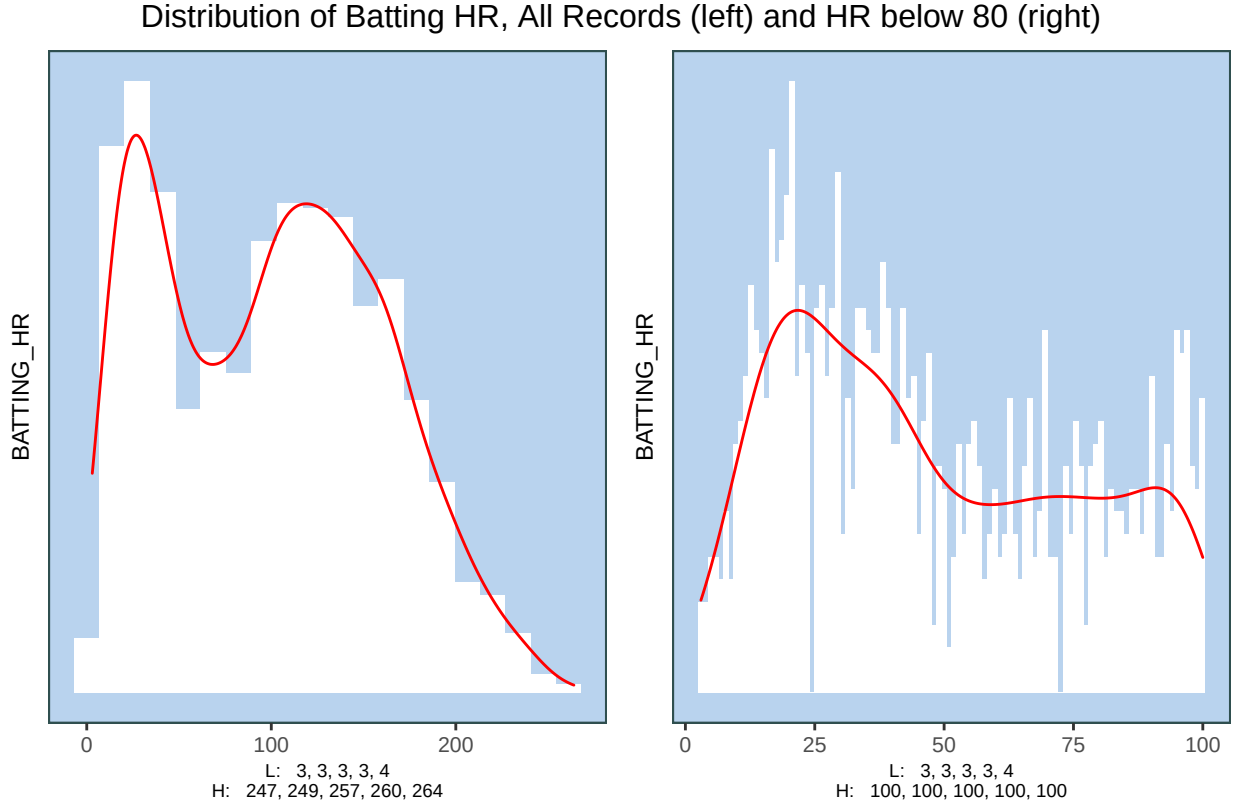


Fig.9

The numbers below represents the r-squareds of a simple linear regression of the column in question on the target variable before and after the transformation:

```
## [1] 0.02231354
```

```
## [1] 0.03503598
```

5. We transform the Fielding_err variable.

While the distributions of a number of columns suggest possible tranformations, we focus here on fielding errors, which has an upside-down u shape when correlated with wins. We therefore add an error squared term to the dataset.

The numbers below represents the r-squareds of a simple linear regression of the column in question on the target variable before and after the transformation:

```
## [1] 0.03072081
```

```
## [1] 0.04825783
```

6. We create interaction terms between the SO missing cohort and the columns identified above in the interaction analysis - Pitching_BB, Fielding E, Batting_H, Batting_HR, Batting_BB, Baserun_SB

The new fields are: Interaction_pbb_With_SO_Missing, Interaction_err_With_SO_Missing, Interaction_bh_With_SO_Missing, Interaction_bhr_With_SO_Missing, and Interaction_bbb_With_SO_Missing, Interaction_sb_With_SO_Missing.

7. For the sake of legibility, we do not create log terms for the many skewed distributions.

We would normally sacrifice some legibility for improved predictability by trying some log transformations on skewed independent variable distributions. However, legibility is already in serious peril with the odd behavior of the many pitching terms which suggest bad defense wins games. We therefore leave our transformations at those described.

4. Model Selection

Here we build and test our models to gain insight into the dataset and ultimately predict outcomes.

According to the assignmrnt: "Since we have not yet covered automated variable selection methods, you should select the variables manually (unless you previously learned Forward or Stepwise selection, etc.)." As I have learned automated and manual selection in another class, I will use automated selection, in particular the "stepAIC" package.

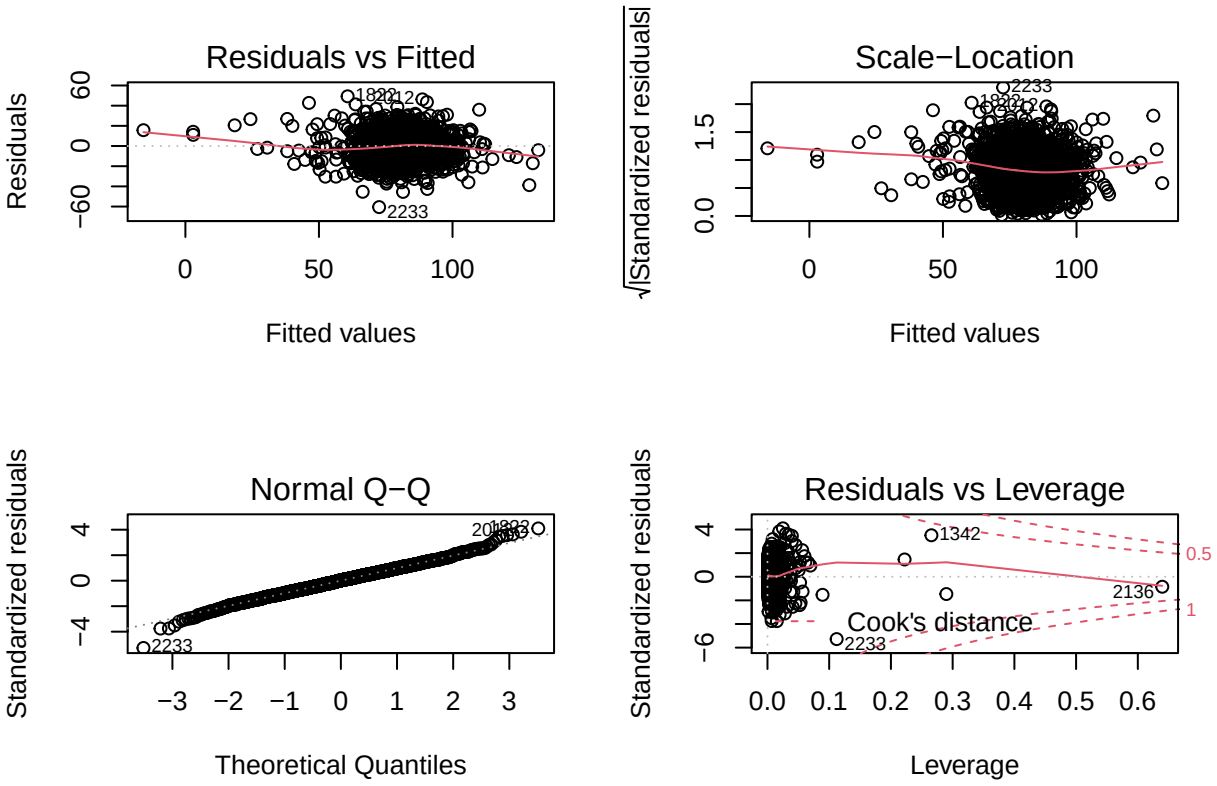
The stepAIC() function performs backward model selection by starting from a "maximal" model, which is then trimmed down. As each variable is eliminated, the Akaike Information Criterion (AIC) is calculated." The process stops when the AIC cannot be reduced by the elimnation of variables.

Because we are interested in interpretation as well as prediciton, we will modify the StepAIC model if we believe it improves readability.

a. Regression 1: Baseline (No transformations except flags for missing data)

```
##
## Call:
## lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +
##     BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +
##     PITCHING_SO + FIELDING_E + FIELDING_DP + BSO_Missing_Flag +
##     BRSB_Missing_Flag + FDP_Missing_Flag, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.531  -8.063   0.330   8.075  49.266
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.7948052  5.0143117   2.751  0.00599 **
## BATTING_H          0.0521109  0.0033520  15.546  < 2e-16 ***
## BATTING_2B        -0.0401259  0.0086621  -4.632 3.82e-06 ***
## BATTING_3B         0.0537762  0.0158617   3.390  0.00071 ***
## BATTING_HR         0.0595856  0.0089648   6.647 3.75e-11 ***
## BATTING_BB         0.0260490  0.0032618   7.986 2.20e-15 ***
## BATTING_SO        -0.0066440  0.0022278  -2.982  0.00289 **
## BASERUN_SB         0.0477764  0.0046194  10.343  < 2e-16 ***
## PITCHING_H         0.0018926  0.0003398   5.569 2.86e-08 ***
## PITCHING_SO       -0.0013966  0.0006654  -2.099  0.03593 *
## FIELDING_E        -0.0560670  0.0033748 -16.613  < 2e-16 ***
## FIELDING_DP       -0.0969459  0.0134629  -7.201 8.10e-13 ***
## BSO_Missing_Flag   8.3474206  1.4721894   5.670 1.61e-08 ***
## BRSB_Missing_Flag 34.1064444  1.8484454  18.451  < 2e-16 ***
## FDP_Missing_Flag   4.2303099  1.4669785   2.884  0.00397 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 12.17 on 2261 degrees of freedom
## Multiple R-squared:  0.4068, Adjusted R-squared:  0.4031
## F-statistic: 110.7 on 14 and 2261 DF,  p-value: < 2.2e-16
##
## [1] "VIF Analysis"
##         BATTING_H        BATTING_2B        BATTING_3B        BATTING_HR
##          3.608349          2.524443          3.016545          4.444255
##        BATTING_BB        BATTING_SO        BASERUN_SB        PITCHING_H
##          2.459248          4.131567          2.380761          3.511045
##       PITCHING_SO        FIELDING_E       FIELDING_DP  BSO_Missing_Flag
##          1.946738          9.076248          1.674220          1.425731
## BRSB_Missing_Flag  FDP_Missing_Flag
##          2.848146          3.633432
```




```
## NULL
```

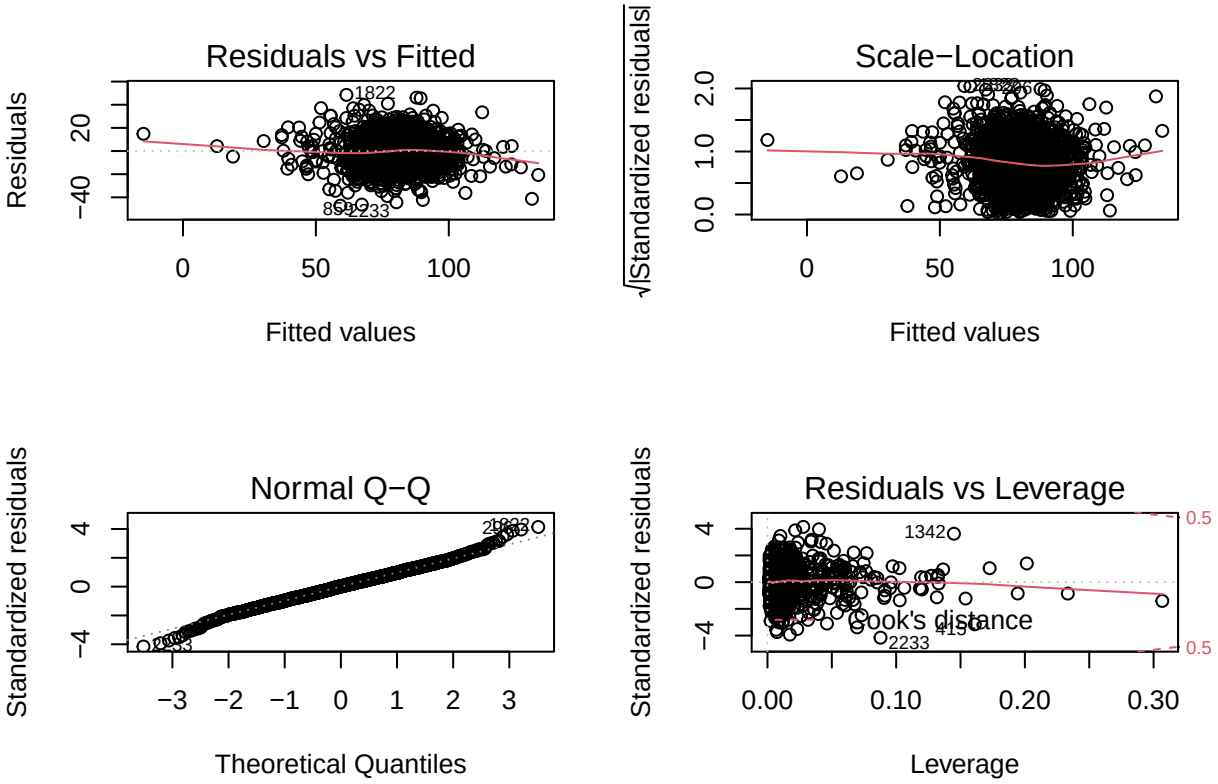
The adjusted r squared is .403. As we expected, many of the signs are in the "wrong" direction, especially for pitching. Without understanding why, we risk proceeding with a faulty model.

