DATA621 - Moneyball

Daniel Hong, Mauricio Alarcon, Maxwell Wagner

September 11, 2016

# NOTES

## Missing Data:

* TEAM\_BATTING\_SO, TEAM\_BASERUN\_SB, TEAM\_PITCHING\_SO and TEAM\_FIELDING\_DP are 90% or so complete, can we fill missing values with mean/median
* TEAM\_BASERUN\_CS, TEAM\_BATTING\_HBP are on the very incomplete side (66%, 8% respectively). Perhaps we can conver to Percent of Games Caught Stealing?, etc?
* I know nothing about baseball, can we tell the Number of Games Played?

## Value Distribution:

* A lot of variables have a long tail. In particular, TEAM\_PITCHING\_H, TEAM\_BATTING\_SO, TEAM\_PITCHING\_BB have a VERY long tail, any ideas about trimming this values? (capping/ transformation)?

## Correlations

The collosal scatterplot <https://github.com/rmalarc/DATA621/blob/master/hw01/Rplot01.png> shows some correlations, particularly TEAM\_BATTING\_H + TEAM\_BATTING\_2B, TEAM\_BATTING\_3B ~ target. Any ideas?

## 1. DATA EXPLORATION (25 Points)

Describe the size and the variables in the moneyball training data set. Consider that too much detail will cause a manager to lose interest while too little detail will make the manager consider that you aren't doing your job. Some suggestions are given below. Please do NOT treat this as a check list of things to do to complete the assignment. You should have your own thoughts on what to tell the boss. These are just ideas.

1. Mean / Standard Deviation / Median

require("plyr")

## Loading required package: plyr

require("knitr")

## Loading required package: knitr

require("psych")

## Loading required package: psych

# Let's load the data  
  
training <- read.csv(url('https://raw.githubusercontent.com/rmalarc/DATA621/master/hw01/moneyball-training-data.csv'))  
metadata <- read.csv(url('https://raw.githubusercontent.com/rmalarc/DATA621/master/hw01/moneyball-metadata.csv'))  
evaluation <- read.csv(url('https://raw.githubusercontent.com/rmalarc/DATA621/master/hw01/moneyball-evaluation-data.csv'))  
  
kable(metadata)

|  |  |  |
| --- | --- | --- |
| Variable | Definiton | Theoretical.Effect |
| INDEX | Identification Variable (do not use) | None |
| TARGET\_WINS | Number of wins |  |
| TEAM\_BATTING\_H | Base Hits by batters (1B,2B,3B,HR) | Positive Impact on Wins |
| TEAM\_BATTING\_2B | Doubles by batters (2B) | Positive Impact on Wins |
| TEAM\_BATTING\_3B | Triples by batters (3B) | Positive Impact on Wins |
| TEAM\_BATTING\_HR | Homeruns by batters (4B) | Positive Impact on Wins |
| TEAM\_BATTING\_BB | Walks by batters | Positive Impact on Wins |
| TEAM\_BATTING\_HBP | Batters hit by pitch (get a free base) | Positive Impact on Wins |
| TEAM\_BATTING\_SO | Strikeouts by batters | Negative Impact on Wins |
| TEAM\_BASERUN\_SB | Stolen bases | Positive Impact on Wins |
| TEAM\_BASERUN\_CS | Caught stealing | Negative Impact on Wins |
| TEAM\_FIELDING\_E | Errors Negative | Impact on Wins |
| TEAM\_FIELDING\_DP | Double Plays | Positive Impact on Wins |
| TEAM\_PITCHING\_BB | Walks allowed | Negative Impact on Wins |
| TEAM\_PITCHING\_H | Hits allowed | Negative Impact on Wins |
| TEAM\_PITCHING\_HR | Homeruns allowed Negative | Impact on Wins |
| TEAM\_PITCHING\_SO | Strikeouts by pitchers | Positive Impact On Wins |

columns <- colnames(training)  
target <- "TARGET\_WINS"  
inputs <- columns[!columns %in% c(target,"INDEX")]  
  
summary <- describe(training[,c(target,inputs)])[,c("n","mean","sd","median","max","min")]  
summary$completeness <- summary$n/nrow(training)  
kable(summary)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | mean | sd | median | max | min | completeness |
| TARGET\_WINS | 2276 | 80.79086 | 15.75215 | 82.0 | 146 | 0 | 1.0000000 |
| TEAM\_BATTING\_H | 2276 | 1469.26977 | 144.59120 | 1454.0 | 2554 | 891 | 1.0000000 |
| TEAM\_BATTING\_2B | 2276 | 241.24692 | 46.80141 | 238.0 | 458 | 69 | 1.0000000 |
| TEAM\_BATTING\_3B | 2276 | 55.25000 | 27.93856 | 47.0 | 223 | 0 | 1.0000000 |
| TEAM\_BATTING\_HR | 2276 | 99.61204 | 60.54687 | 102.0 | 264 | 0 | 1.0000000 |
| TEAM\_BATTING\_BB | 2276 | 501.55888 | 122.67086 | 512.0 | 878 | 0 | 1.0000000 |
| TEAM\_BATTING\_SO | 2174 | 735.60534 | 248.52642 | 750.0 | 1399 | 0 | 0.9551845 |
| TEAM\_BASERUN\_SB | 2145 | 124.76177 | 87.79117 | 101.0 | 697 | 0 | 0.9424429 |
| TEAM\_BASERUN\_CS | 1504 | 52.80386 | 22.95634 | 49.0 | 201 | 0 | 0.6608084 |
| TEAM\_BATTING\_HBP | 191 | 59.35602 | 12.96712 | 58.0 | 95 | 29 | 0.0839192 |
| TEAM\_PITCHING\_H | 2276 | 1779.21046 | 1406.84293 | 1518.0 | 30132 | 1137 | 1.0000000 |
| TEAM\_PITCHING\_HR | 2276 | 105.69859 | 61.29875 | 107.0 | 343 | 0 | 1.0000000 |
| TEAM\_PITCHING\_BB | 2276 | 553.00791 | 166.35736 | 536.5 | 3645 | 0 | 1.0000000 |
| TEAM\_PITCHING\_SO | 2174 | 817.73045 | 553.08503 | 813.5 | 19278 | 0 | 0.9551845 |
| TEAM\_FIELDING\_E | 2276 | 246.48067 | 227.77097 | 159.0 | 1898 | 65 | 1.0000000 |
| TEAM\_FIELDING\_DP | 1990 | 146.38794 | 26.22639 | 149.0 | 228 | 52 | 0.8743409 |

1. Bar Chart or Box Plot of the data

### How are the input values distributed?, do we need to do something about them?

Here's the distribution of the values for each of the variables

require("reshape2")

## Loading required package: reshape2

require("ggplot2")

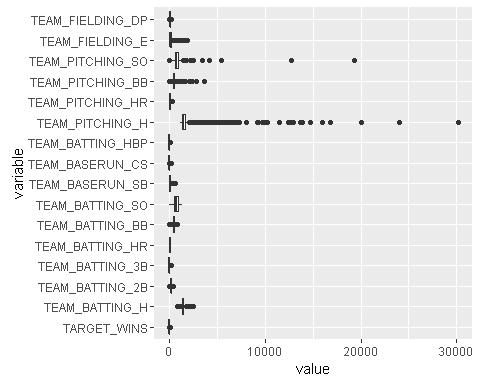
## Loading required package: ggplot2

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

# Let's melt the DF so that we can plot it more easily  
  
ggplot(melt(training, measure.vars = c(target,inputs))  
 ,aes(x=variable,y=value)  
 )+  
 geom\_boxplot() +  
 coord\_flip()

## Warning: Removed 3478 rows containing non-finite values (stat\_boxplot).

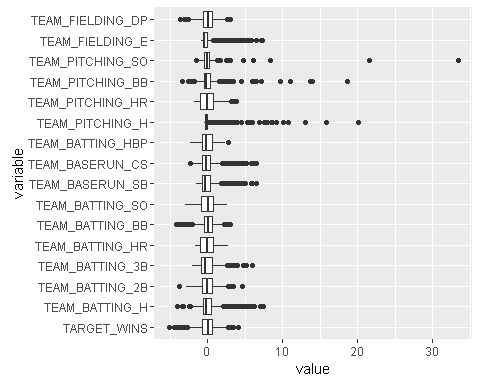


Some of these probably need to be rescaled: TEAM\_PITCHING\_H, TEAM\_PITCHING\_SO (what is this????)

Let's get a view of the normalized values:

require("reshape2")  
require("ggplot2")  
# Let's melt the DF so that we can plot it more easily  
  
ggplot(melt(data.frame(scale(training)), measure.vars = c(target,inputs)),  
 aes(x=variable,y=value)  
 )+  
 geom\_boxplot() +  
 coord\_flip()

## Warning: Removed 3478 rows containing non-finite values (stat\_boxplot).

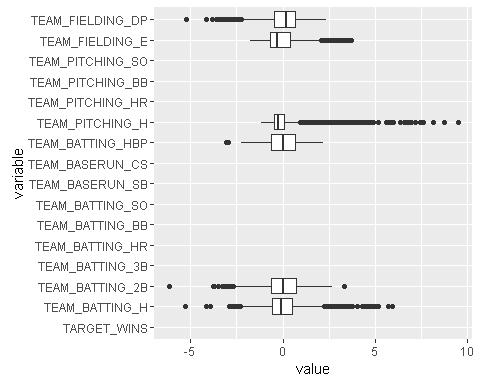


the case for data transformations is a lot more self-evident now!

Let's see what it looks like with a log transformation:

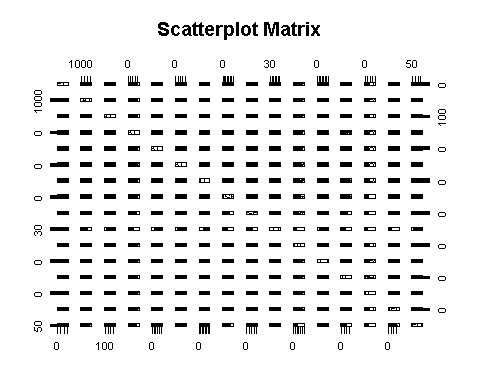
ggplot(melt(data.frame(scale(log(training))), measure.vars = c(target,inputs)),  
 aes(x=variable,y=value)  
 )+  
 geom\_boxplot() +  
 coord\_flip()

## Warning: Removed 25131 rows containing non-finite values (stat\_boxplot).



1. Is the data correlated to the target variable (or to other variables?)

pairs(~.,data=training[,c(target,inputs)],   
 main="Scatterplot Matrix")



# LEt's see the correlation matrix  
  
kable(cor(training[,c(target,inputs)], use="pairwise.complete.obs", method="kendall") )

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TARGET\_WINS | TEAM\_BATTING\_H | TEAM\_BATTING\_2B | TEAM\_BATTING\_3B | TEAM\_BATTING\_HR | TEAM\_BATTING\_BB | TEAM\_BATTING\_SO | TEAM\_BASERUN\_SB | TEAM\_BASERUN\_CS | TEAM\_BATTING\_HBP | TEAM\_PITCHING\_H | TEAM\_PITCHING\_HR | TEAM\_PITCHING\_BB | TEAM\_PITCHING\_SO | TEAM\_FIELDING\_E | TEAM\_FIELDING\_DP |
| TARGET\_WINS | 1.0000000 | 0.2578130 | 0.1624482 | 0.0811299 | 0.1102447 | 0.1661288 | -0.0515269 | 0.0775086 | -0.0052072 | 0.0195018 | 0.1514105 | 0.1155198 | 0.1505342 | -0.0633447 | -0.0884986 | -0.0352128 |
| TEAM\_BATTING\_H | 0.2578130 | 1.0000000 | 0.4277854 | 0.2253093 | 0.0524225 | 0.0543397 | -0.2734543 | 0.0008908 | 0.0133934 | -0.0199818 | 0.6341824 | 0.0783882 | 0.1068980 | -0.2548211 | 0.0671167 | 0.1231197 |
| TEAM\_BATTING\_2B | 0.1624482 | 0.4277854 | 1.0000000 | -0.1055129 | 0.3115901 | 0.1853419 | 0.1136354 | -0.1088553 | -0.0163502 | 0.0118159 | 0.2006267 | 0.3195367 | 0.1483852 | 0.0741572 | -0.2621964 | 0.1708179 |
| TEAM\_BATTING\_3B | 0.0811299 | 0.2253093 | -0.1055129 | 1.0000000 | -0.4737686 | -0.1957625 | -0.5198592 | 0.2383025 | 0.1344331 | -0.1215264 | 0.3820188 | -0.4472653 | -0.0692720 | -0.4282504 | 0.5376888 | -0.1651927 |
| TEAM\_BATTING\_HR | 0.1102447 | 0.0524225 | 0.3115901 | -0.4737686 | 1.0000000 | 0.3467213 | 0.5251584 | -0.2723937 | -0.2595813 | 0.0545571 | -0.1971559 | 0.9289328 | 0.1964719 | 0.4091055 | -0.6029884 | 0.2629838 |
| TEAM\_BATTING\_BB | 0.1661288 | 0.0543397 | 0.1853419 | -0.1957625 | 0.3467213 | 1.0000000 | 0.1790785 | -0.1024013 | -0.1137001 | 0.0059697 | -0.0938700 | 0.3208197 | 0.7639836 | 0.0785126 | -0.3026772 | 0.2194524 |
| TEAM\_BATTING\_SO | -0.0515269 | -0.2734543 | 0.1136354 | -0.5198592 | 0.5251584 | 0.1790785 | 1.0000000 | -0.0740256 | -0.1167562 | 0.1202472 | -0.4303690 | 0.4921480 | 0.0411094 | 0.8257782 | -0.5318635 | 0.0593358 |
| TEAM\_BASERUN\_SB | 0.0775086 | 0.0008908 | -0.1088553 | 0.2383025 | -0.2723937 | -0.1024013 | -0.0740256 | 1.0000000 | 0.4898029 | -0.0209281 | 0.0907725 | -0.2625464 | -0.0129995 | -0.0373803 | 0.2401202 | -0.2895276 |
| TEAM\_BASERUN\_CS | -0.0052072 | 0.0133934 | -0.0163502 | 0.1344331 | -0.2595813 | -0.1137001 | -0.1167562 | 0.4898029 | 1.0000000 | -0.0435396 | 0.0101378 | -0.2534545 | -0.0853003 | -0.1070170 | 0.1437337 | -0.0939819 |
| TEAM\_BATTING\_HBP | 0.0195018 | -0.0199818 | 0.0118159 | -0.1215264 | 0.0545571 | 0.0059697 | 0.1202472 | -0.0209281 | -0.0435396 | 1.0000000 | -0.0159068 | 0.0552225 | 0.0066409 | 0.1211429 | 0.0493273 | -0.0485670 |
| TEAM\_PITCHING\_H | 0.1514105 | 0.6341824 | 0.2006267 | 0.3820188 | -0.1971559 | -0.0938700 | -0.4303690 | 0.0907725 | 0.0101378 | -0.0159068 | 1.0000000 | -0.1295899 | 0.1047002 | -0.2741411 | 0.3301306 | 0.0268468 |
| TEAM\_PITCHING\_HR | 0.1155198 | 0.0783882 | 0.3195367 | -0.4472653 | 0.9289328 | 0.3208197 | 0.4921480 | -0.2625464 | -0.2534545 | 0.0552225 | -0.1295899 | 1.0000000 | 0.2204894 | 0.4190538 | -0.5579081 | 0.2641637 |
| TEAM\_PITCHING\_BB | 0.1505342 | 0.1068980 | 0.1483852 | -0.0692720 | 0.1964719 | 0.7639836 | 0.0411094 | -0.0129995 | -0.0853003 | 0.0066409 | 0.1047002 | 0.2204894 | 1.0000000 | 0.0294722 | -0.1323484 | 0.1847743 |
| TEAM\_PITCHING\_SO | -0.0633447 | -0.2548211 | 0.0741572 | -0.4282504 | 0.4091055 | 0.0785126 | 0.8257782 | -0.0373803 | -0.1070170 | 0.1211429 | -0.2741411 | 0.4190538 | 0.0294722 | 1.0000000 | -0.4012185 | 0.0063776 |
| TEAM\_FIELDING\_E | -0.0884986 | 0.0671167 | -0.2621964 | 0.5376888 | -0.6029884 | -0.3026772 | -0.5318635 | 0.2401202 | 0.1437337 | 0.0493273 | 0.3301306 | -0.5579081 | -0.1323484 | -0.4012185 | 1.0000000 | -0.2502193 |
| TEAM\_FIELDING\_DP | -0.0352128 | 0.1231197 | 0.1708179 | -0.1651927 | 0.2629838 | 0.2194524 | 0.0593358 | -0.2895276 | -0.0939819 | -0.0485670 | 0.0268468 | 0.2641637 | 0.1847743 | 0.0063776 | -0.2502193 | 1.0000000 |

1. Are any of the variables missing and need to be imputed "fixed"?

## 2. DATA PREPARATION (25 Points)

Describe how you have transformed the data by changing the original variables or creating new variables. If you did transform the data or create new variables, discuss why you did this. Here are some possible transformations.