b. Regression 2: Include All transformations

```
##
## Call:
## lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +
```

```
##     BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +
##     FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +
##     FDP_Missing_Flag + Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq +
##     Interaction_pbb_With_SO_Missing + Interaction_err_With_SO_Missing +
##     Interaction_bhr_With_SO_Missing + Interaction_bbb_With_SO_Missing +
##     Interaction_sb_With_SO_Missing, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.202  -7.806   0.193   7.821  48.504
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     2.472e+01  6.702e+00   3.688 0.000231 ***
## BATTING_H                       5.622e-02  3.302e-03  17.023  < 2e-16 ***
## BATTING_2B                     -4.125e-02  8.586e-03  -4.805 1.65e-06 ***
## BATTING_3B                      6.743e-02  1.610e-02   4.188 2.93e-05 ***
## BATTING_HR                      5.825e-02  8.978e-03   6.488 1.06e-10 ***
## BATTING_BB                      2.593e-02  3.247e-03   7.984 2.23e-15 ***
## BATTING_SO                     -1.223e-02  2.218e-03  -5.512 3.95e-08 ***
## BASERUN_SB                      5.238e-02  4.795e-03  10.923  < 2e-16 ***
## PITCHING_H                     -4.629e-03  2.995e-03  -1.546 0.122287
## FIELDING_E                     -8.282e-02  7.453e-03 -11.112  < 2e-16 ***
## FIELDING_DP                    -1.646e-01  3.571e-02  -4.610 4.25e-06 ***
## BSO_Missing_Flag                5.042e+01  1.190e+01   4.237 2.36e-05 ***
## BRSB_Missing_Flag               3.794e+01  2.023e+00  18.752  < 2e-16 ***
## FDP_Missing_Flag                5.282e+00  1.713e+00   3.084 0.002064 **
## Pitch_h_Under1500               2.214e+00  6.829e-01   3.242 0.001206 **
## DP_times_PH                     3.671e-05  2.040e-05   1.799 0.072094 .
## Fielding_Errors_sq              2.143e-05  4.284e-06   5.002 6.11e-07 ***
## Interaction_pbb_With_SO_Missing  1.336e-01  8.560e-02   1.560 0.118847
## Interaction_err_With_SO_Missing -1.938e-01  2.809e-02  -6.899 6.77e-12 ***
## Interaction_bhr_With_SO_Missing  3.652e-01  1.546e-01   2.362 0.018285 *
## Interaction_bbb_With_SO_Missing -1.397e-01  9.536e-02  -1.465 0.143105
## Interaction_sb_With_SO_Missing   3.896e-02  2.653e-02   1.469 0.142097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.89 on 2254 degrees of freedom
## Multiple R-squared:  0.4357, Adjusted R-squared:  0.4305
## F-statistic: 82.88 on 21 and 2254 DF,  p-value: < 2.2e-16
##
## [1] "VIF Analysis"
##                       BATTING_H                       BATTING_2B
##                        3.670470                         2.599487
##                       BATTING_3B                       BATTING_HR
##                        3.258269                         4.671215
##                       BATTING_BB                       BATTING_SO
##                        2.554246                         4.292413
##                       BASERUN_SB                       PITCHING_H
##                        2.688323                       285.743188
##                       FIELDING_E                      FIELDING_DP
##                       46.392975                        12.345591
##                 BSO_Missing_Flag                BRSB_Missing_Flag
```

```
##                       97.619815                        3.575494
##                FDP_Missing_Flag                Pitch_h_Under1500
##                        5.189563                        1.863892
##                     DP_times_PH                Fielding_Errors_sq
##                      282.770320                       24.339858
## Interaction_pbb_With_SO_Missing Interaction_err_With_SO_Missing
##                     1095.501873                       50.949700
## Interaction_bhr_With_SO_Missing Interaction_bbb_With_SO_Missing
##                        7.645153                     1173.699962
##  Interaction_sb_With_SO_Missing
##                       21.438192
```




```
## NULL
```

We note that the StepAIC process included a number of variables even though they were not significant.

The second model has an adjusted r squared of .4305. This is not much better (although an ANOVA shows the p value of the improvement to be near 0). However the interpretive value of the model is greatly increased, as the coefficient signs are much more reasonable (except for Batting_2nd, which we will disregard here.)

We examine the residual plots in the model selection phase.

```
## Analysis of Variance Table
##
## Model 1: TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B + BATTING_HR +
```

```
##     BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H + PITCHING_SO +
##     FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +
##     FDP_Missing_Flag
## Model 2: TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B + BATTING_HR +
##     BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H + FIELDING_E +
##     FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag + FDP_Missing_Flag +
##     Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq + Interaction_pbb_With_SO_Missing +
##     Interaction_err_With_SO_Missing + Interaction_bhr_With_SO_Missing +
##     Interaction_bbb_With_SO_Missing + Interaction_sb_With_SO_Missing
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   2261 334871
## 2   2254 318536  7     16335 16.513 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. Regression 3: Aggregated Power Stats by Hitting and Pitching There are many more transformations possible, but we are interested here in trying a different direction - simplifying as opposed to creating a more complex model.

Throughout the analysis we have been struggling with a multicollinearity issue which we might characterize as follows:

___* Teams have limited budgets. Therefore, those with good batting may have weak pitching and vice-versa. Of course good pitching and good hitting win games - but for an individual team, the question is which wins more games - good hitting or good batting.*___

We begin by creating simple Power Hitting and Pitching Weakness scores for each team. We do this by applying a score of 1 to 5 (1 = 20th percentile and below, 5 = 80th percentile and above) for the Batting and Pitching H and BB columns of each team compared to the overall distribution. We add the pitching scores together to get a Pitching Weakness score and the batting scores for a Batting Strength score. We also subtract weakness from strength to get a Total Power score.

The number below represents the correlation between Batting Power and Pitching Weakness. We can see they are highly correlated, as we suspected. Teams are needing to balance Hitting and pitching given a limited budget:

```
## [1] 0.7257795
```

These boxplots show the relationships in each power/weakness category to overall wins. We can see the paradox at work here - the higher the pitching weakness, the higher the batting power, and the higher the wins (see Fig. 10)

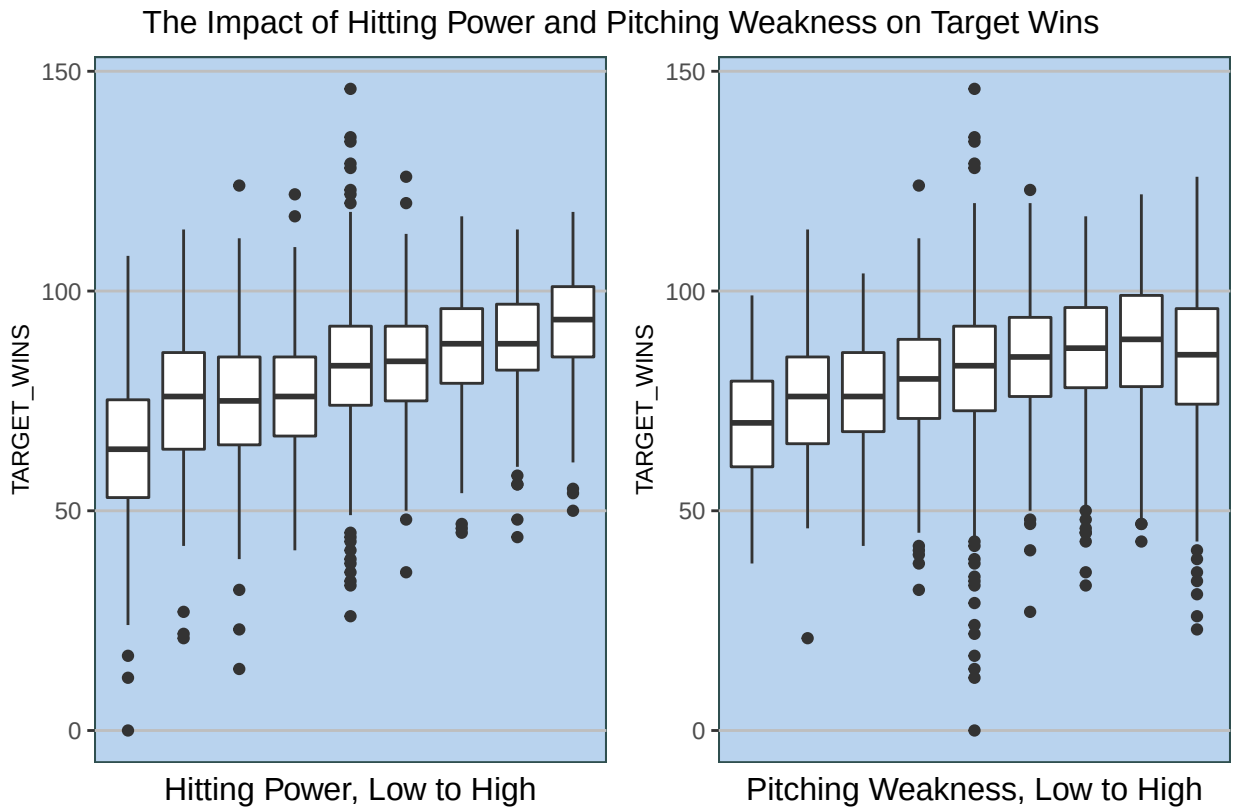


Fig. 10

we run regressions

```
##
## Call:
## lm(formula = TARGET_WINS ~ Total_Power, data = dfCat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6105 -0.6185  0.0210  0.6501  4.1398
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.650e-16  2.055e-02   0.000        1
## Total_Power  2.686e-01  2.775e-02   9.679   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9802 on 2274 degrees of freedom
## Multiple R-squared:  0.03956,    Adjusted R-squared:  0.03914
## F-statistic: 93.68 on 1 and 2274 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = TARGET_WINS ~ Hitting_Power, data = dfCat)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.3660 -0.5613  0.0128  0.5814  4.1365
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.868e-16  1.924e-02    0.00        1
## Hitting_Power  3.971e-01  1.925e-02   20.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.918 on 2274 degrees of freedom
## Multiple R-squared:  0.1577, Adjusted R-squared:  0.1573
## F-statistic: 425.6 on 1 and 2274 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = TARGET_WINS ~ Pitching_Weakness, data = dfCat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1310 -0.5709  0.0746  0.6566  4.1376
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -5.292e-16  2.030e-02     0.0        1
## Pitching_Weakness  2.498e-01  2.031e-02    12.3   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9685 on 2274 degrees of freedom
## Multiple R-squared:  0.06238,    Adjusted R-squared:  0.06196
## F-statistic: 151.3 on 1 and 2274 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = TARGET_WINS ~ Hitting_Power + Pitching_Weakness,
##     data = dfCat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2520 -0.5605  0.0277  0.5817  4.1367
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -9.018e-17  1.921e-02   0.000  1.00000
## Hitting_Power      4.560e-01  2.793e-02  16.325  < 2e-16 ***
## Pitching_Weakness -8.119e-02  2.793e-02  -2.907  0.00369 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9165 on 2273 degrees of freedom
## Multiple R-squared:  0.1608, Adjusted R-squared:   0.16
## F-statistic: 217.7 on 2 and 2273 DF,  p-value: < 2.2e-16
```

The model shows that in the balance between hitting and pitching, ***teams should emphasize good hitting and accept weak pitching.*** The adjusted r squared for the model with Hitting Power alone (.1573) is improved very little when pitching weakness is added to it (.16). The r-squared for Pitching Weakness alone is .06.

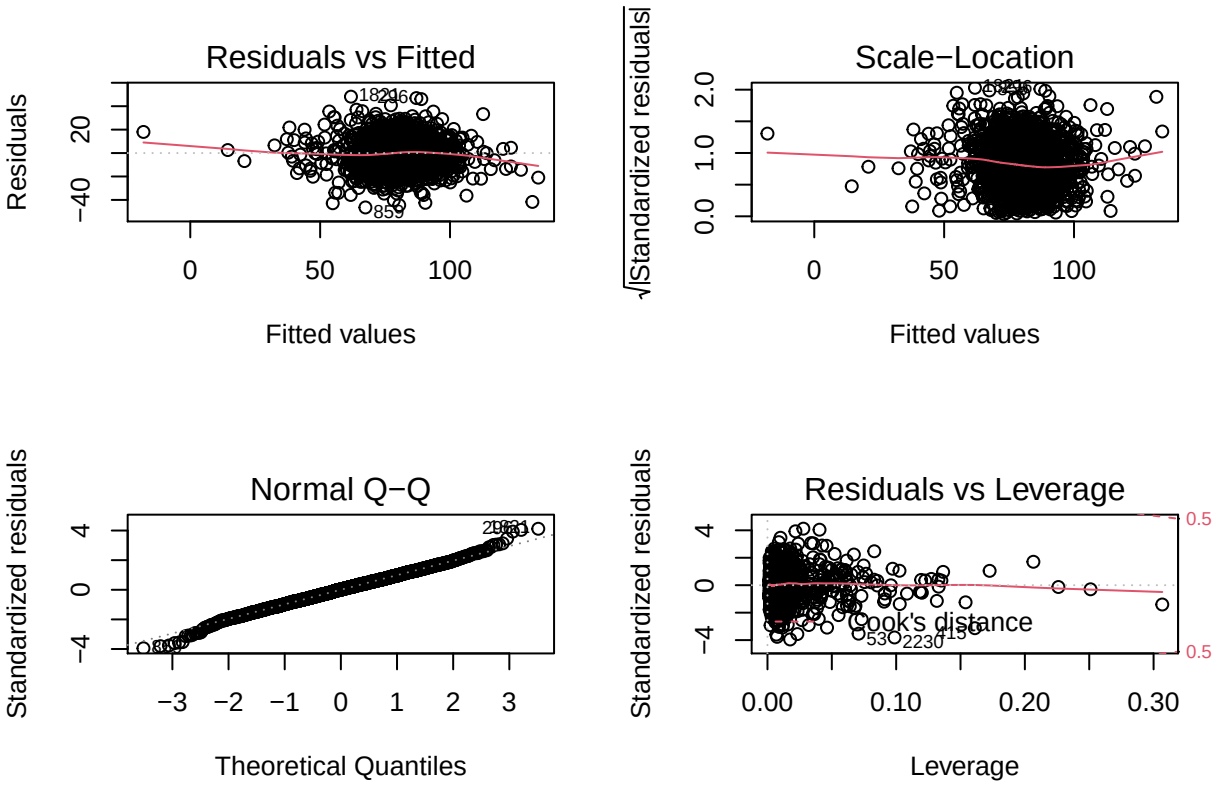
5. Select a Model and Make Predicitons

Now we make predictions. The second model has the highest R squared and reliable interpretability so we will use it for our predictions. We will first eliminate the few influential points indicated by the residual plots.

We examine the new model's output:

```
##
## Call:
## lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +
##     BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +
##     FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +
##     FDP_Missing_Flag + Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq +
##     Interaction_pbb_With_SO_Missing + Interaction_err_With_SO_Missing +
##     Interaction_bhr_With_SO_Missing + Interaction_bbb_With_SO_Missing +
##     Interaction_sb_With_SO_Missing, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.429  -7.819   0.225   7.856  48.134
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     2.374e+01  6.694e+00   3.546 0.000399 ***
## BATTING_H                       5.745e-02  3.312e-03  17.345  < 2e-16 ***
## BATTING_2B                     -4.386e-02  8.587e-03  -5.107 3.54e-07 ***
## BATTING_3B                      7.151e-02  1.610e-02   4.442 9.33e-06 ***
## BATTING_HR                      5.887e-02  8.958e-03   6.571 6.17e-11 ***
## BATTING_BB                      2.550e-02  3.241e-03   7.870 5.46e-15 ***
## BATTING_SO                     -1.222e-02  2.212e-03  -5.526 3.65e-08 ***
## BASERUN_SB                      5.202e-02  4.784e-03  10.872  < 2e-16 ***
## PITCHING_H                     -4.611e-03  2.996e-03  -1.539 0.123868
## FIELDING_E                     -8.444e-02  7.464e-03 -11.313  < 2e-16 ***
## FIELDING_DP                    -1.605e-01  3.568e-02  -4.499 7.16e-06 ***
## BSO_Missing_Flag                5.031e+01  1.187e+01   4.240 2.33e-05 ***
## BRSB_Missing_Flag               3.728e+01  2.027e+00  18.393  < 2e-16 ***
## FDP_Missing_Flag                5.210e+00  1.724e+00   3.023 0.002531 **
## Pitch_h_Under1500               2.243e+00  6.810e-01   3.294 0.001004 **
## DP_times_PH                     3.390e-05  2.042e-05   1.660 0.097107 .
## Fielding_Errors_sq              2.396e-05  4.335e-06   5.527 3.63e-08 ***
## Interaction_pbb_With_SO_Missing  1.347e-01  8.535e-02   1.578 0.114592
## Interaction_err_With_SO_Missing -1.939e-01  2.801e-02  -6.923 5.75e-12 ***
## Interaction_bhr_With_SO_Missing  3.605e-01  1.542e-01   2.338 0.019481 *
## Interaction_bbb_With_SO_Missing -1.403e-01  9.508e-02  -1.476 0.140113
## Interaction_sb_With_SO_Missing   3.857e-02  2.645e-02   1.458 0.144977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 11.85 on 2251 degrees of freedom
## Multiple R-squared:  0.4384, Adjusted R-squared:  0.4331
## F-statistic: 83.67 on 21 and 2251 DF,  p-value: < 2.2e-16
##
## [1] "VIF Analysis"
##                        BATTING_H                      BATTING_2B
##                         3.707874                        2.607807
##                      BATTING_3B                      BATTING_HR
##                         3.260919                        4.673853
##                      BATTING_BB                      BATTING_SO
##                         2.547589                        4.287409
##                      BASERUN_SB                      PITCHING_H
##                         2.691815                      274.568838
##                      FIELDING_E                     FIELDING_DP
##                        46.562944                       12.396781
##                BSO_Missing_Flag                BRSB_Missing_Flag
##                       97.625954                        3.558545
##                FDP_Missing_Flag                Pitch_h_Under1500
##                       5.255244                        1.862688
##                     DP_times_PH              Fielding_Errors_sq
##                      271.957776                       25.000344
## Interaction_pbb_With_SO_Missing Interaction_err_With_SO_Missing
##                     1095.487226                       50.957782
## Interaction_bhr_With_SO_Missing Interaction_bbb_With_SO_Missing
##                        7.645237                     1173.677967
##  Interaction_sb_With_SO_Missing
##                       21.438268
```




```
## NULL
```

There is only a slight improvement with the elimination of influential points. We note that the adjusted r-squared is still best among all models. The F statistic shows the model is significant overall. The residual standard error is small relative to the target variable. We see no patterning in the residuals and the distribution is relatively normal, except at the tails.

There are new influential points after eliminating the others, but we accept them without any better reason to challenge them. A VIF analysis shows a fair amount of multicolinnearity, but we knew this, and even created more with our interaction terms. In all, we can move forward with this model without further modification.

Predictions

```
##      predict(m, newdata = dfEval2)
## 1                         65.55191
## 2                         73.00183
## 3                         74.25101
## 4                         69.95245
## 5                         66.80497
## 6                         70.24417
## 7                         66.29278
## 8                         71.06411
## 9                         74.46557
## 10                        66.37215
```

```
## 11                    69.46976
## 12                    77.74397
## 13                    76.93263
## 14                    70.89365
## 15                    83.13329
## 16                    82.84444
## 17                    88.78233
## 18                    81.68573
## 19                    75.63521
## 20                    86.88036
## 21                    74.96393
## 22                    76.82949
## 23                    78.60665
## 24                    87.13806
## 25                    83.45039
## 26                    77.23266
## 27                    78.86895
## 28                    76.38967
## 29                    76.77439
## 30                    70.04479
## 31                    80.91321
## 32                    83.52546
## 33                    91.89455
## 34                    85.97919
## 35                    88.92170
## 36                    82.22702
## 37                    78.10795
## 38                    86.95044
## 39                    83.69726
## 40                    89.93390
## 41                    87.63146
## 42                    95.15289
## 43                    89.73065
## 44                    78.31297
## 45                    96.70592
## 46                    87.44049
## 47                    89.78451
## 48                    92.55031
## 49                    84.34227
## 50                    71.77717
## 51                    81.43123
## 52                    78.69631
## 53                    76.85425
## 54                    79.21440
## 55                   100.25702
## 56                    99.14342
## 57                    83.16530
## 58                   100.35466
## 59                    97.90465
## 60                    94.72824
## 61                    78.09780
## 62                    57.45653
## 63                    88.51450
## 64                    68.32624
```

```
## 65                      86.80110
## 66                      84.55948
## 67                      87.11471
## 68                     100.41104
## 69                     115.49921
## 70                      93.60006
## 71                      97.35404
## 72                     113.95254
## 73                      92.26079
## 74                      93.98473
## 75                      94.75779
## 76                      77.01866
## 77                      57.46773
## 78                      57.70466
## 79                      61.15589
## 80                      44.13604
## 81                      82.02524
## 82                      81.20671
## 83                      48.95951
## 84                      63.68842
## 85                      74.95106
## 86                      74.49601
## 87                      78.14866
## 88                      72.43367
## 89                      70.70922
## 90                      75.58188
## 91                      71.74207
## 92                      76.26566
## 93                      74.30360
## 94                      91.38236
## 95                      75.23450
## 96                      73.62721
## 97                      71.27857
## 98                      82.87367
## 99                      77.58584
## 100                     80.05358
## 101                     77.53620
## 102                     77.51519
## 103                     79.55527
## 104                     72.22418
## 105                     76.07731
## 106                     72.63919
## 107                     81.77392
## 108                     74.63977
## 109                     71.75797
## 110                     72.39270
## 111                     69.54975
## 112                     68.50319
## 113                     71.87177
## 114                     65.35577
## 115                     67.93539
## 116                     80.53045
## 117                     78.99818
## 118                     84.78329
```

```
## 119                    88.66635
## 120                    79.83718
## 121                    83.09723
## 122                    81.80781
## 123                    70.75139
## 124                    80.68044
## 125                    80.84626
## 126                    86.97092
## 127                    84.79816
## 128                    84.21661
## 129                    84.05194
## 130                    81.29710
## 131                    75.61989
## 132                    82.11703
## 133                    77.91876
## 134                    85.81814
## 135                    73.92727
## 136                    78.22515
## 137                    89.57097
## 138                    79.48114
## 139                    88.44021
## 140                    79.21477
## 141                    73.00361
## 142                    72.64040
## 143                    79.11972
## 144                    78.02837
## 145                    77.01566
## 146                    70.81433
## 147                    75.44452
## 148                    80.55652
## 149                    87.44070
## 150                    76.58422
## 151                    69.97746
## 152                    72.22479
## 153                    83.30746
## 154                    71.31659
## 155                    75.26072
## 156                    86.58829
## 157                    86.33451
## 158                    87.87217
## 159                    84.89918
## 160                    80.70021
## 161                    84.76539
## 162                    88.20931
## 163                    89.48089
## 164                    92.98312
## 165                    90.00451
## 166                    83.03176
## 167                    89.25453
## 168                    80.10568
## 169                    81.92140
## 170                    82.56450
## 171                    66.27968
## 172                    73.35852
```

```
## 173                    61.02738
## 174                    72.28051
## 175                    84.80314
## 176                    77.96885
## 177                    84.73929
## 178                    67.83358
## 179                    37.20481
## 180                    48.14760
## 181                    60.87088
## 182                    64.83300
## 183                    66.87822
## 184                    63.66113
## 185                    79.05949
## 186                    65.35653
## 187                    76.64316
## 188                    75.60494
## 189                    93.95756
## 190                    94.75443
## 191                    99.44551
## 192                    84.28355
## 193                    89.17281
## 194                    74.04649
## 195                    72.01181
## 196                    80.56895
## 197                    76.79435
## 198                    73.20580
## 199                    81.34510
## 200                    75.67742
## 201                    73.85718
## 202                    79.83315
## 203                    84.62222
## 204                    80.06155
## 205                    79.57510
## 206                    77.37831
## 207                    86.77487
## 208                    74.98901
## 209                    74.99894
## 210                    77.05436
## 211                    72.09386
## 212                    74.39570
## 213                    76.96458
## 214                    75.75151
## 215                    71.66229
## 216                    67.26449
## 217                    71.63299
## 218                    72.01206
## 219                    68.80844
## 220                    73.46381
## 221                    77.22493
## 222                    70.22170
## 223                    70.19984
## 224                    77.17477
## 225                    76.88357
## 226                    76.49158
```

```
## 227                    78.90521
## 228                    77.89276
## 229                    72.68706
## 230                    83.51177
## 231                    74.67146
## 232                    69.84811
## 233                    91.66123
## 234                    89.04173
## 235                    87.00029
## 236                    68.06005
## 237                    89.05831
## 238                    80.87576
## 239                    81.45000
## 240                    78.20661
## 241                    80.03334
## 242                    79.36362
## 243                    84.27763
## 244                    88.32514
## 245                    77.14155
## 246                    90.41355
## 247                    85.65255
## 248                    78.14958
## 249                    86.69512
## 250                    80.27654
## 251                    83.14703
## 252                    70.45635
## 253                    81.64373
## 254                    80.74462
## 255                    79.38861
## 256                    86.29034
## 257                    84.13201
## 258                    93.19836
## 259                    87.91209
## 260                    92.72209
## 261                    89.01357
## 262                    90.43686
## 263                    89.65688
## 264                    86.16842
## 265                    78.53071
## 266                    73.42320
## 267                    87.43721
## 268                    83.89223
## 269                    70.97937
## 270                    78.01342
## 271                    76.46185
## 272                    70.01677
## 273                    79.78304
## 274                    49.91156
## 275                    68.89681
## 276                    82.42838
## 277                    79.35215
## 278                    49.23907
## 279                    96.44614
## 280                    62.78509
```

```
## 281  68.31919
## 282  78.69826
## 283  80.67865
## 284  69.58553
## 285  90.73698
## 286 133.66380
## 287 113.08028
## 288 127.12961
## 289 123.52734
## 290 101.66217
## 291  99.40436
## 292  77.12302
## 293  63.92972
## 294  63.82188
## 295 102.06470
## 296  87.85726
## 297  90.71771
## 298 110.03213
## 299 112.56379
## 300  83.43744
## 301  83.80663
## 302  83.40125
## 303  76.09605
## 304  78.50300
## 305  52.20070
## 306  72.91166
## 307  74.42062
## 308  81.81497
## 309  91.59737
## 310  83.32705
## 311  98.57849
## 312  84.12138
## 313  78.26140
## 314  79.18204
## 315  75.95404
## 316  74.90443
## 317  74.61564
## 318  80.67474
## 319  81.92859
## 320  79.55470
## 321  68.79082
## 322  67.90819
## 323  71.65708
## 324  68.67376
## 325  73.03734
## 326  66.92058
## 327  67.22255
## 328  65.41780
## 329  76.53480
## 330  85.41358
## 331  85.24160
## 332  86.33617
## 333  89.95231
## 334  90.18323
```

```
## 335  90.03554
## 336  91.31450
## 337  72.13036
## 338  79.27497
## 339  71.75159
## 340  84.08481
## 341  90.29398
## 342  90.41369
## 343 102.39814
## 344  83.65855
## 345  74.11608
## 346  73.24061
## 347  81.53293
## 348  85.48891
## 349  77.16766
## 350  77.04831
## 351  77.90526
## 352  75.38979
## 353  79.96851
## 354  79.70842
## 355  80.91207
## 356  79.33613
## 357  70.39016
## 358  78.95200
## 359  70.93105
## 360  77.65852
## 361  78.13204
## 362  86.93608
## 363  81.72658
## 364  84.71531
## 365  82.57779
## 366  83.90884
## 367  85.88070
## 368  80.95518
## 369  81.84658
## 370  78.77067
## 371  87.04464
## 372  89.31990
## 373  83.50006
## 374  80.00904
## 375  80.99053
## 376  68.18538
## 377  76.51930
## 378  79.66077
## 379  87.28799
## 380  92.64253
## 381  83.99093
## 382  88.77144
## 383  87.39615
## 384  84.82999
## 385  81.29013
## 386  86.73933
## 387  98.59211
## 388  90.19458
```

```
## 389  88.54704
## 390  84.19735
## 391  30.35911
## 392  37.18246
## 393  50.01622
## 394  55.63881
## 395  52.63914
## 396  99.48903
## 397  77.52753
## 398  80.32902
## 399  82.08776
## 400  65.10853
## 401  79.61325
## 402  98.14394
## 403  96.36429
## 404  96.37953
## 405  83.54787
## 406  84.61772
## 407  76.21535
## 408  89.88489
## 409  64.43448
## 410  80.26479
## 411  86.19696
## 412  75.83988
## 413  78.07354
## 414  71.70651
## 415  57.37223
## 416  63.15380
## 417  64.09999
## 418 123.67045
## 419  93.47920
## 420  92.15568
## 421  77.93867
## 422  94.12695
## 423 102.26343
## 424  90.29505
## 425  74.83494
## 426 101.16465
## 427  97.35455
## 428  77.09618
## 429  91.87652
## 430  76.70486
## 431  81.27233
## 432  83.95916
## 433  70.17944
## 434 106.26012
## 435  99.16509
## 436  92.99587
## 437  81.41676
## 438  82.07542
## 439  66.14591
## 440  72.30282
## 441  69.25086
## 442  84.68774
```

```
## 443  85.36080
## 444  97.23854
## 445 120.42918
## 446  99.14138
## 447 104.74468
## 448  88.16274
## 449  83.79807
## 450  85.50493
## 451  82.31094
## 452  82.38826
## 453  72.55288
## 454  67.35067
## 455  82.15356
## 456  77.14821
## 457  85.83237
## 458  86.97012
## 459  88.76476
## 460  79.54775
## 461  79.50107
## 462  76.33728
## 463  82.74157
## 464  82.12601
## 465  97.93463
## 466 101.02765
## 467  86.24322
## 468  79.09542
## 469  83.72734
## 470  84.38245
## 471  92.38569
## 472  84.14744
## 473  80.05130
## 474  76.41641
## 475  73.56031
## 476  71.94474
## 477  80.03960
## 478  77.16957
## 479  90.41803
## 480  79.64968
## 481  72.45204
## 482  74.41907
## 483  71.93787
## 484  70.36367
## 485  74.01667
## 486  79.84363
## 487  77.48091
## 488  79.15073
## 489  76.29472
## 490  76.27345
## 491  72.87925
## 492  79.80243
## 493  77.70053
## 494  76.31867
## 495  70.65869
## 496  75.45789
```

```
## 497                    64.78003
## 498                    69.28038
## 499                    78.37189
## 500                    79.46774
## 501                    78.23190
## 502                    84.49303
## 503                    80.07494
## 504                    81.78630
## 505                    72.79291
## 506                    77.21512
## 507                    78.37254
## 508                    75.96984
## 509                    80.65548
## 510                    83.12651
## 511                    80.83137
## 512                    68.32050
## 513                    85.53989
## 514                    79.15487
## 515                    87.33907
## 516                    85.10284
## 517                    82.32810
## 518                    86.06251
## 519                    83.68392
## 520                    85.68516
## 521                    86.41419
## 522                    85.56330
## 523                    82.68400
## 524                    79.60344
## 525                    83.34137
## 526                    79.79234
## 527                    83.55843
## 528                    92.18494
## 529                    79.88232
## 530                    85.37272
## 531                    85.78998
## 532                    76.77021
## 533                    76.21532
## 534                    87.64768
## 535                    81.93706
## 536                    89.00685
## 537                    76.65082
## 538                    82.42557
## 539                    79.66558
## 540                    83.08141
## 541                    66.41185
## 542                    96.65180
## 543                    85.55320
## 544                    83.10899
## 545                    84.58586
## 546                    78.22706
## 547                    97.38683
## 548                    67.11364
## 549                    69.67822
## 550                    88.96917
```

```
## 551                    78.23674
## 552                    86.20925
## 553                    77.40755
## 554                    91.01169
## 555                    81.79632
## 556                    85.95612
## 557                    88.27243
## 558                    87.54006
## 559                    84.19177
## 560                    89.62574
## 561                    80.44861
## 562                    79.51266
## 563                    77.56037
## 564                    79.36415
## 565                    76.16739
## 566                    69.06006
## 567                    81.47053
## 568                    77.11984
## 569                    86.89254
## 570                    92.68498
## 571                    81.50845
## 572                    80.12239
## 573                    87.54080
## 574                    84.48123
## 575                    75.80476
## 576                    71.99633
## 577                    77.68747
## 578                    70.30445
## 579                    74.95049
## 580                    71.15396
## 581                    74.85165
## 582                    75.26314
## 583                    76.91339
## 584                    85.79335
## 585                    80.85518
## 586                    84.59574
## 587                    88.38748
## 588                    86.73369
## 589                    86.41457
## 590                    84.26103
## 591                    83.65559
## 592                    84.70819
## 593                    84.27259
## 594                    91.77060
## 595                    84.47208
## 596                    77.04129
## 597                    77.01515
## 598                    75.58869
## 599                    69.66089
## 600                    69.00880
## 601                    64.50346
## 602                    74.30459
## 603                    69.62066
## 604                    79.62800
```

```
## 605  72.26135
## 606  75.24082
## 607  76.62281
## 608  81.64100
## 609  80.32722
## 610  92.35871
## 611  82.25820
## 612  80.26687
## 613  71.49645
## 614  84.88933
## 615  83.03989
## 616  86.15808
## 617  80.72471
## 618  80.58360
## 619  78.35616
## 620  80.58767
## 621  69.38983
## 622  74.32166
## 623  77.14420
## 624  89.67470
## 625  88.23286
## 626  89.92309
## 627 103.65341
## 628  95.12974
## 629  96.54155
## 630  86.03338
## 631  87.48121
## 632  88.59853
## 633  86.16896
## 634  87.19663
## 635  84.66026
## 636  91.89178
## 637  95.37146
## 638 103.81439
## 639 100.75417
## 640  96.46755
## 641  58.82565
## 642  94.60931
## 643  81.61702
## 644  92.16593
## 645  88.68448
## 646  73.16406
## 647  77.79015
## 648  90.75491
## 649  91.40286
## 650 101.24729
## 651  89.96234
## 652  84.98862
## 653  89.41150
## 654  76.17704
## 655  87.02351
## 656  81.76827
## 657  82.69916
## 658  82.04981
```

```
## 659                    89.26857
## 660                    81.88309
## 661                    88.60227
## 662                    86.08641
## 663                    83.66168
## 664                    67.33915
## 665                    73.03391
## 666                    76.42616
## 667                    76.74552
## 668                    79.52607
## 669                    86.70721
## 670                    85.54581
## 671                    81.45759
## 672                    90.55624
## 673                    85.59157
## 674                    86.08666
## 675                    83.77530
## 676                    79.11271
## 677                    76.30380
## 678                    85.45075
## 679                    79.13433
## 680                    69.18043
## 681                    77.08416
## 682                    67.00805
## 683                    74.45282
## 684                    74.73076
## 685                    80.04763
## 686                    72.12182
## 687                    82.94771
## 688                    81.34694
## 689                    81.16054
## 690                    74.85504
## 691                    66.27869
## 692                    74.80237
## 693                    73.13457
## 694                    67.08773
## 695                    80.78265
## 696                    73.97263
## 697                    75.80307
## 698                    78.25924
## 699                    74.20176
## 700                    77.22671
## 701                    79.26142
## 702                    84.19971
## 703                    89.59626
## 704                    81.63889
## 705                    83.27089
## 706                    78.28785
## 707                    82.94053
## 708                    83.58153
## 709                    76.37563
## 710                    77.98625
## 711                    89.34058
## 712                    81.40547
```

```
## 713  75.74777
## 714  85.12236
## 715  70.60691
## 716  84.89606
## 717  86.62123
## 718  89.55312
## 719 101.25862
## 720  93.89528
## 721  85.60490
## 722  83.84923
## 723  87.60087
## 724  79.62832
## 725  76.29905
## 726  79.09720
## 727  81.15734
## 728  84.36782
## 729  84.27746
## 730  90.96940
## 731  81.81429
## 732  81.34202
## 733  88.24698
## 734  85.34721
## 735  84.26292
## 736  98.21037
## 737  90.98895
## 738  86.53305
## 739  83.98766
## 740  80.45527
## 741  91.70571
## 742  87.78883
## 743  78.74698
## 744  75.25194
## 745  71.49615
## 746  79.60788
## 747  82.50879
## 748  83.52548
## 749  66.42791
## 750  70.51211
## 751  81.71638
## 752  67.97215
## 753  74.85185
## 754  93.36413
## 755 103.43174
## 756  73.23282
## 757  69.91817
## 758  84.38538
## 759  78.51857
## 760  76.49596
## 761  67.87962
## 762  72.75711
## 763  71.56563
## 764  71.69259
## 765  81.74595
## 766  81.36179
```

```
## 767                    92.00141
## 768                    93.18933
## 769                    85.65661
## 770                    90.31655
## 771                    86.83804
## 772                    88.40980
## 773                    80.20965
## 774                    72.32151
## 775                    70.44647
## 776                    82.62767
## 777                    82.23854
## 778                    86.45477
## 779                    88.76188
## 780                    80.01928
## 781                    87.15364
## 782                    90.49966
## 783                    83.99484
## 784                    78.09105
## 785                    80.82639
## 786                    73.72852
## 787                    70.69014
## 788                    72.61922
## 789                    73.28473
## 790                    78.74572
## 791                    93.68864
## 792                    81.78455
## 793                    90.23493
## 794                    84.19472
## 795                    81.86079
## 796                    85.40427
## 797                    86.52819
## 798                    81.63614
## 799                    77.37618
## 800                    73.64483
## 801                    80.39197
## 802                    81.87908
## 803                    72.62797
## 804                    73.04517
## 805                    78.56846
## 806                    75.08917
## 807                    70.25776
## 808                    71.80706
## 809                    72.18240
## 810                    74.38915
## 811                    70.18816
## 812                    66.98077
## 813                    78.47174
## 814                    76.39887
## 815                    85.83444
## 816                    82.36609
## 817                    81.99555
## 818                    88.35387
## 819                    93.19957
## 820                    85.68540
```

```
## 821  82.39637
## 822  87.53059
## 823  81.45814
## 824  89.03431
## 825  86.81023
## 826  81.63018
## 827  77.46630
## 828  82.38221
## 829  72.07170
## 830  81.47586
## 831  84.05746
## 832  98.77947
## 833  99.38846
## 834 100.88540
## 835  93.93740
## 836  99.28719
## 837 100.71500
## 838  90.46947
## 839  75.87247
## 840  72.00580
## 841  87.36214
## 842  80.83271
## 843  80.26024
## 844  80.03629
## 845  76.15235
## 846  86.66554
## 847  72.88414
## 848  66.81290
## 849  63.60122
## 850  63.45416
## 851  75.54693
## 852  79.91606
## 853 103.22461
## 854 100.58021
## 855  94.96774
## 856 103.48944
## 857  98.03021
## 858  81.54804
## 859  67.28309
## 860  75.75525
## 861  81.46042
## 862  55.04906
## 863  70.83736
## 864  93.79849
## 865  82.57492
## 866  63.98096
## 867  79.88604
## 868  86.18626
## 869  92.38099
## 870  91.87309
## 871  91.89568
## 872  84.35835
## 873  89.98057
## 874  95.84451
```

```
## 875                    96.21643
## 876                    81.71168
## 877                    80.79532
## 878                    82.03579
## 879                    78.80142
## 880                    78.98589
## 881                    85.89717
## 882                    62.35801
## 883                    73.02447
## 884                    73.20018
## 885                    67.21956
## 886                    73.33752
## 887                    82.23962
## 888                    74.19379
## 889                   102.79514
## 890                    84.51132
## 891                    81.73457
## 892                    82.73832
## 893                    86.43914
## 894                    82.36875
## 895                    78.24434
## 896                    86.18695
## 897                    80.62364
## 898                    96.14499
## 899                    94.53938
## 900                    91.92217
## 901                    98.07684
## 902                    84.20501
## 903                    92.04255
## 904                    83.35374
## 905                    78.32133
## 906                    83.89502
## 907                    95.70668
## 908                    98.95877
## 909                    95.02099
## 910                    97.70491
## 911                    88.79770
## 912                    87.01503
## 913                    93.77405
## 914                    80.36583
## 915                    78.59813
## 916                    72.71940
## 917                    73.44546
## 918                    83.08230
## 919                    84.66109
## 920                    83.46122
## 921                    85.73835
## 922                    88.59571
## 923                    78.33340
## 924                    75.67007
## 925                    81.44338
## 926                    84.87941
## 927                    92.86417
## 928                    79.01968
```