1. Fix missing values (maybe with a Mean or Median value)
2. Create flags to suggest if a variable was missing
3. Transform data by putting it into buckets
4. Mathematical transforms such as log or square root (or use Box-Cox)
5. Combine variables (such as ratios or adding or multiplying) to create new variables

# and a refression on the raw data to see what comes out of it  
complete\_cols <- row.names( summary[summary$completeness == 1,])  
complete\_cols <- c(complete\_cols[complete\_cols!=target])  
  
training\_transformed <- training[,c(target,complete\_cols)]  
  
  
# make up for missing TEAM\_BATTING\_SO  
training[is.na(training$TEAM\_BATTING\_SO),]$TEAM\_BATTING\_SO <- median(training$TEAM\_BATTING\_SO, na.rm = TRUE)  
training\_transformed$TEAM\_BATTING\_SO <- training$TEAM\_BATTING\_SO  
  
# make up for missing TEAM\_BASERUN\_SB  
training[is.na(training$TEAM\_BASERUN\_SB),]$TEAM\_BASERUN\_SB <- median(training$TEAM\_BASERUN\_SB, na.rm = TRUE)  
training\_transformed$TEAM\_BASERUN\_SB <- training$TEAM\_BASERUN\_SB  
  
# make up for missing TEAM\_PITCHING\_SO  
training[is.na(training$TEAM\_PITCHING\_SO),]$TEAM\_PITCHING\_SO <- median(training$TEAM\_PITCHING\_SO, na.rm = TRUE)  
training\_transformed$TEAM\_PITCHING\_SO <- training$TEAM\_PITCHING\_SO  
  
# make up for missing TEAM\_FIELDING\_DP  
training[is.na(training$TEAM\_FIELDING\_DP),]$TEAM\_FIELDING\_DP <- median(training$TEAM\_FIELDING\_DP, na.rm = TRUE)  
training\_transformed$TEAM\_FIELDING\_DP <- training$TEAM\_FIELDING\_DP  
  
# make up for missing TEAM\_BASERUN\_CS  
#training[is.na(training$TEAM\_BASERUN\_CS),]$TEAM\_BASERUN\_CS <- median(training$TEAM\_BASERUN\_CS, na.rm = TRUE)  
#training\_transformed$TEAM\_BASERUN\_CS <- training$TEAM\_BASERUN\_CS   
  
  
#training\_transformed$TEAM\_BATTING\_1B <- (training\_transformed$TEAM\_BATTING\_2B-training\_transformed$TEAM\_BATTING\_3B+training\_transformed$TEAM\_BATTING\_HR)  
  
# validation algo  
#training\_transformed$INVALID <- FALSE  
#training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_1B < 811 | training\_transformed$TEAM\_BATTING\_1B>1338  
#length(training\_transformed$INVALID[training\_transformed$INVALID])  
  
training\_transformed$INVALID <- FALSE  
training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_2B < 110 | training\_transformed$TEAM\_BATTING\_2B>376  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 8

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_3B < 11 | training\_transformed$TEAM\_BATTING\_3B>153  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 20

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_HR < 3 | training\_transformed$TEAM\_BATTING\_HR>264  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 31

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_H < 935 | training\_transformed$TEAM\_BATTING\_H>2131  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 39

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_SO < 308 | training\_transformed$TEAM\_BATTING\_SO>1535  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 74

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_PITCHING\_SO < 308 | training\_transformed$TEAM\_PITCHING\_SO>1535  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 89

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BATTING\_BB < 282 | training\_transformed$TEAM\_BATTING\_BB>835  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 163

training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_PITCHING\_BB < 282 | training\_transformed$TEAM\_PITCHING\_BB>835  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 192

#training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BASERUN\_CS < 8 | training\_transformed$TEAM\_BASERUN\_CS>191  
#length(training\_transformed$INVALID[training\_transformed$INVALID])  
  
training\_transformed$INVALID <- training\_transformed$INVALID | training\_transformed$TEAM\_BASERUN\_SB < 13 | training\_transformed$TEAM\_BASERUN\_SB>638  
length(training\_transformed$INVALID[training\_transformed$INVALID])

## [1] 194

#training\_transformed <- training\_transformed[!training\_transformed$INVALID,]  
  
  
#training\_transformed$TEAM\_BATTING\_1B <- training\_transformed$TEAM\_BATTING\_1B/training\_transformed$TEAM\_BATTING\_H  
  
#training\_transformed$TEAM\_BATTING\_2B <- training\_transformed$TEAM\_BATTING\_2B/training\_transformed$TEAM\_BATTING\_H  
#training\_transformed$TEAM\_BATTING\_3B <- training\_transformed$TEAM\_BATTING\_3B/training\_transformed$TEAM\_BATTING\_H  
#training\_transformed$TEAM\_BATTING\_HR <- training\_transformed$TEAM\_BATTING\_HR/training\_transformed$TEAM\_BATTING\_H  
  
  
#training\_transformed$TEAM\_PITCHING\_HR <- training\_transformed$TEAM\_PITCHING\_HR/training\_transformed$TEAM\_PITCHING\_H  
  
  
library(MASS)  
valid\_data <- training\_transformed[!training\_transformed$INVALID&complete.cases(training\_transformed),colnames(training\_transformed)[!colnames(training\_transformed) %in% c("INVALID")]]

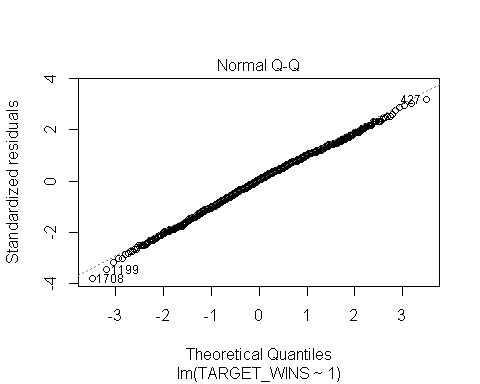
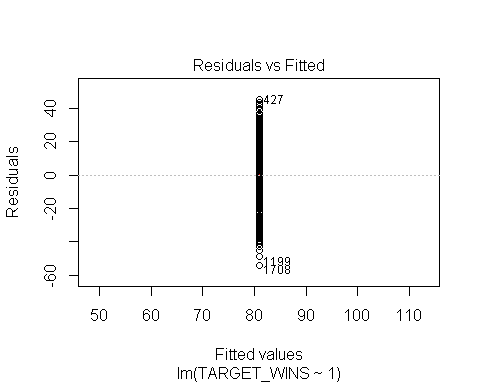
# Model 0: A Mean based model

Let's try a straight up regression whatever linear regression

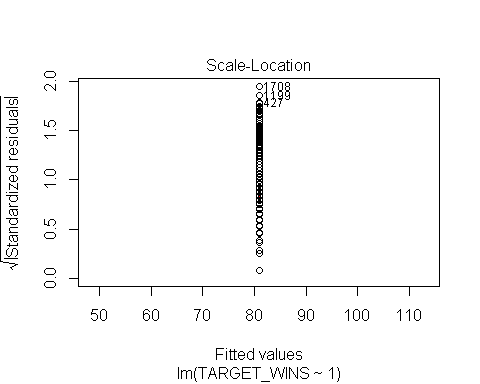
model <- lm(TARGET\_WINS~1,data=valid\_data)  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ 1, data = valid\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -53.899 -8.899 1.101 10.101 45.101   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 80.8987 0.3109 260.2 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.19 on 2081 degrees of freedom

model0 <- model  
model0\_y\_power <- 1  
plot(model)



## hat values (leverages) are all = 0.0004803074  
## and there are no factor predictors; no plot no. 5



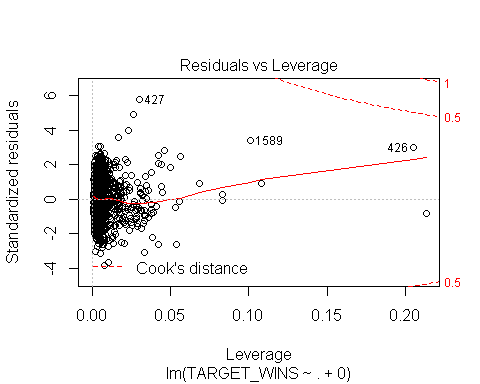
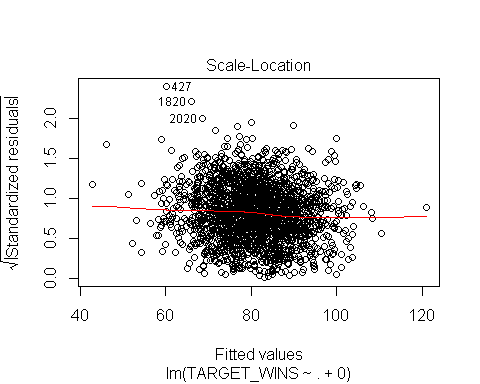
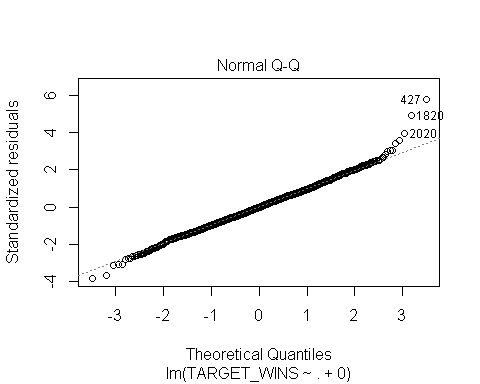
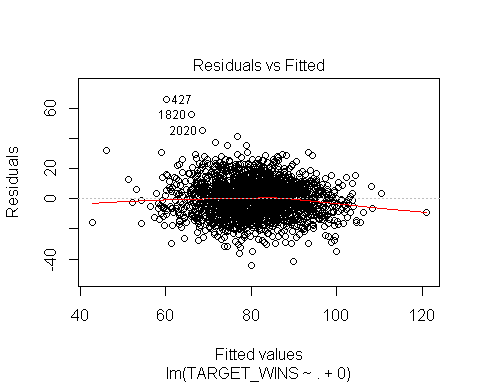
# Model 1: A Straight Up Model

Let's try a straight up regression whatever linear regression

model <- lm(TARGET\_WINS~.+0,data=valid\_data)  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ . + 0, data = valid\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -44.023 -7.526 0.035 7.644 65.759   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## TEAM\_BATTING\_H -0.036425 0.010741 -3.391 0.000709 \*\*\*  
## TEAM\_BATTING\_2B -0.061255 0.008525 -7.185 9.32e-13 \*\*\*  
## TEAM\_BATTING\_3B 0.159143 0.018703 8.509 < 2e-16 \*\*\*  
## TEAM\_BATTING\_HR 0.456476 0.076259 5.986 2.53e-09 \*\*\*  
## TEAM\_BATTING\_BB 0.388537 0.034731 11.187 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_H 0.085238 0.009865 8.640 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_HR -0.384458 0.072216 -5.324 1.13e-07 \*\*\*  
## TEAM\_PITCHING\_BB -0.329688 0.032378 -10.183 < 2e-16 \*\*\*  
## TEAM\_FIELDING\_E -0.063789 0.004234 -15.068 < 2e-16 \*\*\*  
## TEAM\_BATTING\_SO -0.120387 0.016864 -7.139 1.30e-12 \*\*\*  
## TEAM\_BASERUN\_SB 0.057361 0.004670 12.282 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_SO 0.113710 0.015607 7.286 4.53e-13 \*\*\*  
## TEAM\_FIELDING\_DP -0.081378 0.011974 -6.796 1.40e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.6 on 2069 degrees of freedom  
## Multiple R-squared: 0.9802, Adjusted R-squared: 0.9801   
## F-statistic: 7875 on 13 and 2069 DF, p-value: < 2.2e-16

model1 <- model  
model1\_y\_power <- 1  
model1\_powers <- 1  
plot(model)



# Model 2: A Power Model on Just Y

The residual curves look funky. Let's apply some power transformations to the same model:

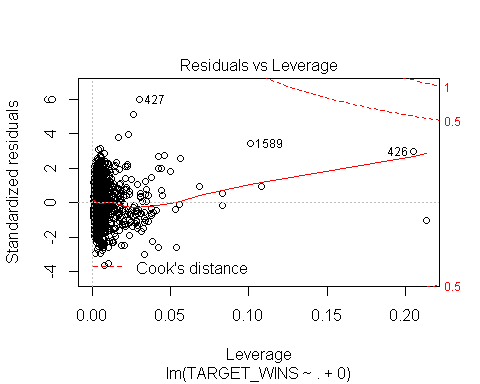
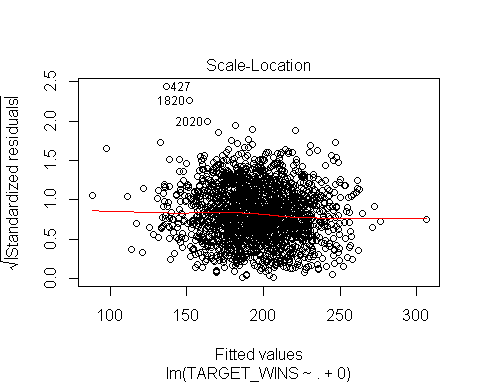
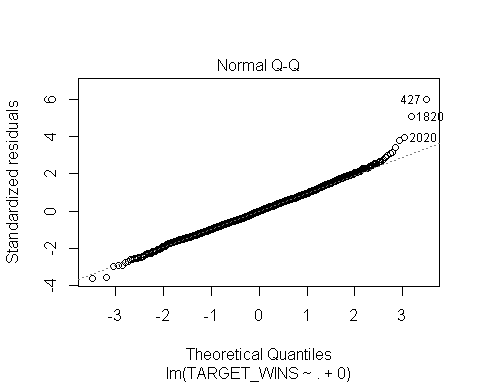
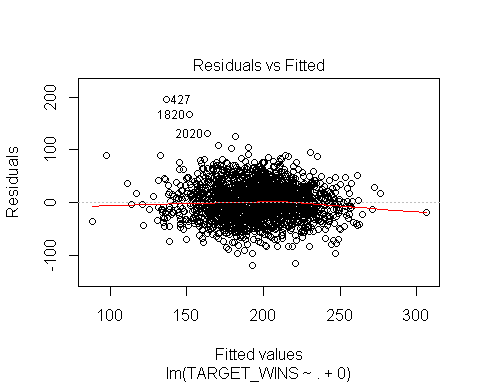
getBoxCoxCoef<-function(x,y){  
 if (min(x)<=0){1} else {  
 d<-boxcox(x~y, plotit = FALSE)  
 d <- data.frame(d)  
 max\_x <- d[d$y == max(d$y),]$x  
 if (abs(max\_x)<=(1/3)) 1 else max\_x  
 }  
}  
  
model\_data <- valid\_data  
  
d<-boxcox(TARGET\_WINS+0.0001~.,data=model\_data, plotit = FALSE)  
  
d <- data.frame(d)  
  
y\_power <- d[d$y == max(d$y),]$x  
y\_power

## [1] 1.2

model\_data$TARGET\_WINS <- with(model\_data,TARGET\_WINS^y\_power)  
model <- lm(TARGET\_WINS~.+0,data=model\_data)  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ . + 0, data = model\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -119.200 -22.007 -0.406 21.254 195.012   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## TEAM\_BATTING\_H -0.13425 0.03071 -4.371 1.30e-05 \*\*\*  
## TEAM\_BATTING\_2B -0.15388 0.02438 -6.312 3.36e-10 \*\*\*  
## TEAM\_BATTING\_3B 0.46912 0.05348 8.771 < 2e-16 \*\*\*  
## TEAM\_BATTING\_HR 1.33714 0.21807 6.132 1.04e-09 \*\*\*  
## TEAM\_BATTING\_BB 1.11073 0.09932 11.184 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_H 0.25192 0.02821 8.930 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_HR -1.09880 0.20651 -5.321 1.14e-07 \*\*\*  
## TEAM\_PITCHING\_BB -0.94701 0.09259 -10.228 < 2e-16 \*\*\*  
## TEAM\_FIELDING\_E -0.18482 0.01211 -15.267 < 2e-16 \*\*\*  
## TEAM\_BATTING\_SO -0.34393 0.04822 -7.132 1.36e-12 \*\*\*  
## TEAM\_BASERUN\_SB 0.16864 0.01335 12.628 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_SO 0.31431 0.04463 7.043 2.56e-12 \*\*\*  
## TEAM\_FIELDING\_DP -0.25300 0.03424 -7.389 2.14e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 33.16 on 2069 degrees of freedom  
## Multiple R-squared: 0.9726, Adjusted R-squared: 0.9724   
## F-statistic: 5650 on 13 and 2069 DF, p-value: < 2.2e-16

model2 <- model  
model2\_y\_power <- y\_power  
model2\_powers <- y\_power  
  
plot(model)



# Model 3: A Power Model on Y and X

The residual curves look funky. Let's apply some power transformations to the same model:

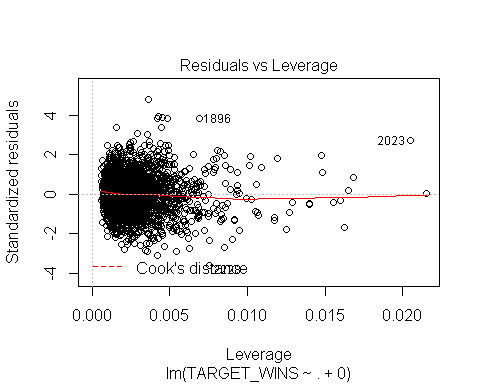
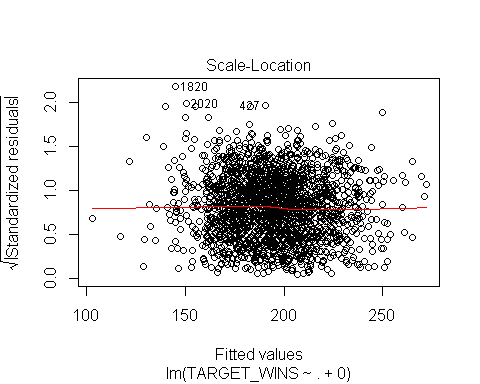
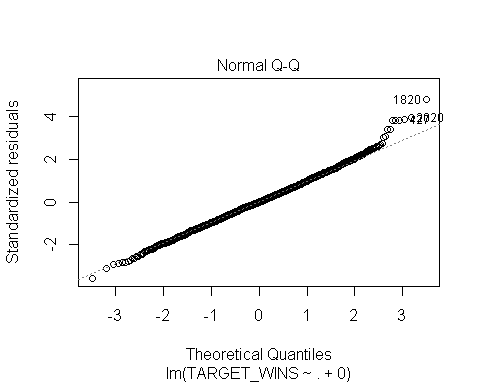
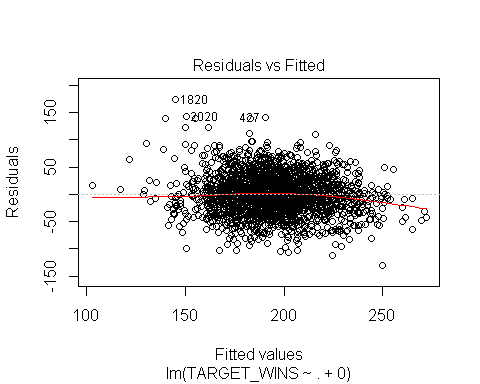
model\_data <- valid\_data[,c("TEAM\_BATTING\_H"  
,"TEAM\_BATTING\_3B"  
,"TEAM\_BATTING\_BB"  
,"TEAM\_FIELDING\_E"  
,"TEAM\_BASERUN\_SB"  
,"TEAM\_FIELDING\_DP","TARGET\_WINS")]  
#,"TEAM\_BATTING\_HR"  
#,"TEAM\_PITCHING\_HR"  
#,"TEAM\_PITCHING\_BB"  
  
source\_cols <- colnames(model\_data)  
  
source\_cols <- source\_cols[source\_cols!="TARGET\_WINS"]  
  
powers <- lapply(source\_cols,function(x){getBoxCoxCoef(model\_data[,x],model\_data$TARGET\_WINS^y\_power)})  
  
params <- c(unlist(powers),y\_power)  
kable(data.frame(variable=c(source\_cols,"TARGET\_WINS"),power=params))

|  |  |
| --- | --- |
| variable | power |
| TEAM\_BATTING\_H | -0.7 |
| TEAM\_BATTING\_3B | 1.0 |
| TEAM\_BATTING\_BB | 1.0 |
| TEAM\_FIELDING\_E | -1.0 |
| TEAM\_BASERUN\_SB | 1.0 |
| TEAM\_FIELDING\_DP | 1.7 |
| TARGET\_WINS | 1.2 |

power\_transforms <- data.frame(t(apply(model\_data,1,function(c) { c^params})))  
  
#model <- lm(TARGET\_WINS~.,data=power\_transforms[,colnames(power\_transforms)[!colnames(power\_transforms) %in% c("TEAM\_PITCHING\_H","TEAM\_BATTING\_H")]])  
#summary(model)  
model <- lm(TARGET\_WINS~.+0,power\_transforms)  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ . + 0, data = power\_transforms)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -129.735 -23.563 -0.472 23.635 173.959   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## TEAM\_BATTING\_H 2.659e+03 9.093e+02 2.925 0.00349 \*\*   
## TEAM\_BATTING\_3B 9.074e-01 4.267e-02 21.265 < 2e-16 \*\*\*  
## TEAM\_BATTING\_BB 1.417e-01 9.579e-03 14.791 < 2e-16 \*\*\*  
## TEAM\_FIELDING\_E 9.786e+03 4.942e+02 19.804 < 2e-16 \*\*\*  
## TEAM\_BASERUN\_SB 7.348e-02 1.182e-02 6.215 6.20e-10 \*\*\*  
## TEAM\_FIELDING\_DP -2.769e-03 6.525e-04 -4.244 2.29e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 36.45 on 2076 degrees of freedom  
## Multiple R-squared: 0.9668, Adjusted R-squared: 0.9667   
## F-statistic: 1.007e+04 on 6 and 2076 DF, p-value: < 2.2e-16

model3 <- model  
model3\_y\_power <- y\_power  
model3\_powers <- powers  
plot(model)



# Models With Advanced Feature/Transformation Selection: Evaluating transformations

Let's get fancy by trying other variables/combinations

combinations <- t(combn(colnames(valid\_data)[colnames(valid\_data)!="TARGET\_WINS"],2))  
  
try\_Variable <- function(x,y,label){  
 y\_power <- getBoxCoxCoef(y+0.00001,x)  
 x\_power <- getBoxCoxCoef(x+0.00001,y)  
 model <- lm(I(y^y\_power)~I(x^x\_power)+0)  
 s <- summary(model)  
 res <- data.frame(try=paste0(label,"^",x\_power),  
 r2=s$r.squared,  
 x\_power=x\_power,  
 y\_power=y\_power,  
 coef=s$coefficients[,"Estimate"][1],  
 coef\_p=s$coefficients[,"Pr(>|t|)"][1],  
 t\_value=s$coefficients[,"t value"][1],  
 min\_x=min(x),  
 max\_x=max(x),  
 f=s$fstatistic["value"],  
 mse=mean(s$residuals^2),  
 rse=sd(s$residuals),  
 stringsAsFactors=FALSE  
 )  
 rownames(res) <- NULL  
 res  
}  
  
evaluate\_variables <- function(data,combination,target){  
 try\_this <- data.frame(x = data[,combination[1]]\*data[,combination[2]],y=data[,target])  
 res\_1 <- try\_Variable(try\_this$x,try\_this$y,paste0("(",combination[1],"\*",combination[2],")"))  
  
 try\_this <- data.frame(x = data[,combination[1]]/(data[,combination[2]]+0.0000001),y=data[,target])  
 res\_2 <- try\_Variable(try\_this$x,try\_this$y,paste0("(",combination[1],"/",combination[2],")"))  
  
 total<-rbind(res\_1,res\_2)  
 rownames(total) <- NULL  
 total  
}  
  
evaluate\_variable <- function(data,variable,target){  
 try\_this <- data.frame(x = data[,variable],y=data[,target])  
 res\_1 <- try\_Variable(try\_this$x,try\_this$y,paste0("(",variable,")"))  
  
  
 rownames(res\_1) <- NULL  
 res\_1  
}  
  
evaluate\_variable\_log <- function(data,variable,target){  
 try\_this <- data.frame(x = log(data[,variable]+0.000001),y=data[,target])  
 res\_1 <- try\_Variable(try\_this$x,try\_this$y,paste0("log(",variable,")"))  
  
  
 rownames(res\_1) <- NULL  
 res\_1  
}  
  
unique\_cols <- c(unique(c(combinations[,1],combinations[,2])))  
  
  
the\_medium\_tamale<-lapply(unique\_cols,function(c){evaluate\_variable(valid\_data,c,"TARGET\_WINS")})  
the\_medium\_tamale<-do.call(rbind,the\_medium\_tamale)  
  
the\_log\_tamale<-lapply(unique\_cols,function(c){evaluate\_variable\_log(valid\_data,c,"TARGET\_WINS")})  
the\_log\_tamale<-do.call(rbind,the\_log\_tamale)  
  
the\_big\_tamale<-apply(combinations,1,function(c){evaluate\_variables(valid\_data,c,"TARGET\_WINS")})  
the\_big\_tamale<-do.call(rbind,the\_big\_tamale)  
  
the\_tamale <- rbind(the\_medium\_tamale,the\_log\_tamale,the\_big\_tamale)  
  
kable(the\_tamale[order(-the\_tamale$r2),])

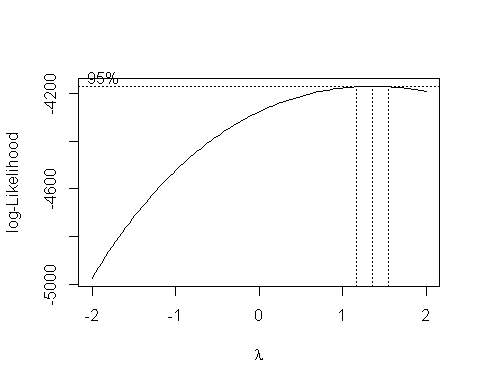
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | try | r2 | x\_power | y\_power | coef | coef\_p | t\_value | min\_x | max\_x | f | mse | rse |
| 21 | log(TEAM\_PITCHING\_BB)^1.9 | 0.9549735 | 1.9 | 1.3 | 9.223931e+00 | 0 | 210.08622 | 5.743003e+00 | 6.726233e+00 | 44136.2208 | 4373.330 | 66.14700 |
| 18 | log(TEAM\_BATTING\_BB)^2 | 0.9548636 | 2.0 | 1.3 | 7.805347e+00 | 0 | 209.81825 | 5.641907e+00 | 6.672033e+00 | 44023.6972 | 4384.004 | 66.22772 |
| 33 | (TEAM\_BATTING\_H\*TEAM\_BATTING\_BB)^0.7 | 0.9547029 | 0.7 | 1.3 | 2.317080e-02 | 0 | 209.42808 | 3.388770e+05 | 1.336995e+06 | 43860.1189 | 4399.613 | 66.29999 |
| 27 | (TEAM\_BATTING\_H\*TEAM\_BATTING\_2B)^0.4 | 0.9544827 | 0.4 | 1.3 | 1.837889e+00 | 0 | 208.89672 | 1.346380e+05 | 6.809880e+05 | 43637.8393 | 4421.004 | 66.49837 |
| 61 | (TEAM\_BATTING\_2B\*TEAM\_PITCHING\_BB)^0.4 | 0.9542975 | 0.4 | 1.3 | 2.719403e+00 | 0 | 208.45286 | 3.936000e+04 | 2.776100e+05 | 43452.5948 | 4438.990 | 66.62385 |
| 15 | log(TEAM\_BATTING\_2B)^2 | 0.9538742 | 2.0 | 1.3 | 1.013745e+01 | 0 | 207.44819 | 4.770685e+00 | 5.921578e+00 | 43034.7526 | 4480.102 | 66.94838 |
| 2 | (TEAM\_BATTING\_2B)^0.6 | 0.9503714 | 0.6 | 1.3 | 1.126700e+01 | 0 | 199.62586 | 1.180000e+02 | 3.730000e+02 | 39850.4841 | 4820.320 | 69.41189 |
| 14 | log(TEAM\_BATTING\_H)^-2 | 0.9481607 | -2.0 | 1.3 | 1.607834e+04 | 0 | 195.09552 | 7.036149e+00 | 7.618251e+00 | 38062.2612 | 5035.046 | 70.97196 |
| 26 | log(TEAM\_FIELDING\_DP)^2 | 0.9461268 | 2.0 | 1.3 | 1.217963e+01 | 0 | 191.17211 | 4.219508e+00 | 5.416100e+00 | 36546.7752 | 5232.586 | 72.33528 |
| 55 | (TEAM\_BATTING\_2B\*TEAM\_BATTING\_BB)^0.6 | 0.9455240 | 0.6 | 1.3 | 2.615050e-01 | 0 | 190.05080 | 3.504600e+04 | 2.466240e+05 | 36119.3064 | 5291.140 | 72.55705 |
| 5 | (TEAM\_BATTING\_BB)^1 | 0.9436301 | 1.0 | 1.3 | 5.739900e-01 | 0 | 186.64364 | 2.820000e+02 | 7.900000e+02 | 34835.8478 | 5475.093 | 73.80265 |
| 28 | (TEAM\_BATTING\_H/TEAM\_BATTING\_2B)^-0.8 | 0.9427479 | -0.8 | 1.3 | 1.270060e+03 | 0 | 185.11351 | 4.253125e+00 | 1.059350e+01 | 34267.0105 | 5560.776 | 74.49935 |
| 8 | (TEAM\_PITCHING\_BB)^1 | 0.9424247 | 1.0 | 1.3 | 5.428864e-01 | 0 | 184.56165 | 3.120000e+02 | 8.340000e+02 | 34063.0010 | 5592.163 | 74.56111 |
| 22 | log(TEAM\_FIELDING\_E)^-2 | 0.9421453 | -2.0 | 1.2 | 4.975845e+03 | 0 | 184.08813 | 4.174387e+00 | 6.872128e+00 | 33888.4397 | 2307.454 | 47.86170 |
| 1 | (TEAM\_BATTING\_H)^-0.7 | 0.9406519 | -0.7 | 1.3 | 4.926394e+04 | 0 | 181.61305 | 1.137000e+03 | 2.035000e+03 | 32983.3004 | 5764.357 | 75.90784 |
| 16 | log(TEAM\_BATTING\_3B)^0.8 | 0.9406243 | 0.8 | 1.4 | 1.596827e+02 | 0 | 181.56814 | 2.397895e+00 | 4.990433e+00 | 32966.9876 | 14058.369 | 118.54831 |
| 19 | log(TEAM\_PITCHING\_H)^-2 | 0.9405439 | -2.0 | 1.4 | 2.535857e+04 | 0 | 181.43769 | 7.036149e+00 | 7.972811e+00 | 32919.6370 | 14077.388 | 118.66866 |
| 111 | (TEAM\_BATTING\_BB\*TEAM\_PITCHING\_H)^1 | 0.9372627 | 1.0 | 1.4 | 5.746000e-04 | 0 | 176.32093 | 3.629340e+05 | 1.641186e+06 | 31089.0696 | 14854.282 | 121.32978 |
| 39 | (TEAM\_BATTING\_H\*TEAM\_PITCHING\_BB)^1 | 0.9372625 | 1.0 | 1.4 | 5.746000e-04 | 0 | 176.32063 | 3.625440e+05 | 1.640210e+06 | 31088.9643 | 14854.329 | 121.32993 |
| 24 | log(TEAM\_BASERUN\_SB)^0.9 | 0.9368255 | 0.9 | 1.4 | 1.182747e+02 | 0 | 175.66878 | 2.890372e+00 | 6.448889e+00 | 30859.5186 | 14957.796 | 122.20623 |
| 25 | log(TEAM\_PITCHING\_SO)^2 | 0.9362389 | 2.0 | 1.4 | 1.059959e+01 | 0 | 174.80399 | 5.831882e+00 | 7.307202e+00 | 30556.4365 | 15096.699 | 122.83648 |
| 43 | (TEAM\_BATTING\_H\*TEAM\_BATTING\_SO)^0.5 | 0.9358626 | 0.5 | 1.3 | 2.871700e-01 | 0 | 174.25544 | 4.859730e+05 | 1.927822e+06 | 30364.9591 | 6229.533 | 78.75367 |
| 182 | (TEAM\_PITCHING\_SO/TEAM\_FIELDING\_DP)^0.4 | 0.9349638 | 0.4 | 1.3 | 1.521065e+02 | 0 | 172.96410 | 1.811594e+00 | 1.300000e+01 | 29916.5797 | 6316.827 | 79.32624 |
| 98 | (TEAM\_BATTING\_HR/TEAM\_PITCHING\_HR)^2 | 0.9349085 | 2.0 | 1.3 | 3.291615e+02 | 0 | 172.88539 | 5.121951e-01 | 1.014815e+00 | 29889.3596 | 6322.205 | 79.34657 |
| 36 | (TEAM\_BATTING\_H/TEAM\_PITCHING\_H)^2 | 0.9347578 | 2.0 | 1.3 | 3.291342e+02 | 0 | 172.67170 | 5.184419e-01 | 1.018544e+00 | 29815.5166 | 6336.841 | 79.43780 |
| 116 | (TEAM\_BATTING\_BB/TEAM\_PITCHING\_BB)^2 | 0.9347576 | 2.0 | 1.3 | 3.291232e+02 | 0 | 172.67152 | 5.183824e-01 | 1.018116e+00 | 29815.4547 | 6336.854 | 79.43791 |
| 23 | log(TEAM\_BATTING\_SO)^2 | 0.9347420 | 2.0 | 1.4 | 1.075618e+01 | 0 | 172.64936 | 5.730100e+00 | 7.243513e+00 | 29807.8009 | 15451.116 | 124.24704 |
| 174 | (TEAM\_BATTING\_SO/TEAM\_PITCHING\_SO)^2 | 0.9347301 | 2.0 | 1.3 | 3.294563e+02 | 0 | 172.63258 | 5.181932e-01 | 1.018735e+00 | 29802.0080 | 6339.526 | 79.45432 |
| 126 | (TEAM\_BATTING\_BB/TEAM\_FIELDING\_DP)^1 | 0.9337496 | 1.0 | 1.3 | 8.198457e+01 | 0 | 171.26042 | 1.883721e+00 | 6.882353e+00 | 29330.1318 | 6434.762 | 79.70940 |
| 139 | (TEAM\_PITCHING\_H\*TEAM\_FIELDING\_DP)^0.7 | 0.9314228 | 0.7 | 1.4 | 8.279110e-02 | 0 | 168.12007 | 8.798400e+04 | 4.265220e+05 | 28264.3595 | 16236.998 | 127.16550 |
| 161 | (TEAM\_PITCHING\_BB\*TEAM\_FIELDING\_DP)^0.7 | 0.9312849 | 0.7 | 1.3 | 1.086170e-01 | 0 | 167.93883 | 2.246400e+04 | 1.660020e+05 | 28203.4498 | 6674.157 | 81.29052 |
| 34 | (TEAM\_BATTING\_H/TEAM\_BATTING\_BB)^-1 | 0.9306758 | -1.0 | 1.3 | 8.264773e+02 | 0 | 167.14479 | 1.831800e+00 | 5.404844e+00 | 27937.3796 | 6733.314 | 81.69830 |
| 130 | (TEAM\_PITCHING\_H/TEAM\_PITCHING\_BB)^-1 | 0.9306711 | -1.0 | 1.3 | 8.264845e+02 | 0 | 167.13874 | 1.830212e+00 | 5.409884e+00 | 27935.3594 | 6733.767 | 81.70091 |
| 123 | (TEAM\_BATTING\_BB\*TEAM\_PITCHING\_SO)^0.5 | 0.9305833 | 0.5 | 1.3 | 4.645181e-01 | 0 | 167.02514 | 1.335180e+05 | 8.856150e+05 | 27897.3958 | 6742.294 | 81.76542 |
| 155 | (TEAM\_PITCHING\_BB\*TEAM\_BATTING\_SO)^0.5 | 0.9305141 | 0.5 | 1.3 | 4.645880e-01 | 0 | 166.93565 | 1.334720e+05 | 8.861600e+05 | 27867.5109 | 6749.022 | 81.80461 |
| 162 | (TEAM\_PITCHING\_BB/TEAM\_FIELDING\_DP)^-0.4 | 0.9292657 | -0.4 | 1.4 | 7.877281e+02 | 0 | 165.34501 | 1.883721e+00 | 7.680000e+00 | 27338.9718 | 16747.724 | 129.32289 |
| 72 | (TEAM\_BATTING\_2B/TEAM\_FIELDING\_DP)^1 | 0.9287250 | 1.0 | 1.3 | 1.755874e+02 | 0 | 164.66863 | 7.919463e-01 | 3.759036e+00 | 27115.7571 | 6922.795 | 82.52841 |
| 119 | (TEAM\_BATTING\_BB\*TEAM\_BATTING\_SO)^0.5 | 0.9262970 | 0.5 | 1.3 | 4.739747e-01 | 0 | 161.72177 | 1.121320e+05 | 8.803300e+05 | 26153.9307 | 7158.621 | 84.13809 |
| 176 | (TEAM\_BATTING\_SO/TEAM\_FIELDING\_DP)^0.5 | 0.9241491 | 0.5 | 1.3 | 1.300300e+02 | 0 | 159.23071 | 1.700483e+00 | 1.146226e+01 | 25354.4190 | 7367.234 | 85.42021 |
| 79 | (TEAM\_BATTING\_3B\*TEAM\_PITCHING\_HR)^0.5 | 0.9218125 | 0.5 | 1.3 | 4.307556e+00 | 0 | 156.63491 | 1.320000e+02 | 2.138400e+04 | 24534.4961 | 7594.191 | 85.92949 |
| 73 | (TEAM\_BATTING\_3B\*TEAM\_BATTING\_HR)^0.5 | 0.9212416 | 0.5 | 1.3 | 4.420346e+00 | 0 | 156.01786 | 1.320000e+02 | 1.792800e+04 | 24341.5738 | 7649.639 | 86.23405 |
| 57 | (TEAM\_BATTING\_2B\*TEAM\_PITCHING\_H)^1 | 0.9208560 | 1.0 | 1.3 | 7.702000e-04 | 0 | 155.60478 | 1.627220e+05 | 8.343320e+05 | 24212.8490 | 7687.089 | 86.50216 |
| 58 | (TEAM\_BATTING\_2B/TEAM\_PITCHING\_H)^1.1 | 0.9206387 | 1.1 | 1.3 | 2.234225e+03 | 0 | 155.37322 | 7.812500e-02 | 2.351212e-01 | 24140.8361 | 7708.200 | 87.14028 |
| 124 | (TEAM\_BATTING\_BB/TEAM\_PITCHING\_SO)^-0.4 | 0.9206046 | -0.4 | 1.4 | 3.906198e+02 | 0 | 155.33699 | 2.149466e-01 | 1.701058e+00 | 24129.5806 | 18798.422 | 136.79963 |
| 40 | (TEAM\_BATTING\_H/TEAM\_PITCHING\_BB)^1 | 0.9165435 | 1.0 | 1.3 | 1.077884e+02 | 0 | 151.17591 | 1.600242e+00 | 4.895385e+00 | 22854.1554 | 8105.952 | 89.47011 |
| 181 | (TEAM\_PITCHING\_SO\*TEAM\_FIELDING\_DP)^0.5 | 0.9151709 | 0.5 | 1.4 | 1.353169e+00 | 0 | 149.83554 | 3.486000e+04 | 2.723040e+05 | 22450.6901 | 20084.941 | 141.09149 |
| 125 | (TEAM\_BATTING\_BB\*TEAM\_FIELDING\_DP)^0.9 | 0.9149586 | 0.9 | 1.3 | 1.158920e-02 | 0 | 149.63095 | 2.138400e+04 | 1.578990e+05 | 22389.4219 | 8259.897 | 89.82237 |
| 132 | (TEAM\_PITCHING\_H/TEAM\_FIELDING\_E)^0.7 | 0.9143641 | 0.7 | 1.2 | 3.851974e+01 | 0 | 149.06225 | 2.135546e+00 | 2.321538e+01 | 22219.5546 | 3415.471 | 57.37624 |
| 101 | (TEAM\_BATTING\_HR\*TEAM\_FIELDING\_E)^0.5 | 0.9109602 | 0.5 | 1.3 | 2.323468e+00 | 0 | 145.91290 | 7.440000e+02 | 1.765050e+05 | 21290.5750 | 8648.247 | 91.83605 |
| 56 | (TEAM\_BATTING\_2B/TEAM\_BATTING\_BB)^1 | 0.9086917 | 1.0 | 1.3 | 6.124455e+02 | 0 | 143.90940 | 2.271341e-01 | 9.858156e-01 | 20709.9141 | 8868.585 | 93.19120 |
| 49 | (TEAM\_BATTING\_H\*TEAM\_FIELDING\_DP)^1.3 | 0.9084966 | 1.3 | 1.3 | 3.330000e-05 | 0 | 143.74046 | 8.366400e+04 | 3.409900e+05 | 20661.3204 | 8887.535 | 92.99003 |
| 12 | (TEAM\_PITCHING\_SO)^0.7 | 0.9071753 | 0.7 | 1.4 | 4.258396e+00 | 0 | 142.61000 | 3.410000e+02 | 1.491000e+03 | 20337.6128 | 21978.057 | 147.21373 |
| 62 | (TEAM\_BATTING\_2B/TEAM\_PITCHING\_BB)^1 | 0.9056863 | 1.0 | 1.3 | 6.436652e+02 | 0 | 141.36362 | 2.104520e-01 | 7.962529e-01 | 19983.6738 | 9160.487 | 94.62303 |
| 168 | (TEAM\_FIELDING\_E/TEAM\_PITCHING\_SO)^-0.4 | 0.9039055 | -0.4 | 1.3 | 1.534208e+02 | 0 | 139.90984 | 6.250000e-02 | 1.354455e+00 | 19574.7644 | 9333.458 | 95.23040 |
| 112 | (TEAM\_BATTING\_BB/TEAM\_PITCHING\_H)^1.3 | 0.9011741 | 1.3 | 1.3 | 1.147193e+03 | 0 | 137.75427 | 9.720790e-02 | 5.443925e-01 | 18976.2376 | 9598.750 | 96.39248 |
| 71 | (TEAM\_BATTING\_2B\*TEAM\_FIELDING\_DP)^1 | 0.9004908 | 1.0 | 1.3 | 7.944700e-03 | 0 | 137.22841 | 9.225000e+03 | 6.676700e+04 | 18831.6371 | 9665.121 | 96.59483 |
| 175 | (TEAM\_BATTING\_SO\*TEAM\_FIELDING\_DP)^0.6 | 0.8993012 | 0.6 | 1.4 | 4.259592e-01 | 0 | 136.32529 | 2.929900e+04 | 2.328150e+05 | 18584.5844 | 23842.421 | 152.95918 |
| 129 | (TEAM\_PITCHING\_H\*TEAM\_PITCHING\_BB)^-0.7 | 0.8986440 | -0.7 | 1.4 | 6.247934e+06 | 0 | 135.83296 | 3.812640e+05 | 2.013388e+06 | 18450.5934 | 23998.018 | 153.94131 |
| 163 | (TEAM\_FIELDING\_E\*TEAM\_BATTING\_SO)^-1 | 0.8974924 | -1.0 | 1.2 | 2.153558e+07 | 0 | 134.98122 | 5.959500e+04 | 6.387040e+05 | 18219.9291 | 4088.376 | 62.13822 |
| 47 | (TEAM\_BATTING\_H\*TEAM\_PITCHING\_SO)^1 | 0.8970305 | 1.0 | 1.3 | 2.469000e-04 | 0 | 134.64351 | 5.192250e+05 | 2.419893e+06 | 18128.8739 | 10001.208 | 98.33906 |
| 133 | (TEAM\_PITCHING\_H\*TEAM\_BATTING\_SO)^1 | 0.8968308 | 1.0 | 1.3 | 2.470000e-04 | 0 | 134.49815 | 5.196440e+05 | 2.418994e+06 | 18089.7518 | 10020.606 | 98.42649 |