```
## 929                    80.72064
## 930                    81.12077
## 931                    92.65951
## 932                    87.75298
## 933                    86.42809
## 934                    84.70339
## 935                    77.99316
## 936                    81.29795
## 937                    79.07916
## 938                    78.49956
## 939                    79.95111
## 940                    76.89701
## 941                    82.44670
## 942                    72.46132
## 943                    82.19255
## 944                    74.58769
## 945                    72.37143
## 946                    76.35055
## 947                    84.04889
## 948                    86.09805
## 949                    88.80999
## 950                    88.83695
## 951                    80.62807
## 952                    86.77966
## 953                    87.17209
## 954                    89.24867
## 955                    82.31457
## 956                    90.91136
## 957                    93.48210
## 958                    84.35778
## 959                    76.33932
## 960                    80.89009
## 961                    82.37508
## 962                    81.07603
## 963                    90.80765
## 964                    87.53160
## 965                    75.63424
## 966                    86.65658
## 967                    80.80470
## 968                    82.73490
## 969                    85.29212
## 970                    79.53801
## 971                    66.81776
## 972                    67.43057
## 973                    82.44908
## 974                    77.59490
## 975                    81.40484
## 976                    78.42875
## 977                    84.37256
## 978                    83.27324
## 979                    86.40346
## 980                   104.54866
## 981                    72.60307
## 982                    12.79520
```

```
## 983                    77.75955
## 984                    77.40356
## 985                    83.19662
## 986                    76.47393
## 987                    80.53532
## 988                    71.26224
## 989                    78.39454
## 990                    83.80685
## 991                    84.05029
## 992                    87.19177
## 993                    83.11745
## 994                    80.62727
## 995                    77.43872
## 996                    86.48362
## 997                    89.87515
## 998                    64.38198
## 999                    84.54957
## 1000                   75.01933
## 1001                   64.66463
## 1002                   64.27674
## 1003                   73.31624
## 1004                   78.53816
## 1005                   76.46953
## 1006                   62.96596
## 1007                   76.95265
## 1008                   85.14973
## 1009                   73.90941
## 1010                   81.07319
## 1011                   80.25418
## 1012                   81.90385
## 1013                   80.64928
## 1014                   82.46304
## 1015                   85.43341
## 1016                   85.76962
## 1017                   80.56681
## 1018                   88.61970
## 1019                   80.86555
## 1020                   73.80244
## 1021                   80.65574
## 1022                   82.17007
## 1023                   78.16364
## 1024                   86.46055
## 1025                   86.19108
## 1026                   80.82279
## 1027                   78.01799
## 1028                   79.47738
## 1029                   73.27587
## 1030                   78.68482
## 1031                   82.31318
## 1032                   89.14470
## 1033                   90.91917
## 1034                   84.03309
## 1035                   80.68875
## 1036                   90.67240
```

```
## 1037                    88.13552
## 1038                    92.44489
## 1039                    86.23471
## 1040                    83.07098
## 1041                    84.23018
## 1042                    81.17885
## 1043                    81.25842
## 1044                    72.23148
## 1045                    70.62632
## 1046                    68.88836
## 1047                    61.16895
## 1048                    80.04411
## 1049                    48.58736
## 1050                    74.78615
## 1051                    72.27362
## 1052                    69.92483
## 1053                    78.63857
## 1054                    72.09716
## 1055                    71.85381
## 1056                    76.18837
## 1057                    79.68996
## 1058                    81.99411
## 1059                    89.50301
## 1060                    93.45649
## 1061                    95.01378
## 1062                    87.46678
## 1063                    80.18852
## 1064                    80.14900
## 1065                    78.39498
## 1066                    84.11826
## 1067                    80.64831
## 1068                    86.81757
## 1069                    82.98826
## 1070                    76.09981
## 1071                    79.95818
## 1072                    88.19643
## 1073                    78.33208
## 1074                    84.59843
## 1075                    87.26213
## 1076                    80.01118
## 1077                    90.29308
## 1078                    75.23493
## 1079                    86.70392
## 1080                    73.84854
## 1081                    73.23477
## 1082                    39.65184
## 1083                    39.52768
## 1084                    72.26446
## 1085                    90.15345
## 1086                    61.46690
## 1087                    81.22847
## 1088                    90.91384
## 1089                    93.98991
## 1090                    78.58497
```

```
## 1091                    95.76005
## 1092                    83.92764
## 1093                   108.47951
## 1094                    89.37229
## 1095                    89.92951
## 1096                    84.10034
## 1097                    73.44865
## 1098                    97.11571
## 1099                    86.81700
## 1100                    74.41367
## 1101                    96.94630
## 1102                    47.22243
## 1103                    79.76360
## 1104                    68.56332
## 1105                    72.78138
## 1106                    81.13607
## 1107                    81.37785
## 1108                    80.11446
## 1109                    73.69191
## 1110                    81.58878
## 1111                    75.95082
## 1112                    73.19684
## 1113                    77.84913
## 1114                    82.43758
## 1115                    80.11967
## 1116                    81.84214
## 1117                    86.33281
## 1118                    75.60293
## 1119                    81.89980
## 1120                    78.63579
## 1121                    79.23169
## 1122                    80.51000
## 1123                    80.59576
## 1124                    79.91008
## 1125                    72.59089
## 1126                    79.80484
## 1127                    77.34458
## 1128                    83.93659
## 1129                    88.50735
## 1130                    81.00054
## 1131                    80.25613
## 1132                    82.48203
## 1133                    83.23569
## 1134                    88.31159
## 1135                    81.65745
## 1136                    95.21848
## 1137                    89.68064
## 1138                    89.23232
## 1139                    88.61033
## 1140                    99.08462
## 1141                    91.42910
## 1142                    91.39095
## 1143                    87.18175
## 1144                    82.91470
```

```
## 1145                    75.02118
## 1146                    85.27623
## 1147                    82.02620
## 1148                    82.88688
## 1149                    91.66510
## 1150                    78.79512
## 1151                    77.64903
## 1152                    78.74921
## 1153                    79.71059
## 1154                    67.30516
## 1155                    62.30735
## 1156                    84.14282
## 1157                    88.87729
## 1158                    81.36676
## 1159                    77.70468
## 1160                    83.18373
## 1161                    90.99038
## 1162                    87.47545
## 1163                    80.19431
## 1164                    89.99785
## 1165                    87.47006
## 1166                    91.13616
## 1167                    86.01049
## 1168                    82.48761
## 1169                    90.93424
## 1170                    82.60900
## 1171                    71.83842
## 1172                    83.91271
## 1173                    75.95272
## 1174                    75.96734
## 1175                    70.96463
## 1176                    86.71127
## 1177                    83.49283
## 1178                    73.95786
## 1179                    83.62713
## 1180                    86.25936
## 1181                    80.13238
## 1182                    91.89137
## 1183                    77.82812
## 1184                    91.11693
## 1185                    84.02232
## 1186                    82.21387
## 1187                    84.13363
## 1188                    67.76529
## 1189                    75.24102
## 1190                    87.52787
## 1191                    71.87848
## 1192                    98.57386
## 1193                    84.52714
## 1194                    89.86159
## 1195                    86.41217
## 1196                    62.77960
## 1197                    93.40529
## 1198                    55.91086
```

```
## 1199                    56.79606
## 1200                    90.46843
## 1201                    72.86501
## 1202                    65.75041
## 1203                    86.04893
## 1204                    75.16521
## 1205                    73.88338
## 1206                    69.39560
## 1207                    73.94204
## 1208                    78.96121
## 1209                    84.99068
## 1210                    52.87414
## 1211                   -14.86831
## 1212                    76.25789
## 1213                    75.50743
## 1214                    67.16884
## 1215                    66.90743
## 1216                    78.73679
## 1217                    70.60007
## 1218                    71.04719
## 1219                    76.07489
## 1220                    86.40975
## 1221                    90.47891
## 1222                    84.40704
## 1223                    74.31993
## 1224                    89.72958
## 1225                    87.90126
## 1226                    79.67031
## 1227                    77.31396
## 1228                    89.91889
## 1229                    79.76720
## 1230                    79.69541
## 1231                    82.23650
## 1232                    84.97907
## 1233                    76.19720
## 1234                    79.83030
## 1235                    86.14125
## 1236                    74.95494
## 1237                    75.20767
## 1238                    85.68961
## 1239                    70.88459
## 1240                    73.86408
## 1241                    74.72470
## 1242                    81.71808
## 1243                    75.74775
## 1244                    76.28133
## 1245                    76.56413
## 1246                    77.13512
## 1247                    78.65494
## 1248                    63.92884
## 1249                    60.69191
## 1250                    87.35683
## 1251                    76.72049
## 1252                    59.58159
```

```
## 1253                    78.27593
## 1254                    79.31386
## 1255                    90.08760
## 1256                    80.81078
## 1257                    75.72528
## 1258                    78.00031
## 1259                    75.66633
## 1260                    71.63302
## 1261                    76.17954
## 1262                    78.85612
## 1263                    90.41550
## 1264                    80.06320
## 1265                    75.46127
## 1266                    77.64302
## 1267                    90.60779
## 1268                    93.69729
## 1269                    96.06713
## 1270                    91.70541
## 1271                    86.69148
## 1272                    82.31295
## 1273                    95.63714
## 1274                    90.44655
## 1275                    91.68412
## 1276                    95.33996
## 1277                    84.19341
## 1278                    89.10651
## 1279                    87.28543
## 1280                    91.92978
## 1281                    82.29826
## 1282                    76.67050
## 1283                    80.45691
## 1284                    76.69684
## 1285                    79.96420
## 1286                    74.58338
## 1287                    81.82592
## 1288                    71.73087
## 1289                    75.14084
## 1290                    76.98514
## 1291                    76.80357
## 1292                    77.25227
## 1293                    78.74897
## 1294                    76.08229
## 1295                    71.49812
## 1296                    75.21147
## 1297                    72.09172
## 1298                    71.92088
## 1299                    74.29483
## 1300                    77.71560
## 1301                    84.87340
## 1302                    83.59490
## 1303                    84.39328
## 1304                    79.46376
## 1305                    75.74019
## 1306                    70.60465
```

```
## 1307                    83.90743
## 1308                    83.81032
## 1309                    82.80087
## 1310                    71.25864
## 1311                    84.54443
## 1312                    82.65051
## 1313                    83.52874
## 1314                    82.23817
## 1315                    86.88477
## 1316                    74.56154
## 1317                    67.27121
## 1318                    79.27253
## 1319                    78.10627
## 1320                    81.04236
## 1321                    87.48365
## 1322                    86.72669
## 1323                    87.24697
## 1324                    80.38180
## 1325                    90.09772
## 1326                    86.48324
## 1327                    82.40340
## 1328                    87.98448
## 1329                    85.22402
## 1330                    93.55616
## 1331                    81.51782
## 1332                    85.66684
## 1333                    81.10750
## 1334                    86.75449
## 1335                    84.73531
## 1336                    90.88728
## 1337                    85.83583
## 1338                    78.76364
## 1339                    93.21848
## 1340                    72.08865
## 1341                    65.31421
## 1342                    68.13908
## 1343                    88.18541
## 1344                    75.38651
## 1345                    39.30175
## 1346                    72.33055
## 1347                    66.39322
## 1348                    51.92644
## 1349                    77.57300
## 1350                    55.51055
## 1351                    92.51743
## 1352                    70.66654
## 1353                    54.83401
## 1354                    54.62291
## 1355                    63.54168
## 1356                    63.12299
## 1357                    62.31852
## 1358                    73.09667
## 1359                    83.48429
## 1360                    70.33134
```

```
## 1361                    74.35867
## 1362                    71.37928
## 1363                    76.58367
## 1364                    80.17380
## 1365                    74.70983
## 1366                    74.71973
## 1367                    75.44644
## 1368                    80.67446
## 1369                    72.90782
## 1370                    72.68845
## 1371                    78.14078
## 1372                    85.63097
## 1373                    85.89099
## 1374                    91.24140
## 1375                    86.85887
## 1376                    80.84092
## 1377                    86.25345
## 1378                    81.56670
## 1379                    72.18818
## 1380                    75.01936
## 1381                    70.71780
## 1382                    79.16251
## 1383                    76.89407
## 1384                    80.09622
## 1385                    78.20693
## 1386                    96.48085
## 1387                    88.02606
## 1388                    76.97589
## 1389                    78.33184
## 1390                    69.00315
## 1391                    75.44951
## 1392                    83.59121
## 1393                    91.03650
## 1394                    96.85602
## 1395                    43.67512
## 1396                    62.04538
## 1397                    55.07884
## 1398                    79.74197
## 1399                    69.45314
## 1400                    76.53649
## 1401                    79.68342
## 1402                    88.81726
## 1403                    64.54058
## 1404                    60.94187
## 1405                    92.82356
## 1406                    89.13504
## 1407                    68.33606
## 1408                    70.52905
## 1409                    77.25037
## 1410                    74.18765
## 1411                    77.59405
## 1412                    71.52221
## 1413                    71.98010
## 1414                    78.57186
```