| 64 | (TEAM\_BATTING\_2B/TEAM\_FIELDING\_E)^0.6 | 0.8964228 | 0.6 | 1.2 | 1.420119e+02 | 0 | 134.20247 | 1.987768e-01 | 4.954545e+00 | 18010.3037 | 4131.032 | 62.57527 |
| 154 | (TEAM\_PITCHING\_BB/TEAM\_FIELDING\_E)^0.7 | 0.8953688 | 0.7 | 1.2 | 7.665523e+01 | 0 | 133.44626 | 5.502591e-01 | 1.042647e+01 | 17807.9030 | 4173.072 | 62.73178 |
| 173 | (TEAM\_BATTING\_SO\*TEAM\_PITCHING\_SO)^0.4 | 0.8951724 | 0.4 | 1.4 | 2.194191e+00 | 0 | 133.30655 | 1.091200e+05 | 1.957201e+06 | 17770.6373 | 24819.996 | 155.79961 |
| 20 | log(TEAM\_PITCHING\_HR)^2 | 0.8931147 | 2.0 | 1.3 | 1.368940e+01 | 0 | 131.86536 | 1.098613e+00 | 5.837731e+00 | 17388.4733 | 10381.542 | 99.57654 |
| 115 | (TEAM\_BATTING\_BB\*TEAM\_PITCHING\_BB)^1 | 0.8927732 | 1.0 | 1.3 | 9.548000e-04 | 0 | 131.63005 | 9.266400e+04 | 6.564900e+05 | 17326.4690 | 10414.710 | 99.12873 |
| 84 | (TEAM\_BATTING\_3B/TEAM\_FIELDING\_E)^1 | 0.8920061 | 1.0 | 1.3 | 9.628175e+02 | 0 | 131.10535 | 5.360130e-02 | 8.015267e-01 | 17188.6139 | 10489.217 | 99.16643 |
| 50 | (TEAM\_BATTING\_H/TEAM\_FIELDING\_DP)^-1.2 | 0.8912777 | -1.2 | 1.4 | 6.979993e+03 | 0 | 130.61208 | 6.538813e+00 | 1.980488e+01 | 17059.5159 | 25742.125 | 158.90123 |
| 35 | (TEAM\_BATTING\_H\*TEAM\_PITCHING\_H)^-1 | 0.8908310 | -1.0 | 1.4 | 9.895888e+08 | 0 | 130.31187 | 1.292769e+06 | 5.083430e+06 | 16981.1823 | 25847.909 | 159.42048 |
| 118 | (TEAM\_BATTING\_BB/TEAM\_FIELDING\_E)^0.7 | 0.8874507 | 0.7 | 1.2 | 7.821970e+01 | 0 | 128.09637 | 3.326425e-01 | 1.048529e+01 | 16408.6793 | 4488.873 | 64.79859 |
| 17 | log(TEAM\_BATTING\_HR)^2 | 0.8858472 | 2.0 | 1.3 | 1.387135e+01 | 0 | 127.07853 | 1.098613e+00 | 5.575949e+00 | 16148.9533 | 11087.421 | 102.54990 |
| 137 | (TEAM\_PITCHING\_H\*TEAM\_PITCHING\_SO)^1 | 0.8856014 | 1.0 | 1.3 | 2.304000e-04 | 0 | 126.92431 | 5.499300e+05 | 3.438246e+06 | 16109.7799 | 11111.297 | 103.10365 |
| 120 | (TEAM\_BATTING\_BB/TEAM\_BATTING\_SO)^-0.7 | 0.8852021 | -0.7 | 1.4 | 3.409643e+02 | 0 | 126.67484 | 3.136033e-01 | 2.087662e+00 | 16046.5156 | 27180.644 | 162.91433 |
| 160 | (TEAM\_PITCHING\_BB/TEAM\_PITCHING\_SO)^-0.7 | 0.8850550 | -0.7 | 1.4 | 3.407588e+02 | 0 | 126.58321 | 3.138790e-01 | 2.087302e+00 | 16023.3098 | 27215.484 | 163.00953 |
| 70 | (TEAM\_BATTING\_2B/TEAM\_PITCHING\_SO)^-0.7 | 0.8847217 | -0.7 | 1.4 | 1.912530e+02 | 0 | 126.37626 | 1.237762e-01 | 9.536785e-01 | 15970.9597 | 27294.408 | 163.21951 |
| 13 | (TEAM\_FIELDING\_DP)^1.7 | 0.8846216 | 1.7 | 1.3 | 5.726780e-02 | 0 | 126.31433 | 6.800000e+01 | 2.250000e+02 | 15955.3099 | 11206.459 | 103.81673 |
| 141 | (TEAM\_PITCHING\_HR\*TEAM\_PITCHING\_BB)^0.5 | 0.8799533 | 0.5 | 1.3 | 1.177697e+00 | 0 | 123.50677 | 1.575000e+03 | 2.695980e+05 | 15253.9227 | 11659.882 | 104.44118 |
| 10 | (TEAM\_BATTING\_SO)^0.9 | 0.8792554 | 0.9 | 1.4 | 1.128272e+00 | 0 | 123.10047 | 3.080000e+02 | 1.399000e+03 | 15153.7267 | 28588.649 | 166.17348 |
| 6 | (TEAM\_PITCHING\_H)^-2 | 0.8786705 | -2.0 | 1.4 | 1.020274e+09 | 0 | 122.76252 | 1.137000e+03 | 2.901000e+03 | 15070.6360 | 28727.146 | 167.30925 |
| 156 | (TEAM\_PITCHING\_BB/TEAM\_BATTING\_SO)^-0.7 | 0.8779145 | -0.7 | 1.4 | 3.496237e+02 | 0 | 122.32918 | 3.150457e-01 | 2.561688e+00 | 14964.4287 | 28906.140 | 167.53693 |
| 42 | (TEAM\_BATTING\_H/TEAM\_FIELDING\_E)^0.9 | 0.8760511 | 0.9 | 1.2 | 2.408561e+01 | 0 | 121.27727 | 1.665285e+00 | 2.321538e+01 | 14708.1774 | 4943.532 | 67.58122 |
| 169 | (TEAM\_FIELDING\_E\*TEAM\_FIELDING\_DP)^-0.9 | 0.8760103 | -0.9 | 1.3 | 2.362580e+06 | 0 | 121.25448 | 9.250000e+03 | 1.437850e+05 | 14702.6488 | 12042.862 | 106.18616 |
| 152 | (TEAM\_PITCHING\_HR/TEAM\_FIELDING\_DP)^0.6 | 0.8746460 | 0.6 | 1.3 | 3.456226e+02 | 0 | 120.49890 | 2.985070e-02 | 2.289256e+00 | 14519.9837 | 12175.373 | 106.57679 |
| 164 | (TEAM\_FIELDING\_E/TEAM\_BATTING\_SO)^-0.5 | 0.8726889 | -0.5 | 1.3 | 1.278938e+02 | 0 | 119.43530 | 6.250000e-02 | 2.178330e+00 | 14264.7918 | 12365.455 | 108.00129 |
| 167 | (TEAM\_FIELDING\_E\*TEAM\_PITCHING\_SO)^-1.1 | 0.8724395 | -1.1 | 1.3 | 1.097220e+08 | 0 | 119.30143 | 6.264000e+04 | 8.974650e+05 | 14232.8307 | 12389.681 | 107.15735 |
| 149 | (TEAM\_PITCHING\_HR\*TEAM\_PITCHING\_SO)^0.4 | 0.8723648 | 0.4 | 1.3 | 3.031076e+00 | 0 | 119.26140 | 2.440500e+03 | 4.434990e+05 | 14223.2822 | 12396.937 | 107.90338 |
| 140 | (TEAM\_PITCHING\_H/TEAM\_FIELDING\_DP)^-1.3 | 0.8718654 | -1.3 | 1.4 | 9.115771e+03 | 0 | 118.99470 | 6.652850e+00 | 2.287952e+01 | 14159.7384 | 30338.371 | 171.26208 |
| 110 | (TEAM\_BATTING\_HR/TEAM\_FIELDING\_DP)^0.6 | 0.8718002 | 0.6 | 1.2 | 2.260290e+02 | 0 | 118.95998 | 2.985070e-02 | 1.990566e+00 | 14151.4774 | 5113.072 | 68.65961 |
| 59 | (TEAM\_BATTING\_2B\*TEAM\_PITCHING\_HR)^0.5 | 0.8708639 | 0.5 | 1.3 | 1.747809e+00 | 0 | 118.46424 | 4.650000e+02 | 1.179920e+05 | 14033.7752 | 12542.723 | 107.89800 |
| 95 | (TEAM\_BATTING\_HR\*TEAM\_PITCHING\_H)^0.6 | 0.8696911 | 0.6 | 1.2 | 1.376472e-01 | 0 | 117.85054 | 3.840000e+03 | 5.721660e+05 | 13888.7491 | 5197.190 | 69.01747 |
| 37 | (TEAM\_BATTING\_H\*TEAM\_PITCHING\_HR)^0.6 | 0.8696812 | 0.6 | 1.2 | 1.376426e-01 | 0 | 117.84537 | 3.603000e+03 | 5.728100e+05 | 13887.5308 | 5197.587 | 69.02049 |
| 107 | (TEAM\_BATTING\_HR\*TEAM\_PITCHING\_SO)^0.4 | 0.8682998 | 0.4 | 1.3 | 3.071723e+00 | 0 | 117.13258 | 2.440500e+03 | 3.090270e+05 | 13720.0403 | 12791.763 | 109.39180 |
| 145 | (TEAM\_PITCHING\_HR\*TEAM\_BATTING\_SO)^0.4 | 0.8681925 | 0.4 | 1.3 | 3.071620e+00 | 0 | 117.07767 | 2.250000e+03 | 3.093860e+05 | 13707.1803 | 12802.182 | 109.43142 |
| 127 | (TEAM\_PITCHING\_H\*TEAM\_PITCHING\_HR)^0.6 | 0.8681608 | 0.6 | 1.3 | 2.088862e-01 | 0 | 117.06142 | 3.840000e+03 | 8.211420e+05 | 13703.3752 | 12805.268 | 108.72596 |
| 66 | (TEAM\_BATTING\_2B/TEAM\_BATTING\_SO)^-0.8 | 0.8678681 | -0.8 | 1.4 | 1.727150e+02 | 0 | 116.91199 | 1.447124e-01 | 1.043750e+00 | 13668.4142 | 31284.818 | 173.67346 |
| 146 | (TEAM\_PITCHING\_HR/TEAM\_BATTING\_SO)^0.8 | 0.8665868 | 0.8 | 1.3 | 1.327875e+03 | 0 | 116.26332 | 4.000000e-03 | 4.843750e-01 | 13517.1591 | 12958.142 | 108.94723 |
| 48 | (TEAM\_BATTING\_H/TEAM\_PITCHING\_SO)^-0.9 | 0.8653463 | -0.9 | 1.4 | 7.432392e+02 | 0 | 115.64365 | 9.154930e-01 | 5.178117e+00 | 13373.4534 | 31881.914 | 174.95570 |
| 53 | (TEAM\_BATTING\_2B\*TEAM\_BATTING\_HR)^0.5 | 0.8650931 | 0.5 | 1.3 | 1.774729e+00 | 0 | 115.51818 | 4.650000e+02 | 8.829800e+04 | 13344.4507 | 13103.226 | 109.96521 |
| 159 | (TEAM\_PITCHING\_BB\*TEAM\_PITCHING\_SO)^1 | 0.8641337 | 1.0 | 1.3 | 6.221000e-04 | 0 | 115.04576 | 1.449000e+05 | 1.167453e+06 | 13235.5266 | 13196.410 | 110.98448 |
| 31 | (TEAM\_BATTING\_H\*TEAM\_BATTING\_HR)^0.6 | 0.8631175 | 0.6 | 1.2 | 1.402228e-01 | 0 | 114.55051 | 3.603000e+03 | 4.155360e+05 | 13121.8185 | 5459.371 | 70.49801 |
| 103 | (TEAM\_BATTING\_HR\*TEAM\_BATTING\_SO)^0.4 | 0.8625657 | 0.4 | 1.3 | 3.104020e+00 | 0 | 114.28376 | 2.250000e+03 | 2.944550e+05 | 13060.7785 | 13348.707 | 111.43059 |
| 144 | (TEAM\_PITCHING\_HR/TEAM\_FIELDING\_E)^0.4 | 0.8622188 | 0.4 | 1.2 | 2.083609e+02 | 0 | 114.11685 | 1.209680e-02 | 3.069444e+00 | 13022.6549 | 5495.215 | 70.85034 |
| 85 | (TEAM\_BATTING\_3B\*TEAM\_BATTING\_SO)^1 | 0.8613825 | 1.0 | 1.4 | 1.170850e-02 | 0 | 113.71691 | 9.933000e+03 | 1.066240e+05 | 12931.5359 | 32820.408 | 175.33443 |
| 117 | (TEAM\_BATTING\_BB\*TEAM\_FIELDING\_E)^-0.9 | 0.8609176 | -0.9 | 1.3 | 7.305154e+06 | 0 | 113.49603 | 3.515000e+04 | 4.203420e+05 | 12881.3493 | 13508.785 | 112.38091 |
| 165 | (TEAM\_FIELDING\_E\*TEAM\_BASERUN\_SB)^-0.4 | 0.8578217 | -0.4 | 1.3 | 1.277776e+04 | 0 | 112.05156 | 1.800000e+03 | 4.028760e+05 | 12555.5511 | 13809.480 | 113.64333 |
| 102 | (TEAM\_BATTING\_HR/TEAM\_FIELDING\_E)^0.4 | 0.8568362 | 0.4 | 1.2 | 2.103567e+02 | 0 | 111.60105 | 1.209680e-02 | 3.069444e+00 | 12454.7941 | 5709.893 | 72.01366 |
| 170 | (TEAM\_FIELDING\_E/TEAM\_FIELDING\_DP)^-0.8 | 0.8533684 | -0.8 | 1.3 | 2.862422e+02 | 0 | 110.05013 | 3.793103e-01 | 6.476510e+00 | 12111.0319 | 14242.017 | 114.47516 |
| 63 | (TEAM\_BATTING\_2B\*TEAM\_FIELDING\_E)^-1 | 0.8532671 | -1.0 | 1.3 | 1.034411e+07 | 0 | 110.00560 | 1.818300e+04 | 2.373900e+05 | 12101.2323 | 14251.858 | 114.83529 |
| 9 | (TEAM\_FIELDING\_E)^-1 | 0.8503593 | -1.0 | 1.3 | 4.244388e+04 | 0 | 108.74580 | 6.500000e+01 | 9.650000e+02 | 11825.6487 | 14534.282 | 115.16612 |
| 69 | (TEAM\_BATTING\_2B\*TEAM\_PITCHING\_SO)^1 | 0.8499779 | 1.0 | 1.3 | 1.387000e-03 | 0 | 108.58309 | 6.245800e+04 | 5.055200e+05 | 11790.2870 | 14571.334 | 115.85653 |
| 99 | (TEAM\_BATTING\_HR\*TEAM\_PITCHING\_BB)^0.6 | 0.8487577 | 0.6 | 1.3 | 3.848267e-01 | 0 | 108.06654 | 1.575000e+03 | 1.878540e+05 | 11678.3764 | 14689.849 | 115.08275 |
| 113 | (TEAM\_BATTING\_BB\*TEAM\_PITCHING\_HR)^0.6 | 0.8487536 | 0.6 | 1.3 | 3.848165e-01 | 0 | 108.06484 | 1.479000e+03 | 1.879640e+05 | 11678.0086 | 14690.242 | 115.08532 |
| 153 | (TEAM\_PITCHING\_BB\*TEAM\_FIELDING\_E)^-0.9 | 0.8467487 | -0.9 | 1.3 | 7.436759e+06 | 0 | 107.22873 | 3.537200e+04 | 5.396880e+05 | 11498.0003 | 14884.981 | 117.11743 |
| 60 | (TEAM\_BATTING\_2B/TEAM\_PITCHING\_HR)^-0.7 | 0.8463835 | -0.7 | 1.3 | 4.870846e+02 | 0 | 107.07812 | 7.695652e-01 | 5.166666e+01 | 11465.7230 | 14920.447 | 116.61598 |
| 93 | (TEAM\_BATTING\_HR\*TEAM\_BATTING\_BB)^0.6 | 0.8453542 | 0.6 | 1.2 | 2.512561e-01 | 0 | 106.65626 | 1.479000e+03 | 1.812000e+05 | 11375.5567 | 6167.836 | 74.08289 |
| 104 | (TEAM\_BATTING\_HR/TEAM\_BATTING\_SO)^0.9 | 0.8448459 | 0.9 | 1.3 | 1.625278e+03 | 0 | 106.44938 | 4.000000e-03 | 3.937500e-01 | 11331.4701 | 15069.794 | 115.92568 |
| 150 | (TEAM\_PITCHING\_HR/TEAM\_PITCHING\_SO)^0.9 | 0.8447183 | 0.9 | 1.3 | 1.625107e+03 | 0 | 106.39761 | 3.687800e-03 | 3.944020e-01 | 11320.4509 | 15082.185 | 115.96680 |
| 151 | (TEAM\_PITCHING\_HR\*TEAM\_FIELDING\_DP)^0.6 | 0.8445797 | 0.6 | 1.3 | 8.209250e-01 | 0 | 106.34143 | 2.610000e+02 | 5.865300e+04 | 11308.5003 | 15095.646 | 116.96809 |
| 134 | (TEAM\_PITCHING\_H/TEAM\_BATTING\_SO)^-0.9 | 0.8440114 | -0.9 | 1.4 | 7.862272e+02 | 0 | 106.11182 | 9.849893e-01 | 7.806250e+00 | 11259.7185 | 36933.360 | 186.46465 |
| 114 | (TEAM\_BATTING\_BB/TEAM\_PITCHING\_HR)^-0.7 | 0.8435083 | -0.7 | 1.3 | 8.297678e+02 | 0 | 105.90953 | 1.286195e+00 | 1.643333e+02 | 11216.8276 | 15199.713 | 117.61545 |
| 44 | (TEAM\_BATTING\_H/TEAM\_BATTING\_SO)^-1 | 0.8432612 | -1.0 | 1.4 | 8.027682e+02 | 0 | 105.81054 | 9.849893e-01 | 6.359375e+00 | 11195.8696 | 37110.973 | 186.95987 |