```
## 1415  83.22347
## 1416  93.39679
## 1417  91.50523
## 1418  97.07548
## 1419  93.35404
## 1420  85.68483
## 1421  91.98329
## 1422 103.56845
## 1423  92.43988
## 1424  93.40845
## 1425 102.19696
## 1426 101.20930
## 1427  93.70389
## 1428  93.42425
## 1429  99.68818
## 1430  84.70107
## 1431  83.68351
## 1432  82.49972
## 1433  80.46820
## 1434  81.63863
## 1435  78.82205
## 1436  76.33139
## 1437  89.16750
## 1438  84.71805
## 1439  95.45757
## 1440  83.08476
## 1441  80.02298
## 1442  85.73764
## 1443  81.73079
## 1444  85.90304
## 1445  83.90321
## 1446  80.43332
## 1447  77.81319
## 1448  81.28784
## 1449  87.00260
## 1450  86.27996
## 1451  80.49646
## 1452  82.64941
## 1453  68.40495
## 1454  73.97850
## 1455  65.11257
## 1456  67.00220
## 1457  72.77595
## 1458  82.64995
## 1459  80.54151
## 1460  71.47096
## 1461  76.16283
## 1462  79.38334
## 1463  82.82390
## 1464  90.79761
## 1465  92.57407
## 1466  89.66948
## 1467  90.42414
## 1468  81.43143
```

```
## 1469                    84.69947
## 1470                    86.15017
## 1471                    82.11231
## 1472                    89.47976
## 1473                    89.87134
## 1474                    88.14762
## 1475                    84.24673
## 1476                    82.87699
## 1477                    76.08768
## 1478                    86.77008
## 1479                    95.37445
## 1480                    89.66387
## 1481                    95.17080
## 1482                    92.91091
## 1483                    99.90620
## 1484                    96.07490
## 1485                    92.02609
## 1486                    89.56345
## 1487                    90.11809
## 1488                    92.06957
## 1489                    92.75542
## 1490                    93.40726
## 1491                    95.36751
## 1492                    83.14946
## 1493                    90.32770
## 1494                    79.10407
## 1495                    75.15981
## 1496                    85.45343
## 1497                    84.91113
## 1498                    83.94183
## 1499                    70.26753
## 1500                    94.48332
## 1501                   103.28924
## 1502                   121.45508
## 1503                    91.39908
## 1504                    60.27100
## 1505                    62.30368
## 1506                    73.60100
## 1507                    59.16067
## 1508                    66.35067
## 1509                    63.55810
## 1510                    74.54077
## 1511                    78.49194
## 1512                    76.06463
## 1513                    77.49844
## 1514                    90.41574
## 1515                    88.73248
## 1516                    97.92793
## 1517                    95.88305
## 1518                    88.32950
## 1519                    99.04154
## 1520                    88.86333
## 1521                    84.02294
## 1522                    81.27678
```

```
## 1523                    77.26068
## 1524                    80.14478
## 1525                    83.86352
## 1526                    74.56434
## 1527                    77.98810
## 1528                    72.07848
## 1529                    67.24270
## 1530                    75.02149
## 1531                    68.31295
## 1532                    71.89396
## 1533                    82.00599
## 1534                    78.14076
## 1535                    77.04762
## 1536                    76.30899
## 1537                    79.61284
## 1538                    66.90099
## 1539                    74.96476
## 1540                    72.20948
## 1541                    74.32697
## 1542                    73.55172
## 1543                    72.41824
## 1544                    87.70961
## 1545                    76.75329
## 1546                    70.38464
## 1547                    75.46643
## 1548                    68.11055
## 1549                    71.35640
## 1550                    74.23048
## 1551                    80.20165
## 1552                    80.11817
## 1553                    73.28357
## 1554                    83.18685
## 1555                    79.93114
## 1556                    87.51005
## 1557                    76.08595
## 1558                    71.78502
## 1559                    72.31305
## 1560                    90.87099
## 1561                    85.20843
## 1562                    81.57663
## 1563                    85.98846
## 1564                    82.60436
## 1565                    87.31671
## 1566                    85.08796
## 1567                    83.97403
## 1568                    88.24729
## 1569                    83.19463
## 1570                    87.12874
## 1571                    82.52175
## 1572                    82.86340
## 1573                    84.54527
## 1574                    84.15981
## 1575                    80.79247
## 1576                    78.03604
```

```
## 1577                    85.77454
## 1578                    85.08992
## 1579                    82.94755
## 1580                    78.74724
## 1581                    81.91254
## 1582                    80.73980
## 1583                    83.25665
## 1584                    44.89652
## 1585                    80.26286
## 1586                    80.36965
## 1587                    89.09074
## 1588                    77.93892
## 1589                    83.00193
## 1590                    95.18983
## 1591                    95.13485
## 1592                    51.36249
## 1593                    79.80377
## 1594                    63.95358
## 1595                    71.26201
## 1596                    64.04355
## 1597                    90.35811
## 1598                    55.44163
## 1599                    77.13828
## 1600                    89.22636
## 1601                    72.26745
## 1602                    80.79500
## 1603                   111.81219
## 1604                   131.29061
## 1605                    95.94909
## 1606                    93.89723
## 1607                    80.09299
## 1608                    90.77829
## 1609                    88.18465
## 1610                    83.06276
## 1611                    64.85960
## 1612                    72.44562
## 1613                    75.09178
## 1614                    75.61855
## 1615                    87.51030
## 1616                    83.99103
## 1617                    76.53121
## 1618                    78.84924
## 1619                    78.87754
## 1620                    83.79980
## 1621                    76.38788
## 1622                    72.62525
## 1623                    69.42170
## 1624                    72.91944
## 1625                    66.78497
## 1626                    69.41925
## 1627                    71.43095
## 1628                    80.06691
## 1629                    77.63455
## 1630                    78.05276
```

```
## 1631                    84.59928
## 1632                    76.30619
## 1633                    81.14714
## 1634                    74.17113
## 1635                    98.28792
## 1636                    89.31541
## 1637                    76.32556
## 1638                    88.61179
## 1639                    71.71472
## 1640                    79.29136
## 1641                    78.90375
## 1642                    66.33320
## 1643                    72.27326
## 1644                    67.63748
## 1645                    69.61910
## 1646                    61.72377
## 1647                    72.75563
## 1648                    74.32216
## 1649                    75.69109
## 1650                    77.80176
## 1651                    78.91773
## 1652                    83.45507
## 1653                    83.02456
## 1654                    80.42750
## 1655                    80.44534
## 1656                    82.08588
## 1657                    81.62892
## 1658                    70.80280
## 1659                    66.90872
## 1660                    67.91063
## 1661                    79.47662
## 1662                    73.89015
## 1663                    73.37765
## 1664                    68.74694
## 1665                    65.31852
## 1666                    70.22005
## 1667                    69.71952
## 1668                    70.00281
## 1669                    70.87700
## 1670                    77.73193
## 1671                    82.37303
## 1672                    86.69303
## 1673                    92.54408
## 1674                    84.69542
## 1675                    90.04192
## 1676                    87.44192
## 1677                    91.57678
## 1678                    76.96811
## 1679                    79.81322
## 1680                    83.89263
## 1681                    70.80974
## 1682                    78.23753
## 1683                    80.94462
## 1684                    76.91876
```

```
## 1685                    79.40930
## 1686                    89.91181
## 1687                    78.26627
## 1688                    80.49813
## 1689                    83.10824
## 1690                    90.15640
## 1691                    80.96437
## 1692                    85.02245
## 1693                    83.97359
## 1694                    82.13321
## 1695                    89.93923
## 1696                    92.08662
## 1697                    87.54202
## 1698                    58.62330
## 1699                    82.99815
## 1700                    78.75513
## 1701                    83.27171
## 1702                    80.35481
## 1703                    62.15965
## 1704                    85.67398
## 1705                    64.52600
## 1706                    70.19915
## 1707                    56.72306
## 1708                    47.10289
## 1709                    71.50721
## 1710                   109.16681
## 1711                    89.19625
## 1712                    98.45044
## 1713                    85.02313
## 1714                    72.13163
## 1715                    78.74191
## 1716                    89.75651
## 1717                    94.70675
## 1718                    99.68939
## 1719                    82.20003
## 1720                    94.01725
## 1721                    94.77412
## 1722                   102.23011
## 1723                   103.98130
## 1724                   103.97227
## 1725                    79.13519
## 1726                    95.76770
## 1727                    85.18453
## 1728                    73.49835
## 1729                    77.83862
## 1730                    74.56013
## 1731                    68.28687
## 1732                    81.68564
## 1733                    84.11331
## 1734                    85.57622
## 1735                    91.37602
## 1736                   104.24516
## 1737                    95.52619
## 1738                    91.22868
```

```
## 1739                    94.78600
## 1740                    83.55382
## 1741                    93.48625
## 1742                    96.75604
## 1743                   100.40223
## 1744                    90.73385
## 1745                    74.76422
## 1746                    85.65299
## 1747                    85.31663
## 1748                    90.86962
## 1749                    90.58249
## 1750                    89.41231
## 1751                    86.53021
## 1752                    84.22554
## 1753                    79.53790
## 1754                    81.29795
## 1755                    80.03760
## 1756                    73.64911
## 1757                    77.50623
## 1758                    85.36075
## 1759                    82.71380
## 1760                    77.19572
## 1761                    84.79545
## 1762                    85.59784
## 1763                    81.54402
## 1764                    82.54350
## 1765                    76.13430
## 1766                    76.28119
## 1767                    72.34871
## 1768                    73.94165
## 1769                    76.00197
## 1770                    74.62276
## 1771                    74.06178
## 1772                    74.96876
## 1773                    75.72543
## 1774                    67.22022
## 1775                    71.07864
## 1776                    72.21657
## 1777                    77.73381
## 1778                    79.94965
## 1779                    76.24118
## 1780                    80.02158
## 1781                    84.87503
## 1782                    77.22274
## 1783                    76.64299
## 1784                    81.81345
## 1785                    78.53895
## 1786                    83.84706
## 1787                    84.17435
## 1788                    88.62038
## 1789                    87.14151
## 1790                    80.59349
## 1791                    86.70901
## 1792                    82.57953
```

```
## 1793                    75.99284
## 1794                    79.51426
## 1795                    85.35239
## 1796                    77.80860
## 1797                    84.08457
## 1798                    87.79439
## 1799                    83.66026
## 1800                    82.77320
## 1801                    69.85413
## 1802                    81.41726
## 1803                    79.75396
## 1804                    76.23201
## 1805                    69.58414
## 1806                    68.00580
## 1807                    83.14810
## 1808                    71.83989
## 1809                    73.68056
## 1810                   103.55689
## 1811                    98.56856
## 1812                    80.71717
## 1813                   104.74975
## 1814                    76.82983
## 1815                   106.74587
## 1816                    85.17167
## 1817                    79.67060
## 1818                    81.68417
## 1819                    94.25169
## 1820                    91.36705
## 1821                    72.34573
## 1822                    61.49556
## 1823                    82.18661
## 1824                    69.23555
## 1825                    18.80203
## 1826                    62.18674
## 1827                    75.93525
## 1828                    62.03754
## 1829                    93.21808
## 1830                    79.77183
## 1831                    60.32556
## 1832                    71.71609
## 1833                    65.12902
## 1834                    65.53250
## 1835                    67.78414
## 1836                    69.94814
## 1837                    67.01801
## 1838                    73.00962
## 1839                    77.85395
## 1840                    74.24686
## 1841                    87.66879
## 1842                    72.43845
## 1843                    81.28324
## 1844                    83.62708
## 1845                    84.18550
## 1846                    76.45624
```

```
## 1847                    87.70459
## 1848                    76.98064
## 1849                    77.46274
## 1850                    82.47850
## 1851                    77.16459
## 1852                    79.46878
## 1853                    75.51107
## 1854                    81.71258
## 1855                    84.03095
## 1856                    87.19759
## 1857                    86.46741
## 1858                    79.04322
## 1859                    80.45795
## 1860                    80.67275
## 1861                    74.92599
## 1862                    81.86840
## 1863                    82.58899
## 1864                    86.98529
## 1865                    77.36641
## 1866                    74.80178
## 1867                    85.82635
## 1868                    72.91180
## 1869                    78.06480
## 1870                    79.96291
## 1871                    71.11005
## 1872                    83.13746
## 1873                    81.40171
## 1874                    72.26757
## 1875                    90.61998
## 1876                    77.05026
## 1877                    76.36588
## 1878                    82.04913
## 1879                    77.77679
## 1880                    80.78692
## 1881                    83.66812
## 1882                    84.24362
## 1883                    92.65908
## 1884                    94.55569
## 1885                    92.01758
## 1886                    89.81358
## 1887                    89.46349
## 1888                    89.99140
## 1889                   100.67663
## 1890                    92.45367
## 1891                    87.01803
## 1892                    84.41864
## 1893                    79.65028
## 1894                    86.82086
## 1895                    85.55919
## 1896                   108.70581
## 1897                    71.35281
## 1898                    99.26440
## 1899                    74.84199
## 1900                    94.23312
```

```
## 1901                    76.83131
## 1902                    82.83056
## 1903                    70.57475
## 1904                   102.51035
## 1905                   102.07675
## 1906                    89.96004
## 1907                   100.36135
## 1908                   101.73733
## 1909                    77.61398
## 1910                    78.26324
## 1911                    70.60512
## 1912                    64.37255
## 1913                    55.67408
## 1914                    93.81420
## 1915                    95.79974
## 1916                   117.11241
## 1917                   107.12225
## 1918                    96.03197
## 1919                    89.02749
## 1920                    87.34414
## 1921                    97.70865
## 1922                    92.72954
## 1923                    89.28686
## 1924                    79.74015
## 1925                    71.66500
## 1926                    81.37575
## 1927                    79.35560
## 1928                    79.84553
## 1929                    79.34939
## 1930                    80.31298
## 1931                    97.06207
## 1932                    96.66579
## 1933                    92.64241
## 1934                    86.89358
## 1935                    79.64500
## 1936                    88.82754
## 1937                    88.17389
## 1938                    95.50298
## 1939                   103.52014
## 1940                    92.73955
## 1941                    81.72361
## 1942                    75.96567
## 1943                    80.55155
## 1944                    91.15125
## 1945                    82.26535
## 1946                    83.12106
## 1947                    81.62433
## 1948                    80.64365
## 1949                    84.82151
## 1950                    77.83518
## 1951                    83.58269
## 1952                    77.38139
## 1953                    81.54834
## 1954                    88.87408
```

```
## 1955                    83.67142
## 1956                    89.89394
## 1957                    86.76129
## 1958                    86.86648
## 1959                    86.24192
## 1960                    78.85478
## 1961                    86.77730
## 1962                    81.78580
## 1963                    81.89456
## 1964                    76.35250
## 1965                    78.76773
## 1966                    79.96449
## 1967                    81.44875
## 1968                    80.15261
## 1969                    85.96076
## 1970                    85.08263
## 1971                    78.84692
## 1972                    82.09936
## 1973                    76.16900
## 1974                    74.11347
## 1975                    73.15330
## 1976                    75.28360
## 1977                    84.13038
## 1978                    75.71428
## 1979                    78.71630
## 1980                    87.68198
## 1981                    74.60858
## 1982                    79.77265
## 1983                    71.22685
## 1984                    78.94526
## 1985                    79.51838
## 1986                    77.52557
## 1987                    83.26182
## 1988                    82.12379
## 1989                    70.66966
## 1990                    76.39111
## 1991                    80.75423
## 1992                    78.00667
## 1993                    79.21319
## 1994                    85.70446
## 1995                    70.03010
## 1996                    86.60789
## 1997                    82.60220
## 1998                    79.34914
## 1999                    77.30414
## 2000                    83.43552
## 2001                    87.93789
## 2002                    88.28898
## 2003                    91.14773
## 2004                    85.51529
## 2005                    86.78964
## 2006                    83.95283
## 2007                    89.17882
## 2008                    77.62991
```

```
## 2009  83.86297
## 2010  71.90911
## 2011  88.12751
## 2012  89.40111
## 2013  76.00431
## 2014  44.65090
## 2015  46.54715
## 2016  72.01402
## 2017  69.31108
## 2018  89.11690
## 2019  84.66067
## 2020  95.25301
## 2021  80.80391
## 2022 112.89646
## 2023  83.81413
## 2024  86.45394
## 2025  82.21161
## 2026  87.36365
## 2027  68.75727
## 2028  90.97261
## 2029  90.23974
## 2030  77.58696
## 2031  80.29470
## 2032  63.83707
## 2033  89.95607
## 2034  96.20271
## 2035  84.88836
## 2036  67.98186
## 2037  56.19358
## 2038  77.59880
## 2039  83.51087
## 2040  70.02379
## 2041  46.75252
## 2042  48.84109
## 2043  70.56914
## 2044  78.00975
## 2045  79.51433
## 2046  79.37326
## 2047  76.97936
## 2048  75.51561
## 2049  79.16711
## 2050  67.16412
## 2051  71.49356
## 2052  71.36621
## 2053  84.91706
## 2054  91.95770
## 2055  92.40683
## 2056  85.42620
## 2057  83.84230
## 2058  88.43629
## 2059  79.00310
## 2060  92.33296
## 2061  90.42476
## 2062  92.82913
```

```
## 2063                    82.86310
## 2064                    77.73821
## 2065                    85.10907
## 2066                    88.49718
## 2067                    85.78150
## 2068                    85.23119
## 2069                    87.19529
## 2070                    83.67927
## 2071                    86.51123
## 2072                    88.66307
## 2073                    82.48833
## 2074                    82.80572
## 2075                    80.49059
## 2076                    79.93105
## 2077                    87.32657
## 2078                    85.58132
## 2079                    86.54277
## 2080                    76.20930
## 2081                    79.64144
## 2082                    79.08269
## 2083                    84.14945
## 2084                    86.03111
## 2085                    79.03993
## 2086                    81.12575
## 2087                    85.14312
## 2088                    76.40638
## 2089                    79.42712
## 2090                    77.15059
## 2091                    76.08445
## 2092                    82.99992
## 2093                    84.84673
## 2094                    77.75520
## 2095                    79.43344
## 2096                    74.33193
## 2097                    83.18418
## 2098                    74.44958
## 2099                    77.68838
## 2100                    84.93724
## 2101                    87.40402
## 2102                    77.90750
## 2103                    76.85843
## 2104                    81.93931
## 2105                    82.36836
## 2106                    75.46751
## 2107                    82.17710
## 2108                    71.89975
## 2109                    85.09388
## 2110                    85.93681
## 2111                    85.26535
## 2112                    85.48008
## 2113                    78.69280
## 2114                    94.46427
## 2115                    77.04686
## 2116                    89.85809
```

```
## 2117                    85.60405
## 2118                    83.53485
## 2119                    86.96302
## 2120                    85.28335
## 2121                    88.09081
## 2122                    83.56187
## 2123                    81.66703
## 2124                    65.89695
## 2125                    83.34374
## 2126                    80.16882
## 2127                    85.28029
## 2128                    84.29207
## 2129                    89.51183
## 2130                    84.20566
## 2131                    83.57110
## 2132                    92.41966
## 2133                    87.59234
## 2134                    80.84726
## 2135                    81.04132
## 2136                    50.06806
## 2137                    62.07784
## 2138                    79.56881
## 2139                    77.67244
## 2140                    75.80818
## 2141                    78.69893
## 2142                    73.06391
## 2143                    72.82676
## 2144                    80.52420
## 2145                    82.73718
## 2146                    84.41271
## 2147                    79.09130
## 2148                    70.81323
## 2149                    78.01184
## 2150                    63.94425
## 2151                    66.80312
## 2152                    66.48291
## 2153                    67.88544
## 2154                    61.94055
## 2155                    65.09533
## 2156                    80.28228
## 2157                    75.04591
## 2158                    65.12067
## 2159                    76.08670
## 2160                    81.62543
## 2161                    77.37865
## 2162                    77.12585
## 2163                    89.67867
## 2164                    86.59012
## 2165                    86.13036
## 2166                    79.55490
## 2167                    73.40650
## 2168                    81.35782
## 2169                    83.23258
## 2170                    84.67323
```

```
## 2171                    85.59296
## 2172                    79.71178
## 2173                    80.36302
## 2174                    76.87710
## 2175                    87.08616
## 2176                    74.59007
## 2177                    89.81382
## 2178                    80.33642
## 2179                    96.02359
## 2180                    80.99649
## 2181                    92.60314
## 2182                    96.09116
## 2183                    85.57263
## 2184                    86.80469
## 2185                    85.41643
## 2186                    84.68321
## 2187                    83.13571
## 2188                    85.69452
## 2189                    79.62012
## 2190                    86.87461
## 2191                    70.03476
## 2192                    75.02736
## 2193                    68.03381
## 2194                    70.89891
## 2195                    62.15289
## 2196                    80.61343
## 2197                    92.02857
## 2198                    90.84946
## 2199                    87.57983
## 2200                    89.27920
## 2201                    90.54054
## 2202                    85.18860
## 2203                    83.11234
## 2204                    90.21774
## 2205                    83.17718
## 2206                    93.67170
## 2207                    93.52825
## 2208                    85.33928
## 2209                    80.60595
## 2210                    78.60144
## 2211                    76.22472
## 2212                    90.46434
## 2213                    87.64183
## 2214                    86.19946
## 2215                    81.10451
## 2216                    79.15053
## 2217                    81.20221
## 2218                    86.83673
## 2219                   101.60105
## 2220                    72.20146
## 2221                    79.96315
## 2222                    83.75027
## 2223                    73.32756
## 2224                    61.27572
```