| 138 | (TEAM\_PITCHING\_H/TEAM\_PITCHING\_SO)^-1 | 0.8432073 | -1.0 | 1.4 | 8.022485e+02 | 0 | 105.78896 | 9.849893e-01 | 6.356234e+00 | 11191.3046 | 37123.738 | 186.98672 |
| 131 | (TEAM\_PITCHING\_H\*TEAM\_FIELDING\_E)^-0.9 | 0.8411121 | -0.9 | 1.3 | 1.850878e+07 | 0 | 104.95847 | 9.808500e+04 | 2.563040e+06 | 11016.2808 | 15432.452 | 118.58927 |
| 7 | (TEAM\_PITCHING\_HR)^0.7 | 0.8398316 | 0.7 | 1.3 | 1.024867e+01 | 0 | 104.45849 | 3.000000e+00 | 3.430000e+02 | 10911.5759 | 15556.820 | 118.18804 |
| 65 | (TEAM\_BATTING\_2B\*TEAM\_BATTING\_SO)^1 | 0.8396773 | 1.0 | 1.3 | 1.434700e-03 | 0 | 104.39861 | 5.129100e+04 | 4.434970e+05 | 10899.0697 | 15571.809 | 119.14106 |
| 41 | (TEAM\_BATTING\_H\*TEAM\_FIELDING\_E)^-1 | 0.8385972 | -1.0 | 1.3 | 6.086968e+07 | 0 | 103.98179 | 9.808500e+04 | 1.550755e+06 | 10812.2123 | 15676.711 | 119.31786 |
| 54 | (TEAM\_BATTING\_2B/TEAM\_BATTING\_HR)^-0.7 | 0.8383408 | -0.7 | 1.3 | 4.969926e+02 | 0 | 103.88340 | 8.083333e-01 | 5.166666e+01 | 10791.7614 | 15701.616 | 119.12764 |
| 109 | (TEAM\_BATTING\_HR\*TEAM\_FIELDING\_DP)^0.6 | 0.8379481 | 0.6 | 1.3 | 8.355073e-01 | 0 | 103.73315 | 2.610000e+02 | 4.827800e+04 | 10760.5673 | 15739.758 | 119.02258 |
| 89 | (TEAM\_BATTING\_3B\*TEAM\_PITCHING\_SO)^1 | 0.8372696 | 1.0 | 1.4 | 1.063720e-02 | 0 | 103.47475 | 9.933000e+03 | 1.629440e+05 | 10707.0233 | 38529.610 | 187.53105 |
| 108 | (TEAM\_BATTING\_HR/TEAM\_PITCHING\_SO)^0.9 | 0.8362213 | 0.9 | 1.3 | 1.671042e+03 | 0 | 103.07846 | 3.687800e-03 | 3.688394e-01 | 10625.1699 | 15907.481 | 118.56475 |
| 94 | (TEAM\_BATTING\_HR/TEAM\_BATTING\_BB)^0.7 | 0.8361173 | 0.7 | 1.3 | 8.479074e+02 | 0 | 103.03935 | 6.085200e-03 | 5.366492e-01 | 10617.1073 | 15917.582 | 119.90731 |
| 142 | (TEAM\_PITCHING\_HR/TEAM\_PITCHING\_BB)^0.7 | 0.8361101 | 0.7 | 1.3 | 8.478646e+02 | 0 | 103.03663 | 5.714300e-03 | 5.370705e-01 | 10616.5474 | 15918.284 | 119.91115 |
| 38 | (TEAM\_BATTING\_H/TEAM\_PITCHING\_HR)^-0.7 | 0.8350756 | -0.7 | 1.3 | 1.676227e+03 | 0 | 102.64943 | 4.725000e+00 | 4.003333e+02 | 10536.9048 | 16018.758 | 119.97245 |
| 4 | (TEAM\_BATTING\_HR)^0.7 | 0.8319066 | 0.7 | 1.3 | 1.045178e+01 | 0 | 101.48409 | 3.000000e+00 | 2.640000e+02 | 10299.0208 | 16326.561 | 120.56806 |
| 32 | (TEAM\_BATTING\_H/TEAM\_BATTING\_HR)^-0.7 | 0.8267414 | -0.7 | 1.3 | 1.707679e+03 | 0 | 99.64913 | 5.876923e+00 | 4.003333e+02 | 9929.9499 | 16828.241 | 122.41826 |
| 128 | (TEAM\_PITCHING\_H/TEAM\_PITCHING\_HR)^-0.7 | 0.8267350 | -0.7 | 1.3 | 1.707624e+03 | 0 | 99.64690 | 5.876923e+00 | 4.266667e+02 | 9929.5041 | 16828.866 | 122.42151 |
| 100 | (TEAM\_BATTING\_HR/TEAM\_PITCHING\_BB)^0.7 | 0.8251668 | 0.7 | 1.3 | 8.601561e+02 | 0 | 99.10485 | 5.714300e-03 | 5.252525e-01 | 9821.7707 | 16981.186 | 123.13218 |
| 78 | (TEAM\_BATTING\_3B/TEAM\_PITCHING\_H)^1 | 0.8173856 | 1.0 | 1.4 | 1.205833e+04 | 0 | 96.51216 | 7.236800e-03 | 8.103640e-02 | 9314.5968 | 43237.538 | 196.94890 |
| 96 | (TEAM\_BATTING\_HR/TEAM\_PITCHING\_H)^0.7 | 0.8155937 | 0.7 | 1.3 | 1.728992e+03 | 0 | 95.93676 | 2.343700e-03 | 1.701571e-01 | 9203.8612 | 17911.001 | 125.52543 |
| 75 | (TEAM\_BATTING\_3B\*TEAM\_BATTING\_BB)^1 | 0.8070687 | 1.0 | 1.4 | 1.463130e-02 | 0 | 93.30178 | 6.523000e+03 | 8.396800e+04 | 8705.2226 | 45680.269 | 198.74486 |
| 51 | (TEAM\_BATTING\_2B\*TEAM\_BATTING\_3B)^1 | 0.7974889 | 1.0 | 1.4 | 3.090690e-02 | 0 | 90.52614 | 2.822000e+03 | 3.841000e+04 | 8194.9824 | 47948.462 | 202.32142 |
| 3 | (TEAM\_BATTING\_3B)^1 | 0.7916217 | 1.0 | 1.4 | 7.375421e+00 | 0 | 88.91370 | 1.100000e+01 | 1.470000e+02 | 7905.6467 | 49337.637 | 206.16140 |
| 91 | (TEAM\_BATTING\_3B\*TEAM\_FIELDING\_DP)^1 | 0.7909152 | 1.0 | 1.4 | 5.097390e-02 | 0 | 88.72372 | 1.416000e+03 | 2.193000e+04 | 7871.8988 | 49504.930 | 206.76649 |
| 81 | (TEAM\_BATTING\_3B\*TEAM\_PITCHING\_BB)^1 | 0.7754049 | 1.0 | 1.4 | 1.305350e-02 | 0 | 84.76180 | 6.523000e+03 | 1.062400e+05 | 7184.5620 | 53177.299 | 210.41709 |
| 30 | (TEAM\_BATTING\_H/TEAM\_BATTING\_3B)^1 | 0.7747867 | 1.0 | 1.3 | 7.290309e+00 | 0 | 84.61163 | 9.388889e+00 | 1.381818e+02 | 7159.1275 | 21874.499 | 136.89352 |
| 29 | (TEAM\_BATTING\_H\*TEAM\_BATTING\_3B)^1 | 0.7739950 | 1.0 | 1.4 | 4.846600e-03 | 0 | 84.42016 | 1.672000e+04 | 2.340250e+05 | 7126.7628 | 53511.104 | 211.30134 |
| 82 | (TEAM\_BATTING\_3B/TEAM\_PITCHING\_BB)^1 | 0.7451137 | 1.0 | 1.4 | 3.714546e+03 | 0 | 77.99630 | 1.851850e-02 | 3.674033e-01 | 6083.4235 | 60349.338 | 223.48377 |
| 143 | (TEAM\_PITCHING\_HR\*TEAM\_FIELDING\_E)^1 | 0.7427166 | 1.0 | 1.3 | 1.344020e-02 | 0 | 77.50714 | 7.440000e+02 | 2.557170e+05 | 6007.3570 | 24989.397 | 141.71903 |
| 77 | (TEAM\_BATTING\_3B\*TEAM\_PITCHING\_H)^1 | 0.7363186 | 1.0 | 1.4 | 4.255500e-03 | 0 | 76.23056 | 1.672000e+04 | 2.920320e+05 | 5811.0990 | 62431.750 | 222.82144 |
| 92 | (TEAM\_BATTING\_3B/TEAM\_FIELDING\_DP)^1 | 0.7331299 | 1.0 | 1.4 | 9.527461e+02 | 0 | 75.60953 | 7.407410e-02 | 1.320513e+00 | 5716.8013 | 63186.728 | 224.82049 |
| 52 | (TEAM\_BATTING\_2B/TEAM\_BATTING\_3B)^1 | 0.7236252 | 1.0 | 1.3 | 3.982849e+01 | 0 | 73.81484 | 1.118056e+00 | 2.754545e+01 | 5448.6309 | 26843.701 | 146.64732 |
| 76 | (TEAM\_BATTING\_3B/TEAM\_BATTING\_BB)^1 | 0.7166419 | 1.0 | 1.4 | 3.312092e+03 | 0 | 72.54697 | 1.854970e-02 | 4.897959e-01 | 5263.0627 | 67090.598 | 231.25871 |
| 171 | (TEAM\_BATTING\_SO\*TEAM\_BASERUN\_SB)^1 | 0.7132491 | 1.0 | 1.3 | 2.531500e-03 | 0 | 71.94561 | 9.126000e+03 | 4.129200e+05 | 5176.1705 | 27851.511 | 147.12809 |
| 67 | (TEAM\_BATTING\_2B\*TEAM\_BASERUN\_SB)^1 | 0.7107128 | 1.0 | 1.4 | 1.211750e-02 | 0 | 71.50204 | 3.348000e+03 | 1.515000e+05 | 5112.5421 | 68494.428 | 228.99389 |
| 105 | (TEAM\_BATTING\_HR\*TEAM\_BASERUN\_SB)^1 | 0.7032535 | 1.0 | 1.3 | 2.043260e-02 | 0 | 70.22620 | 5.500000e+02 | 5.467700e+04 | 4931.7198 | 28822.364 | 146.00321 |
| 147 | (TEAM\_PITCHING\_HR\*TEAM\_BASERUN\_SB)^1 | 0.6929458 | 1.0 | 1.3 | 1.910500e-02 | 0 | 68.52960 | 5.750000e+02 | 7.307300e+04 | 4696.3059 | 29823.526 | 147.41859 |
| 177 | (TEAM\_BASERUN\_SB\*TEAM\_PITCHING\_SO)^1 | 0.6835918 | 1.0 | 1.4 | 3.549500e-03 | 0 | 67.05182 | 9.594000e+03 | 4.882500e+05 | 4495.9471 | 74915.845 | 238.06371 |
| 121 | (TEAM\_BATTING\_BB\*TEAM\_BASERUN\_SB)^1 | 0.6781572 | 1.0 | 1.4 | 5.298700e-03 | 0 | 66.21850 | 7.854000e+03 | 3.081320e+05 | 4384.8891 | 76202.598 | 236.31094 |
| 179 | (TEAM\_BASERUN\_SB\*TEAM\_FIELDING\_DP)^1 | 0.6756041 | 1.0 | 1.4 | 1.918770e-02 | 0 | 65.83314 | 2.214000e+03 | 9.416800e+04 | 4334.0017 | 76807.081 | 239.30078 |
| 11 | (TEAM\_BASERUN\_SB)^1 | 0.6706848 | 1.0 | 1.4 | 2.705475e+00 | 0 | 65.10125 | 1.800000e+01 | 6.320000e+02 | 4238.1732 | 77971.839 | 239.82482 |
| 45 | (TEAM\_BATTING\_H\*TEAM\_BASERUN\_SB)^1 | 0.6639580 | 1.0 | 1.4 | 1.819600e-03 | 0 | 64.12237 | 2.253600e+04 | 9.870000e+05 | 4111.6779 | 79564.538 | 239.69915 |
| 136 | (TEAM\_PITCHING\_H/TEAM\_BASERUN\_SB)^1 | 0.6570397 | 1.0 | 1.3 | 1.148516e+01 | 0 | 63.14077 | 2.927215e+00 | 9.305555e+01 | 3986.7575 | 33311.015 | 157.15382 |
| 46 | (TEAM\_BATTING\_H/TEAM\_BASERUN\_SB)^1 | 0.6537578 | 1.0 | 1.3 | 1.196969e+01 | 0 | 62.68368 | 2.439873e+00 | 8.844444e+01 | 3929.2438 | 33629.777 | 157.57446 |
| 68 | (TEAM\_BATTING\_2B/TEAM\_BASERUN\_SB)^1 | 0.6501236 | 1.0 | 1.3 | 7.079655e+01 | 0 | 62.18372 | 2.821869e-01 | 1.644444e+01 | 3866.8146 | 33982.761 | 157.63468 |
| 90 | (TEAM\_BATTING\_3B/TEAM\_PITCHING\_SO)^1 | 0.6398063 | 1.0 | 1.4 | 4.009433e+03 | 0 | 60.79841 | 1.006710e-02 | 3.393782e-01 | 3696.4471 | 85282.919 | 244.78166 |
| 166 | (TEAM\_FIELDING\_E/TEAM\_BASERUN\_SB)^1 | 0.6356577 | 1.0 | 1.3 | 1.017607e+02 | 0 | 60.25498 | 3.439490e-01 | 9.952381e+00 | 3630.6623 | 35387.800 | 162.04831 |
| 158 | (TEAM\_PITCHING\_BB/TEAM\_BASERUN\_SB)^1 | 0.6351494 | 1.0 | 1.3 | 3.054407e+01 | 0 | 60.18890 | 8.095238e-01 | 3.500000e+01 | 3622.7039 | 35437.176 | 159.08545 |
| 122 | (TEAM\_BATTING\_BB/TEAM\_BASERUN\_SB)^1 | 0.6318856 | 1.0 | 1.3 | 3.172413e+01 | 0 | 59.76733 | 6.738095e-01 | 3.327778e+01 | 3572.1332 | 35754.183 | 159.41371 |
| 157 | (TEAM\_PITCHING\_BB\*TEAM\_BASERUN\_SB)^1 | 0.6311528 | 1.0 | 1.4 | 4.560600e-03 | 0 | 59.67330 | 8.316000e+03 | 3.779270e+05 | 3560.9028 | 87331.804 | 244.71724 |
| 180 | (TEAM\_BASERUN\_SB/TEAM\_FIELDING\_DP)^1 | 0.6237073 | 1.0 | 1.4 | 3.449580e+02 | 0 | 58.73048 | 1.052632e-01 | 4.241611e+00 | 3449.2689 | 89094.683 | 247.40375 |
| 88 | (TEAM\_BATTING\_3B/TEAM\_BASERUN\_SB)^1 | 0.6210397 | 1.0 | 1.3 | 3.409340e+02 | 0 | 58.39813 | 8.163270e-02 | 3.166667e+00 | 3410.3411 | 36807.614 | 160.13578 |
| 172 | (TEAM\_BATTING\_SO/TEAM\_BASERUN\_SB)^1 | 0.6201045 | 1.0 | 1.4 | 3.417356e+01 | 0 | 58.28226 | 6.380000e-01 | 5.094444e+01 | 3396.8216 | 89947.719 | 255.35360 |
| 86 | (TEAM\_BATTING\_3B/TEAM\_BATTING\_SO)^1 | 0.6126631 | 1.0 | 1.4 | 3.532960e+03 | 0 | 57.37233 | 1.006710e-02 | 4.225806e-01 | 3291.5844 | 91709.598 | 248.66010 |
| 135 | (TEAM\_PITCHING\_H\*TEAM\_BASERUN\_SB)^1 | 0.6123523 | 1.0 | 1.4 | 1.543000e-03 | 0 | 57.33478 | 2.340000e+04 | 1.230000e+06 | 3287.2770 | 91783.184 | 247.95116 |
| 178 | (TEAM\_BASERUN\_SB/TEAM\_PITCHING\_SO)^1 | 0.5677306 | 1.0 | 1.4 | 1.636999e+03 | 0 | 52.27933 | 1.962920e-02 | 1.256281e+00 | 2733.1279 | 102348.247 | 253.49058 |
| 97 | (TEAM\_BATTING\_HR\*TEAM\_PITCHING\_HR)^1 | 0.5628724 | 1.0 | 1.3 | 1.157690e-02 | 0 | 51.76508 | 9.000000e+00 | 8.197700e+04 | 2679.6239 | 42457.284 | 159.15999 |
| 148 | (TEAM\_PITCHING\_HR/TEAM\_BASERUN\_SB)^1 | 0.5460020 | 1.0 | 1.3 | 1.183565e+02 | 0 | 50.02720 | 1.345290e-02 | 1.033333e+01 | 2502.7211 | 44095.868 | 163.48350 |
| 106 | (TEAM\_BATTING\_HR/TEAM\_BASERUN\_SB)^1 | 0.5425575 | 1.0 | 1.3 | 1.216254e+02 | 0 | 49.68104 | 1.345290e-02 | 1.033333e+01 | 2468.2054 | 44430.431 | 163.63626 |
| 83 | (TEAM\_BATTING\_3B\*TEAM\_FIELDING\_E)^1 | 0.4253998 | 1.0 | 1.3 | 1.069850e-02 | 0 | 39.25110 | 8.910000e+02 | 9.360000e+04 | 1540.6487 | 55809.709 | 165.42409 |
| 87 | (TEAM\_BATTING\_3B\*TEAM\_BASERUN\_SB)^1 | 0.4215036 | 1.0 | 1.4 | 2.715110e-02 | 0 | 38.93914 | 5.220000e+02 | 7.710400e+04 | 1516.2568 | 136970.360 | 254.11218 |
| 80 | (TEAM\_BATTING\_3B/TEAM\_PITCHING\_HR)^1 | 0.3246277 | 1.0 | 1.3 | 1.009383e+02 | 0 | 31.62694 | 5.140190e-02 | 1.850000e+01 | 1000.2633 | 65597.497 | 160.67193 |
| 74 | (TEAM\_BATTING\_3B/TEAM\_BATTING\_HR)^1 | 0.3218945 | 1.0 | 1.3 | 9.105530e+01 | 0 | 31.42999 | 5.140190e-02 | 1.850000e+01 | 987.8442 | 65862.959 | 160.37749 |