```
## 2225                    67.42931
## 2226                    75.88428
## 2227                    88.51002
## 2228                    89.67653
## 2229                    88.46107
## 2230                    74.46896
## 2231                    78.20332
## 2232                    46.63564
## 2233                    59.20192
## 2234                    65.50040
## 2235                    37.78302
## 2236                    69.68197
## 2237                    55.35096
## 2238                    65.34020
## 2239                    40.85866
## 2240                    74.82410
## 2241                    87.04544
## 2242                    64.74947
## 2243                    64.40598
## 2244                    69.42575
## 2245                    69.80910
## 2246                    79.15849
## 2247                    69.42327
## 2248                    63.74421
## 2249                    84.05834
## 2250                    73.41059
## 2251                    84.61742
## 2252                    90.25716
## 2253                    85.04350
## 2254                    89.54356
## 2255                    87.65199
## 2256                    77.90718
## 2257                    80.27853
## 2258                    84.84810
## 2259                    84.67466
## 2260                    78.94493
## 2261                    79.70711
## 2262                    87.93593
## 2263                    80.34373
## 2264                    91.78796
## 2265                    77.09548
## 2266                    81.08794
## 2267                    71.90745
## 2268                    79.29134
## 2269                    78.28669
## 2270                    73.41430
## 2271                    78.85509
## 2272                    78.55816
## 2273                    76.03423
## 2274                    70.06265
## 2275                    82.28133
## 2276                    46.35624
```

Conclusion

We examined ~2200 records of baseball teams to create a predictive model of wins. However, if this were an actual workplace project, it seems unlikely that the point would be the passive prediction of wins from sample performance statistics. Rather, the data would need to serve the question of what strategies should be employed to improve wins. Answering this question require more insight than ability to predict. Throughout this analysis we have confronted the counter-intuitive phenomenon that weaker pitching is correlated with better outcomes. ***Analysis shows that this is most likely because teams need to trade off pitching and hitting, and better hitting compensates more for poor pitching than vice versa. This is the most important finding of this examination.***

Appendix

Libraries included:

library(Hmisc) library(psych) library(tidyverse) library(skimr) library(purrr) library(tidyr) library(tidyverse) library(gridExtra) library(lubridate) library(fastDummies) library(data.table) library(mltools) library(MASS) library(car) library(patchwork) library(ggthemes) library(tinytex) library(stats) library(EHData) library(ggsci)

1. Fix Column Names colnames(dfTrain)<-gsub("TEAM_","“,colnames(dfTrain)) colnames(dfEval)<-gsub("TEAM_“,"",colnames(dfEval))

2. Summarize summary(dfTrain)

df_NoIndex <- dfTrain %>% dplyr::select(-INDEX)

a <- EHSummarize_SingleColumn_Histograms(df_NoIndex, font_size = 9) grid.arrange(grobs=a[c(1:16)], ncol=4, top = “Column Distributions”, bottom=“Fig. 1”)

a <- EHSummarize_SingleColumn_Boxplots(df_NoIndex, font_size=9) grid.arrange(grobs=a[c(1:16)], ncol=4, top = “Boxplots for Outlier Analysis”, bottom=“Fig. 2”)

a <- EHExplore_TwoContinuousColumns_Scatterplots(df_NoIndex, “TARGET_WINS”) grid.arrange(grobs=a[c(2:16)], ncol=3, top = “Scatterplots Against TARGET_WINS”, bottom=“Fig.3”)

EHExplore_Multicollinearity(dfTrain, title=“Correlations, Fig. 4”, run_all = FALSE)

3. Prepare Data dfTrain1 <- dfTrain %>% dplyr::select(-BATTING_HBP, -BASERUN_CS)

dfTrain2 <- dfTrain1 %>% mutate(PSO_Missing_Flag = ifelse(is.na(PITCHING_SO),1,0), BSO_Missing_Flag = ifelse(is.na(BATTING_SO),1,0), BRSB_Missing_Flag = ifelse(is.na(BASERUN_SB),1,0), FDP_Missing_Flag = ifelse(is.na(FIELDING_DP),1,0))

dfEval1 <- dfTrain %>% dplyr::select(-BATTING_HBP, -BASERUN_CS)

dfEval2 <- dfEval1 %>% mutate(PSO_Missing_Flag = ifelse(is.na(PITCHING_SO),1,0), BSO_Missing_Flag = ifelse(is.na(BATTING_SO),1,0), BRSB_Missing_Flag = ifelse(is.na(BASERUN_SB),1,0), FDP_Missing_Flag = ifelse(is.na(FIELDING_DP),1,0))

dfTrain22 <- dfTrain2 %>% dplyr::select(PSO_Missing_Flag, BSO_Missing_Flag, BRSB_Missing_Flag, FDP_Missing_Flag)

z1 <- EHExplore_TwoCategoricalColumns_Barcharts(dfTrain22, “BSO_Missing_Flag”) z2 <- EHExplore_TwoCategoricalColumns_Barcharts(dfTrain22, “BRSB_Missing_Flag”) z3 <- c(z1, z2)

dfTrain2 <- dfTrain2 %>% dplyr::select(-PSO_Missing_Flag)

dfEval2 <- dfEval2 %>% dplyr::select(-PSO_Missing_Flag)

grid.arrange(grobs=z3[c(1,3:4,8)], ncol=2, top="Overlap of NA's Among Columns", bottom = "Fig. 5")

a <- EHExplore_Interactions_Scatterplots(dfTrain2, "TARGET_WINS", "BSO_Missing_Flag")

grid.arrange(a[[6]], a[[10]], a[[11]], a[[12]], a[[14]], ncol=2, top = "Selected Interactions with Missing Batting_SO", bottom = "Fig. 6")

dfTrain2 <- dfTrain2 %>% mutate(PITCHING_SO = ifelse(PITCHING_SO==0, NA, PITCHING_SO)) %>% mutate(BATTING_SO = ifelse(BATTING_SO==0, NA, BATTING_SO)) %>% mutate(BATTING_HR = ifelse(BATTING_HR==0, NA, BATTING_HR))

dfTrain2 <- EHPrepare_MissingValues_Imputation(dfTrain2, "TARGET_WINS")

dfEval2 <- dfEval2 %>% mutate(PITCHING_SO = ifelse(PITCHING_SO==0, NA, PITCHING_SO)) %>% mutate(BATTING_SO = ifelse(BATTING_SO==0, NA, BATTING_SO)) %>% mutate(BATTING_HR = ifelse(BATTING_HR==0, NA, BATTING_HR))

dfEval2 <- EHPrepare_MissingValues_Imputation(dfEval2, "TARGET_WINS")

dfTrain_NoTransformations <- dfTrain2

3. Data Modeling 1. We create a flag for hits under 1500

dfPH <- dfTrain2 %>% dplyr::select(TARGET_WINS, PITCHING_H)

dfPH2 <- dfPH %>% dplyr::filter(PITCHING_H <= 3000)

x1 <- EHExplore_TwoContinuousColumns_Scatterplots(dfPH, "TARGET_WINS") x2 <- EHExplore_TwoContinuousColumns_Scatterplots(dfPH2, "TARGET_WINS")

grid.arrange(x1[[2]], x2[[2]], ncol=2, top="Pitching_H Against Wins, All Records (left) and Hits Below 3000 (right)", bottom="Fig.7")

dfTrain2 <- dfTrain2 %>% mutate(Pitch_h_Under1500 = ifelse(PITCHING_H<=1500, 1, 0))

dfEval2 <- dfEval2 %>% mutate(Pitch_h_Under1500 = ifelse(PITCHING_H<=1500, 1, 0))

EHExplore_OneContinuousAndOneCategoricalColumn_Boxplots(dfTrain2, "Pitch_h_Under1500")

dfTrain2 <- dfTrain2 %>% mutate(DP_times_PH = FIELDING_DP*PITCHING_H)

dfEval2 <- dfEval2 %>% mutate(DP_times_PH = FIELDING_DP*PITCHING_H)

a <- summary(lm(TARGET_WINS ~ FIELDING_DP, dfTrain2))$adj.r.squared b <- summary(lm($TARGET_WINS FIELDING_DP + PITCHING_H$, $dfTrain2$))$adj.r.squared c <- summary(lm(TARGET_WINS ~ FIELDING_DP + PITCHING_H + DP_times_PH, dfTrain2))$adj.r.squared

a <- ggplot(dfTrain2, aes(BATTING_HR, PITCHING_HR)) + EHTheme() + geom_point(fill="navy", color="white") + geom_smooth(method = "loess", color="red", fill="lightcoral") + ggtitle("Batting_HR vs Ptching_HR")

grid.arrange(a, bottom="Fig. 8")

dfTrain2 <- dfTrain2 %>% dplyr::select(-PITCHING_HR)

dfEval2 <- dfEval2 %>% dplyr::select(-PITCHING_HR)

dfPH3 <- dfTrain2 %>% dplyr::select(BATTING_HR)

dfPH4 <- dfPH3 %>% dplyr::filter(BATTING_HR <= 100)

x1 <- EHSummarize_SingleColumn_Histograms(dfPH3) x2 <- EHSummarize_SingleColumn_Histograms(dfPH4, hist_nbins = 100)

grid.arrange(x1[[1]], x2[[1]], ncol=2, top="Distribution of Batting HR, All Records (left) and HR below 80 (right)", bottom="Fig.9")

dfTrain2 <- dfTrain2 %>% mutate(Bat_hr_Under60 = ifelse(BATTING_HR<=80, 1, 0))

dfEval2 <- dfEval2 %>% mutate(Bat_hr_Under60 = ifelse(BATTING_HR<=80, 1, 0))

The numbers below represents the r-squareds of a simple linear regression of the column in question on t

[1] 0.02231354

[1] 0.03503598

__5. We transform the Fielding_err variable.__

While the distributions of a number of columns suggest possible tranformations, we focus here on fieldi

The numbers below represents the r-squareds of a simple linear regression of the column in question on t

[1] 0.03072081

[1] 0.04825783

__6. We create interaction terms between the SO missing cohort and the columns identified above in the i

The new fields are: Interaction_pbb_With_SO_Missing, Interaction_err_With_SO_Missing, Interaction_bh_Wi

__7. For the sake of legibility, we do not create log terms for the many skewed distributions.__

We would normally sacrifice some legibility for improved predictability by trying some log transformati

4. Model Selection

Here we build and test our models to gain insight into the dataset and ultimately predict outcomes.

According to the assignmrnt: "Since we have not yet covered automated variable
selection methods, you should select the variables manually (unless you previously learned Forward or St

The stepAIC() function performs backward model selection by starting from a "maximal" model, which is th

Because we are interested in interpretation as well as prediciton, we will modify the StepAIC model if w

__*a. Regression 1: Baseline (No transformations except flags for missing data)*__

Call:

lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +

BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +

PITCHING_SO + FIELDING_E + FIELDING_DP + BSO_Missing_Flag +

BRSB_Missing_Flag + FDP_Missing_Flag, data = df)

Residuals:

Min 1Q Median 3Q Max

-60.531 -8.063 0.330 8.075 49.266

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13.7948052 5.0143117 2.751 0.00599 **

BATTING_H 0.0521109 0.0033520 15.546 < 2e-16 ***

BATTING_2B -0.0401259 0.0086621 -4.632 3.82e-06 ***

BATTING_3B 0.0537762 0.0158617 3.390 0.00071 ***

BATTING_HR 0.0595856 0.0089648 6.647 3.75e-11 ***

BATTING_BB 0.0260490 0.0032618 7.986 2.20e-15 ***

BATTING_SO -0.0066440 0.0022278 -2.982 0.00289 **

BASERUN_SB 0.0477764 0.0046194 10.343 < 2e-16 ***

PITCHING_H 0.0018926 0.0003398 5.569 2.86e-08 ***

PITCHING_SO -0.0013966 0.0006654 -2.099 0.03593 *

FIELDING_E -0.0560670 0.0033748 -16.613 < 2e-16 ***

FIELDING_DP -0.0969459 0.0134629 -7.201 8.10e-13 ***

BSO_Missing_Flag 8.3474206 1.4721894 5.670 1.61e-08 ***

BRSB_Missing_Flag 34.1064444 1.8484454 18.451 < 2e-16 ***

FDP_Missing_Flag 4.2303099 1.4669785 2.884 0.00397 **

NULL

The adjusted r squared is .403. As we expected, many of the signs are in the "wrong" direction, especia

__*b. Regression 2: Include All transformations*__

Call:

lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +

BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +

FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +

FDP_Missing_Flag + Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq +

Interaction_pbb_With_SO_Missing + Interaction_err_With_SO_Missing +

Interaction_bhr_With_SO_Missing + Interaction_bbb_With_SO_Missing +

Interaction_sb_With_SO_Missing, data = df)

Residuals:

Min 1Q Median 3Q Max

-47.202 -7.806 0.193 7.821 48.504

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.472e+01 6.702e+00 3.688 0.000231 *

BATTING_H 5.622e-02 3.302e-03 17.023 < 2e-16 ***

BATTING_2B -4.125e-02 8.586e-03 -4.805 1.65e-06 ***

BATTING_3B 6.743e-02 1.610e-02 4.188 2.93e-05 ***

BATTING_HR 5.825e-02 8.978e-03 6.488 1.06e-10 ***

BATTING_BB 2.593e-02 3.247e-03 7.984 2.23e-15 ***

BATTING_SO -1.223e-02 2.218e-03 -5.512 3.95e-08 ***

BASERUN_SB 5.238e-02 4.795e-03 10.923 < 2e-16 ***

PITCHING_H -4.629e-03 2.995e-03 -1.546 0.122287

FIELDING_E -8.282e-02 7.453e-03 -11.112 < 2e-16 ***

NULL

__We note that the StepAIC process included a number of variables even though they were not significant

The second model has an adjusted r squared of .4305. This is not much better (although an ANOVA shows

We examine the residual plots in the model selection phase.

Analysis of Variance Table

Model 1: TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B + BATTING_HR +

BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H + PITCHING_SO +

FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +

FDP_Missing_Flag

Model 2: TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B + BATTING_HR +

BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H + FIELDING_E +

FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag + FDP_Missing_Flag +

Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq + Interaction_pbb_With_SO_Missing +

Interaction_err_With_SO_Missing + Interaction_bhr_With_SO_Missing +

Interaction_bbb_With_SO_Missing + Interaction_sb_With_SO_Missing

Res.Df RSS Df Sum of Sq F Pr(>F)

1 2261 334871

2 2254 318536 7 16335 16.513 < 2.2e-16 *

—

Signif. codes: 0 ‘’ *0.001* ’’ *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

__*c. Regression 3: Aggregated Power Stats by Hitting and Pitching*__

There are many more transformations possible, but we are interested here in trying a different direction

Throughout the analysis we have been struggling with a multicollinearity issue which we might characteri

__* Teams have limited budgets. Therefore, those with good batting may have weak pitching and vice-versa

We begin by creating simple Power Hitting and Pitching Weakness scores for each team. We do this by appl

The number below represents the correlation between Batting Power and Pitching Weakness. We can see the

[1] 0.7257795

These boxplots show the relationships in each power/weakness category to overall wins. We can see the p

we run regressions

Call:

lm(formula = TARGET_WINS ~ Total_Power, data = dfCat)

Residuals:

Min 1Q Median 3Q Max

-4.6105 -0.6185 0.0210 0.6501 4.1398

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -4.650e-16 2.055e-02 0.000 1

Total_Power 2.686e-01 2.775e-02 9.679 <2e-16 *

—

Signif. codes: 0 ‘’ *0.001* ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.9802 on 2274 degrees of freedom

Multiple R-squared: 0.03956, Adjusted R-squared: 0.03914

F-statistic: 93.68 on 1 and 2274 DF, p-value: < 2.2e-16

Call:

lm(formula = TARGET_WINS ~ Hitting_Power, data = dfCat)

Residuals:

Min 1Q Median 3Q Max

-4.3660 -0.5613 0.0128 0.5814 4.1365

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.868e-16 1.924e-02 0.00 1

Hitting_Power 3.971e-01 1.925e-02 20.63 <2e-16 *

—

Signif. codes: 0 ‘’ *0.001* ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.918 on 2274 degrees of freedom

Multiple R-squared: 0.1577, Adjusted R-squared: 0.1573

F-statistic: 425.6 on 1 and 2274 DF, p-value: < 2.2e-16

Call:

lm(formula = TARGET_WINS ~ Pitching_Weakness, data = dfCat)

Residuals:

Min 1Q Median 3Q Max

-5.1310 -0.5709 0.0746 0.6566 4.1376

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.292e-16 2.030e-02 0.0 1

Pitching_Weakness 2.498e-01 2.031e-02 12.3 <2e-16 *

—

Signif. codes: 0 ‘’ *0.001* ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.9685 on 2274 degrees of freedom

Multiple R-squared: 0.06238, Adjusted R-squared: 0.06196

F-statistic: 151.3 on 1 and 2274 DF, p-value: < 2.2e-16

Call:

lm(formula = TARGET_WINS ~ Hitting_Power + Pitching_Weakness,

data = dfCat)

Residuals:

Min 1Q Median 3Q Max

-4.2520 -0.5605 0.0277 0.5817 4.1367

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -9.018e-17 1.921e-02 0.000 1.00000

Hitting_Power 4.560e-01 2.793e-02 16.325 < 2e-16 *

**Pitching_Weakness -8.119e-02 2.793e-02 -2.907 0.00369 **

—

Signif. codes: 0 ‘’ *0.001* ’’ *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.9165 on 2273 degrees of freedom

Multiple R-squared: 0.1608, Adjusted R-squared: 0.16

F-statistic: 217.7 on 2 and 2273 DF, p-value: < 2.2e-16

The model shows that in the balance between hitting and pitching, __*teams should emphasize good hittin

Select a Model and Make Predicitons

Now we make predictions. The second model has the highest R squared and reliable interpretability so w

We examine the new model’s output:

Call:

lm(formula = TARGET_WINS ~ BATTING_H + BATTING_2B + BATTING_3B +

BATTING_HR + BATTING_BB + BATTING_SO + BASERUN_SB + PITCHING_H +

FIELDING_E + FIELDING_DP + BSO_Missing_Flag + BRSB_Missing_Flag +

FDP_Missing_Flag + Pitch_h_Under1500 + DP_times_PH + Fielding_Errors_sq +

Interaction_pbb_With_SO_Missing + Interaction_err_With_SO_Missing +

Interaction_bhr_With_SO_Missing + Interaction_bbb_With_SO_Missing +

Interaction_sb_With_SO_Missing, data = df)

Residuals:

Min 1Q Median 3Q Max

-46.429 -7.819 0.225 7.856 48.134

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.374e+01 6.694e+00 3.546 0.000399 *

BATTING_H 5.745e-02 3.312e-03 17.345 < 2e-16 ***

BATTING_2B -4.386e-02 8.587e-03 -5.107 3.54e-07 ***

BATTING_3B 7.151e-02 1.610e-02 4.442 9.33e-06 ***

BATTING_HR 5.887e-02 8.958e-03 6.571 6.17e-11 ***

BATTING_BB 2.550e-02 3.241e-03 7.870 5.46e-15 ***

BATTING_SO -1.222e-02 2.212e-03 -5.526 3.65e-08 ***

BASERUN_SB 5.202e-02 4.784e-03 10.872 < 2e-16 ***

PITCHING_H -4.611e-03 2.996e-03 -1.539 0.123868

FIELDING_E -8.444e-02 7.464e-03 -11.313 < 2e-16 *****

NULL

There is only a slight improvement with the elimination of influential points. We note that the adjuste

There are new influential points after eliminating the others, but we accept them without any better re

Predictions

predict(m, newdata = dfEval2)

1 65.55191

2 73.00183

3 74.25101

4 69.95245

5 66.80497

6 70.24417

7 66.29278

8 71.06411

9 74.46557

10 66.37215

11 69.46976

12 77.74397

13 76.93263

14 70.89365

15 83.13329

16 82.84444

17 88.78233

18 81.68573

19 75.63521

20 86.88036

21 74.96393

22 76.82949