### Models Based on Calculated Measures - Model 4

########  
# select columns for model  
  
head(the\_tamale[order(-the\_tamale$r2),c("try","r2","coef\_p")],5)

## try r2 coef\_p  
## 21 log(TEAM\_PITCHING\_BB)^1.9 0.9549735 0  
## 18 log(TEAM\_BATTING\_BB)^2 0.9548636 0  
## 33 (TEAM\_BATTING\_H\*TEAM\_BATTING\_BB)^0.7 0.9547029 0  
## 27 (TEAM\_BATTING\_H\*TEAM\_BATTING\_2B)^0.4 0.9544827 0  
## 61 (TEAM\_BATTING\_2B\*TEAM\_PITCHING\_BB)^0.4 0.9542975 0

model\_data <- data.frame(LOG\_TEAM\_PITCHING\_BB = with(valid\_data,log(TEAM\_PITCHING\_BB)),  
LOG\_TEAM\_BATTING\_BB = with(valid\_data,log(TEAM\_BATTING\_BB)),  
TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_BB = with(valid\_data,(TEAM\_BATTING\_H\*TEAM\_BATTING\_BB)),  
TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_2B = with(valid\_data,(TEAM\_BATTING\_H\*TEAM\_BATTING\_2B)),  
TEAM\_BATTING\_2B\_BY\_TEAM\_PITCHING\_BB = with(valid\_data,(TEAM\_BATTING\_2B\*TEAM\_PITCHING\_BB)),  
 TARGET\_WINS=with(valid\_data,TARGET\_WINS)  
)  
  
d<-boxcox(TARGET\_WINS+0.0001~.,data=model\_data, plotit = TRUE)



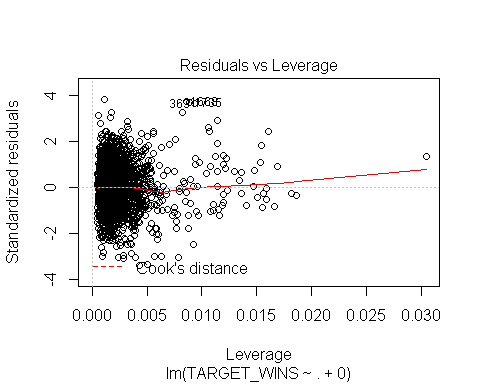
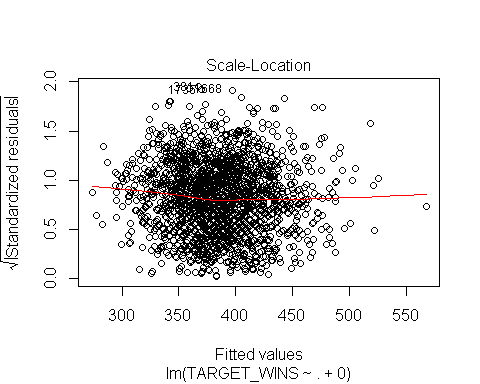
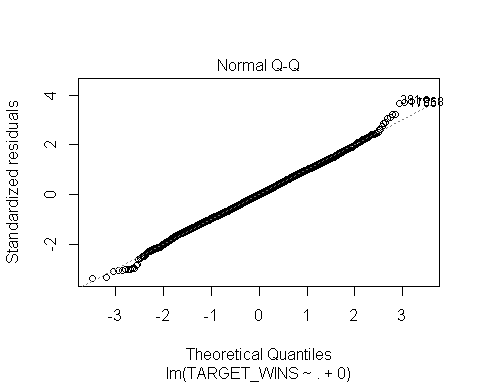
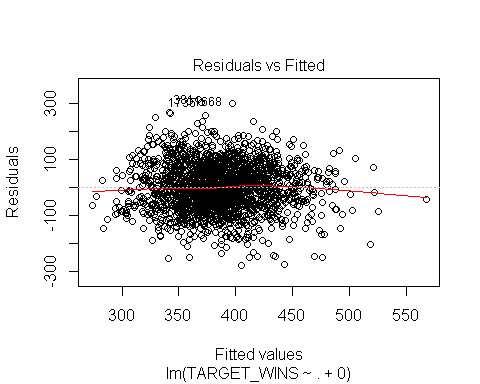
d <- data.frame(d)  
  
y\_power <- d[d$y == max(d$y),]$x  
  
source\_cols <- colnames(model\_data)  
  
source\_cols <- source\_cols[source\_cols!="TARGET\_WINS"]  
  
powers <- lapply(source\_cols,function(x){getBoxCoxCoef(model\_data[,x],model\_data$TARGET\_WINS^y\_power)})  
  
params <- c(unlist(powers),y\_power)  
kable(data.frame(variable=c(source\_cols,"TARGET\_WINS"),power=params))

|  |  |
| --- | --- |
| variable | power |
| LOG\_TEAM\_PITCHING\_BB | 1.900000 |
| LOG\_TEAM\_BATTING\_BB | 2.000000 |
| TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_BB | 0.700000 |
| TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_2B | 0.400000 |
| TEAM\_BATTING\_2B\_BY\_TEAM\_PITCHING\_BB | 0.400000 |
| TARGET\_WINS | 1.353535 |

power\_transforms <- data.frame(t(apply(model\_data,1,function(c) { c^params})))  
  
  
#model <- lm(TARGET\_WINS~.,data=power\_transforms[,colnames(power\_transforms)[!colnames(power\_transforms) %in% c("TEAM\_PITCHING\_H","TEAM\_BATTING\_H")]])  
#summary(model)  
model <- lm(TARGET\_WINS~.+0,power\_transforms[,colnames(power\_transforms)[!colnames(power\_transforms) %in% c("TEAM\_BATTING\_3B\_BY\_TEAM\_PITCHING\_HR","TEAM\_BATTING\_BB\_BY\_TEAM\_PITCHING\_BB")]])  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ . + 0, data = power\_transforms[, colnames(power\_transforms)[!colnames(power\_transforms) %in%   
## c("TEAM\_BATTING\_3B\_BY\_TEAM\_PITCHING\_HR", "TEAM\_BATTING\_BB\_BY\_TEAM\_PITCHING\_BB")]])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -277.599 -56.184 -0.364 55.354 314.072   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## LOG\_TEAM\_PITCHING\_BB 16.635307 2.417770 6.880 7.87e-12  
## LOG\_TEAM\_BATTING\_BB -11.954161 2.143930 -5.576 2.78e-08  
## TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_BB 0.030138 0.002491 12.097 < 2e-16  
## TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_2B 2.112645 0.184609 11.444 < 2e-16  
## TEAM\_BATTING\_2B\_BY\_TEAM\_PITCHING\_BB -3.939456 0.454639 -8.665 < 2e-16  
##   
## LOG\_TEAM\_PITCHING\_BB \*\*\*  
## LOG\_TEAM\_BATTING\_BB \*\*\*  
## TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_BB \*\*\*  
## TEAM\_BATTING\_H\_BY\_TEAM\_BATTING\_2B \*\*\*  
## TEAM\_BATTING\_2B\_BY\_TEAM\_PITCHING\_BB \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 82.3 on 2077 degrees of freedom  
## Multiple R-squared: 0.9568, Adjusted R-squared: 0.9567   
## F-statistic: 9204 on 5 and 2077 DF, p-value: < 2.2e-16

model4 <- model  
model4\_y\_power <- y\_power  
model4\_powers <- powers  
  
plot(model)



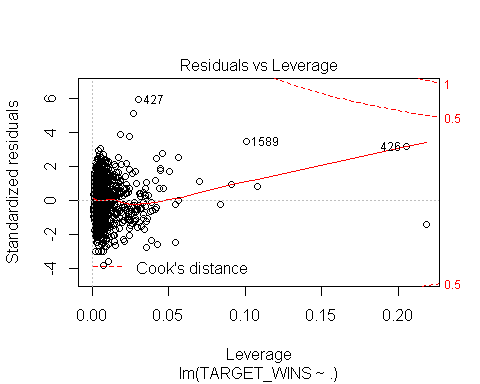
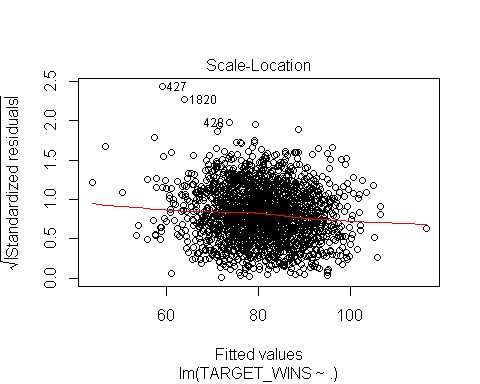
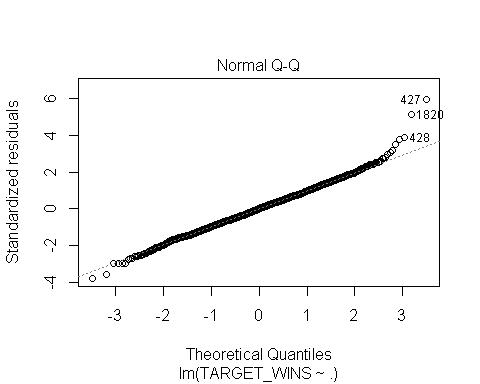
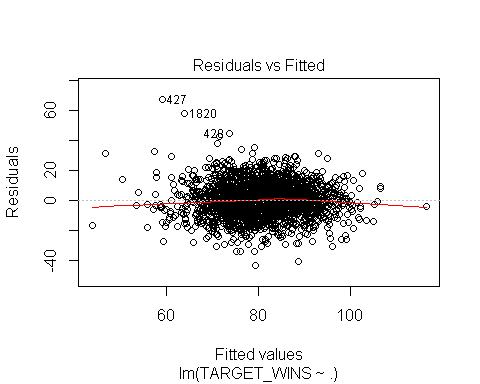
# Model 5: A Straight Up Model, w/o intercept calcellation

Let's try a straight up regression whatever linear regression

model <- lm(TARGET\_WINS~.,data=valid\_data)  
summary(model)

##   
## Call:  
## lm(formula = TARGET\_WINS ~ ., data = valid\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -43.316 -7.494 0.150 7.530 66.987   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 39.906223 5.842738 6.830 1.11e-11 \*\*\*  
## TEAM\_BATTING\_H -0.068565 0.011620 -5.901 4.22e-09 \*\*\*  
## TEAM\_BATTING\_2B -0.037171 0.009140 -4.067 4.95e-05 \*\*\*  
## TEAM\_BATTING\_3B 0.172414 0.018602 9.268 < 2e-16 \*\*\*  
## TEAM\_BATTING\_HR 0.475803 0.075484 6.303 3.55e-10 \*\*\*  
## TEAM\_BATTING\_BB 0.379235 0.034381 11.030 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_H 0.093178 0.009827 9.482 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_HR -0.370954 0.071460 -5.191 2.30e-07 \*\*\*  
## TEAM\_PITCHING\_BB -0.326856 0.032029 -10.205 < 2e-16 \*\*\*  
## TEAM\_FIELDING\_E -0.068667 0.004248 -16.164 < 2e-16 \*\*\*  
## TEAM\_BATTING\_SO -0.114314 0.016705 -6.843 1.02e-11 \*\*\*  
## TEAM\_BASERUN\_SB 0.062922 0.004691 13.414 < 2e-16 \*\*\*  
## TEAM\_PITCHING\_SO 0.096919 0.015632 6.200 6.80e-10 \*\*\*  
## TEAM\_FIELDING\_DP -0.101177 0.012194 -8.298 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.47 on 2068 degrees of freedom  
## Multiple R-squared: 0.3505, Adjusted R-squared: 0.3464   
## F-statistic: 85.85 on 13 and 2068 DF, p-value: < 2.2e-16

model5 <- model  
model5\_y\_power <- 1  
model5\_powers <- powers  
  
plot(model)



## 3. BUILD MODELS (25 Points)

We built/evaluated the following models:

Using the training data set, build at least three different multiple linear regression models, using different variables (or the same variables with different transformations). 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.). Since you manually selected a variable for inclusion into the model or exclusion into the model, indicate why this was done.

Discuss the coefficients in the models, do they make sense? For example, if a team hits a lot of Home Runs, it would be reasonably expected that such a team would win more games. However, if the coefficient is negative (suggesting that the team would lose more games), then that needs to be discussed. Are you keeping the model even though it is counter intuitive? Why? The boss needs to know.

## 4. SELECT MODELS (25 Points)

Decide on the criteria for selecting the best multiple linear regression model. Will you select a model with slightly worse performance if it makes more sense or is more parsimonious? Discuss why you selected your model. For the multiple linear regression model, will you use a metric such as Adjusted R2, RMSE, etc.? Be sure to explain how you can make inferences from the model, discuss multi-collinearity issues (if any), and discuss other relevant model output. Using the training data set, evaluate the multiple linear regression model based on

1. mean squared error
2. R2,
3. F-statistic,
4. residual plots.

Make predictions using the training\_transformed data set.

models <- list(model0,model1,model2,model3,model4,model5)  
  
model\_metrics <- lapply(models,function (m){   
 s <- summary(m)  
 res <- data.frame(r2=s$r.squared,  
 rse=s$sigma,  
 me=sqrt(mean(s$residuals^2)),  
 fstatistic=head(c(s$fstatistic[1],1),1),  
 fstatistic\_df1=head(c(s$fstatistic[2],1),1),  
 fstatistic\_df2=head(c(s$fstatistic[3],nrow(valid\_data)),1)  
 )  
 res  
 })  
  
combined\_results <- do.call(rbind,model\_metrics)  
rownames(combined\_results) <- NULL  
combined\_results <- data.frame(  
 model = c("model1","model3","model3-1","model3-2","model4","model2"),  
 description = c("Mean-based model","Model with Straight-Up Variables, Zero Intercept","Model with Y Power Transformation Using Boxcox","Model with X&Y Power Transformation Using Boxcox", "Model with Y & Automatic Feature Set 1 w/Power Transformation Using Boxcox", "Model with Straight-Up Variables"),  
 combined\_results)  
  
y\_powers<- c(model0\_y\_power,model1\_y\_power,model2\_y\_power,model3\_y\_power,model4\_y\_power,model5\_y\_power)  
combined\_results$me <- combined\_results$me ^ (1/y\_powers)  
kable(combined\_results)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| model | description | r2 | rse | me | fstatistic | fstatistic\_df1 | fstatistic\_df2 |
| model1 | Mean-based model | 0.0000000 | 14.18806 | 14.18465 | 1.0000 | 1 | 2082 |
| model3 | Model with Straight-Up Variables, Zero Intercept | 0.9801910 | 11.59599 | 11.55973 | 7875.2786 | 13 | 2069 |
| model3-1 | Model with Y Power Transformation Using Boxcox | 0.9726027 | 33.15973 | 18.45196 | 5649.9620 | 13 | 2069 |
| model3-2 | Model with X&Y Power Transformation Using Boxcox | 0.9667822 | 36.45101 | 19.99413 | 10070.0951 | 6 | 2076 |
| model4 | Model with Y & Automatic Feature Set 1 w/Power Transformation Using Boxcox | 0.9568147 | 82.30408 | 25.98597 | 9203.6234 | 5 | 2077 |
| model2 | Model with Straight-Up Variables | 0.3505140 | 11.47014 | 11.43151 | 85.8505 | 13 | 2068 |

combined\_results <-combined\_results[order(combined\_results$model),]  
kable(combined\_results[combined\_results$model %in% c("model1","model2","model3","model4"),])

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | model | description | r2 | rse | me | fstatistic | fstatistic\_df1 | fstatistic\_df2 |
| 1 | model1 | Mean-based model | 0.0000000 | 14.18806 | 14.18465 | 1.0000 | 1 | 2082 |
| 6 | model2 | Model with Straight-Up Variables | 0.3505140 | 11.47014 | 11.43151 | 85.8505 | 13 | 2068 |
| 2 | model3 | Model with Straight-Up Variables, Zero Intercept | 0.9801910 | 11.59599 | 11.55973 | 7875.2786 | 13 | 2069 |
| 5 | model4 | Model with Y & Automatic Feature Set 1 w/Power Transformation Using Boxcox | 0.9568147 | 82.30408 | 25.98597 | 9203.6234 | 5 | 2077 |

# Predictions

# make up for missing values in the evaluation dataset by imputing wiht the training medians  
  
# make up for missing TEAM\_BATTING\_SO  
evaluation[is.na(evaluation$TEAM\_BATTING\_SO),]$TEAM\_BATTING\_SO <- median(training$TEAM\_BATTING\_SO, na.rm = TRUE)  
  
# make up for missing TEAM\_BASERUN\_SB  
evaluation[is.na(evaluation$TEAM\_BASERUN\_SB),]$TEAM\_BASERUN\_SB <- median(training$TEAM\_BASERUN\_SB, na.rm = TRUE)  
  
# make up for missing TEAM\_PITCHING\_SO  
evaluation[is.na(evaluation$TEAM\_PITCHING\_SO),]$TEAM\_PITCHING\_SO <- median(training$TEAM\_PITCHING\_SO, na.rm = TRUE)  
  
# make up for missing TEAM\_FIELDING\_DP  
evaluation[is.na(evaluation$TEAM\_FIELDING\_DP),]$TEAM\_FIELDING\_DP <- median(training$TEAM\_FIELDING\_DP, na.rm = TRUE)  
  
  
# apply validation rules  
  
evaluation$INVALID <- FALSE  
evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_2B < 110 | evaluation$TEAM\_BATTING\_2B>376  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 2

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_3B < 11 | evaluation$TEAM\_BATTING\_3B>153  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 3

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_HR < 3 | evaluation$TEAM\_BATTING\_HR>264  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 3

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_H < 935 | evaluation$TEAM\_BATTING\_H>2131  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 4

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_SO < 308 | evaluation$TEAM\_BATTING\_SO>1535  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 10

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_PITCHING\_SO < 308 | evaluation$TEAM\_PITCHING\_SO>1535  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 10

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_BATTING\_BB < 282 | evaluation$TEAM\_BATTING\_BB>835  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 16

evaluation$INVALID <- evaluation$INVALID | evaluation$TEAM\_PITCHING\_BB < 282 | evaluation$TEAM\_PITCHING\_BB>835  
length(evaluation$INVALID[evaluation$INVALID])

## [1] 20

model1\_predictions <- predict(model1,evaluation)^(1/model1\_y\_power)  
model1\_predictions[evaluation$INVALID] <- NA  
  
model2\_predictions <- predict(model2,evaluation)^(1/model2\_y\_power)  
model2\_predictions[evaluation$INVALID] <- NA  
  
  
kable(data.frame(model1\_predictions = model1\_predictions))

## model1\_predictions

60.85668  
 66.60762  
 73.68598  
 82.80724  
 NA  
 NA  
 73.77146  
 74.94659  
 72.15503  
 68.94785  
 65.95122  
 79.61884  
 80.62423  
 77.27399  
 81.45970  
 73.89162  
 70.67914  
 78.44014  
 67.00079  
 86.68226  
 82.60369  
 82.87553  
 81.45499  
 68.21000  
 76.20223  
 83.07728  
 NA  
 66.31804  
 81.90593  
 65.89571  
 91.66097  
 86.86108  
 86.56662  
 87.86919  
 80.25817  
 85.92277  
 74.97824  
 91.95978  
 NA  
 88.77946  
 79.25278  
 92.32462  
 NA  
 NA  
 89.47400  
 84.63332  
 96.31432  
 71.77025  
 65.48920  
 74.96805  
 70.61387  
 81.13160  
 75.79310  
 69.51143  
 70.51507  
 74.47321  
 92.40763  
 72.76270  
 63.45855  
 82.16635  
 88.57546  
 84.90677  
 88.18881  
 82.34184  
 80.60836  
 NA  
 74.61024  
 82.95597  
 83.49474  
 98.16283  
 89.32630  
 69.36656  
 79.26399  
 88.37175  
 71.45491  
 83.22354  
 88.65491  
 80.54019  
 70.41694  
 76.95837  
 78.54072  
 84.94984  
 94.18019  
 73.88323  
 88.45191  
 75.44434  
 81.87707  
 84.94723  
 94.78531  
 92.55442  
 75.40361  
 NA  
 67.60360  
 84.80045  
 85.24743  
 86.69439  
 98.19906  
 105.84229  
 90.64975  
 92.55917  
 79.17806  
 69.15397  
 81.27237  
 83.80423  
 78.53226  
 NA  
 64.48545  
 72.52681  
 88.00314  
 58.36451  
 88.45862  
 84.84879  
 95.66201  
 87.99926  
 75.86895  
 75.45672  
 88.30109  
 80.35430  
 67.34131  
 67.16138  
 NA  
 65.50159  
 65.96929  
 63.99127  
 69.98507  
 81.96967  
 88.10467  
 73.35652  
 89.40027  
 90.20046  
 83.67803  
 78.10502  
 73.94994  
 89.09536  
 89.97603  
 NA  
 75.30262  
 71.86075  
 88.90790  
 81.84130  
 59.17268  
 73.58108  
 93.63329  
 71.86006  
 73.22482  
 71.06135  
 77.46782  
 82.80107  
 80.11898  
 79.60650  
 81.57136  
 79.07958  
 NA  
 60.87316  
 78.63707  
 69.12305  
 88.96947  
 72.81828  
 96.75684  
 72.86030  
 110.02899  
 NA  
 94.04398  
 108.72433  
 101.76542  
 93.32500  
 86.26306  
 81.48590  
 69.11451  
 78.17223  
 98.86393  
 93.52629  
 78.07919  
 93.00183  
 77.60508  
 76.66838  
 80.87211  
 69.84972  
 72.72402  
 84.35035  
 94.32573  
 90.26498  
 87.67509  
 86.43137  
 NA  
 94.62872  
 76.01496  
 NA  
 55.88220  
 108.01302  
 64.40556  
 83.71560  
 66.96904  
 70.76306  
 72.06455  
 60.24939  
 71.55934  
 88.01133  
 81.14871  
 87.91060  
 76.19451  
 80.76934  
 76.18787  
 NA  
 78.82828  
 89.95658  
 78.93555  
 75.33750  
 73.48528  
 65.09778  
 NA  
 93.30727  
 87.87688  
 62.56424  
 67.43990  
 84.37875  
 79.89277  
 93.82903  
 72.66308  
 80.97510  
 78.30875  
 67.90866  
 80.00406  
 69.31442  
 NA  
 78.43955  
 77.79125  
 82.55309  
 84.65543  
 76.90449  
 82.99904  
 95.76164  
 83.32970  
 87.84930  
 80.76349  
 75.20516  
 81.41895  
 77.77083  
 84.75150  
 76.36607  
 88.48320  
 86.50174  
 82.09133  
 78.01211  
 61.81926  
 82.83755  
 74.63854  
 81.94208  
 68.53639  
 83.01730  
 79.35538  
 NA  
 86.10966  
 NA  
 68.22252  
 83.71438  
 79.40289  
 84.17469  
 67.71176