# Project 1

### Contents

Data Exploration	1
Data Preparation	1
Build Model  Model 1	9
R Code	6

## **Data Exploration**

## Data Preparation

##		variable	q_zeros	p_zeros	q_na	p_na	$q_{inf}$	p_inf	type	unique
##	1	INDEX	0	0.00	0	0.00	0	0	${\tt integer}$	259
##	2	TEAM_BATTING_H	0	0.00	0	0.00	0	0	${\tt integer}$	199
##	3	TEAM_BATTING_2B	0	0.00	0	0.00	0	0	${\tt integer}$	137
##	4	TEAM_BATTING_3B	0	0.00	0	0.00	0	0	${\tt integer}$	91
##	5	TEAM_BATTING_HR	1	0.39	0	0.00	0	0	${\tt integer}$	141
	6	TEAM_BATTING_BB	0	0.00	0	0.00	0	0	${\tt integer}$	185
##	7	TEAM_BATTING_SO	2	0.77	18	6.95	0	0	integer	206
##	8	TEAM_BASERUN_SB	1	0.39	13	5.02	0	0	integer	153
##	9	TEAM_BASERUN_CS	1	0.39	87	33.59	0	0	integer	68
##	10	TEAM_BATTING_HBP	0	0.00	240	92.66	0	0	integer	17
##	11	TEAM_PITCHING_H	0	0.00	0	0.00	0	0	integer	210
##		TEAM_PITCHING_HR	1	0.39	0	0.00	0	0	integer	144
##		TEAM_PITCHING_BB	0	0.00	0	0.00	0	0	integer	199
##		TEAM_PITCHING_SO	2	0.77	18	6.95	0	0	integer	205
##	15	TEAM_FIELDING_E	0	0.00	0	0.00	0	0	integer	168
##	16	TEAM_FIELDING_DP	0	0.00	31	11.97	0	0	integer	90
##		variable	q_zeros	p_zeros	q_na	p_na	q_inf	p_inf	type	unique
	7	variable TEAM_BATTING_SO	q_zeros 2	p_zeros 0.77	q_na 18	p_na 6.95	$\begin{array}{c} q\_{\text{inf}} \\ 0 \end{array}$	<pre>p_inf 0</pre>	type integer	unique 206
			_	_	_	_	_	0		_
##		TEAM_BATTING_SO	2	0.77	18	6.95	0	0	integer	206
## ## ##	14	TEAM_BATTING_SO TEAM_PITCHING_SO	2	0.77	18 18	6.95 6.95	0	0	integer integer	206 205
## ## ##	14 5	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR	2 2 1	0.77 0.77 0.39	18 18 0	6.95 6.95 0.00 5.02	0 0	0 0	integer integer integer	206 205 141
## ## ## ##	14 5 8	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_SB TEAM_BASERUN_CS	2 2 1 1	0.77 0.77 0.39 0.39	18 18 0 13	6.95 6.95 0.00 5.02	0 0 0	0 0 0	integer integer integer integer	206 205 141 153
## ## ## ##	14 5 8 9	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_SB TEAM_BASERUN_CS	2 2 1 1 1	0.77 0.77 0.39 0.39 0.39	18 18 0 13 87	6.95 6.95 0.00 5.02 33.59	0 0 0 0	0 0 0 0	integer integer integer integer integer	206 205 141 153 68
## ## ## ## ##	14 5 8 9 12	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H	2 2 1 1 1	0.77 0.77 0.39 0.39 0.39 0.39	18 18 0 13 87 0	6.95 6.95 0.00 5.02 33.59 0.00	0 0 0 0 0	0 0 0 0 0 0 0	integer integer integer integer integer integer	206 205 141 153 68 144 259 199
## ## ## ## ## ##	14 5 8 9 12	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_ZB	2 2 1 1 1 1 0	0.77 0.77 0.39 0.39 0.39 0.39	18 18 0 13 87 0 0 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137
## ## ## ## ## ##	14 5 8 9 12 1 2 3 4	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B	2 2 1 1 1 1 0 0 0	0.77 0.77 0.39 0.39 0.39 0.00 0.00 0.00	18 18 0 13 87 0 0 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91
## ## ## ## ## ##	14 5 8 9 12 1 2	TEAM_BATTING_SO TEAM_PITCHING_HR TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_B TEAM_BATTING_B TEAM_BATTING_BB	2 2 1 1 1 1 0 0	0.77 0.77 0.39 0.39 0.39 0.39 0.00 0.00	18 18 0 13 87 0 0 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00	0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91
## ## ## ## ## ## ##	14 5 8 9 12 1 2 3 4 6 10	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B TEAM_BATTING_BB TEAM_BATTING_HBP	2 2 1 1 1 1 0 0 0 0 0	0.77 0.77 0.39 0.39 0.39 0.00 0.00 0.00 0.00	18 18 0 13 87 0 0 0 0 0 0 240	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00 0.00 92.66	0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91 185
## ## ## ## ## ## ## ## ## ## ## ## ##	14 5 8 9 12 1 2 3 4 6 10	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_B TEAM_BATTING_3B TEAM_BATTING_BB TEAM_BATTING_HBP TEAM_PITCHING_H	2 2 1 1 1 1 0 0 0 0 0 0	0.77 0.77 0.39 0.39 0.39 0.00 0.00 0.00 0.00 0.00	18 18 0 13 87 0 0 0 0 0 240 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00 92.66 0.00	0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91 185 17 210
######################################	14 5 8 9 12 1 2 3 4 6 10 11	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_B TEAM_BATTING_3B TEAM_BATTING_BB TEAM_BATTING_BB TEAM_BATTING_HBP TEAM_PITCHING_H	2 2 1 1 1 1 0 0 0 0 0 0 0 0	0.77 0.77 0.39 0.39 0.39 0.00 0.00 0.00 0.00 0.00	18 18 0 13 87 0 0 0 0 0 240 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00 92.66 0.00	0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91 185 17 210 199
## ## ## ## ## ## ## ## ## ## ## ## ##	14 5 8 9 12 1 2 3 4 6 10 11 13 15	TEAM_BATTING_SO TEAM_PITCHING_SO TEAM_BATTING_HR TEAM_BASERUN_CS TEAM_PITCHING_HR INDEX TEAM_BATTING_H TEAM_BATTING_B TEAM_BATTING_3B TEAM_BATTING_BB TEAM_BATTING_HBP TEAM_PITCHING_H	2 2 1 1 1 1 0 0 0 0 0 0	0.77 0.77 0.39 0.39 0.39 0.00 0.00 0.00 0.00 0.00	18 18 0 13 87 0 0 0 0 0 240 0	6.95 6.95 0.00 5.02 33.59 0.00 0.00 0.00 0.00 92.66 0.00	0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0	integer integer integer integer integer integer integer integer integer integer integer integer	206 205 141 153 68 144 259 199 137 91 185 17 210

```
##
               variable q_zeros p_zeros q_na p_na q_inf p_inf
                                                                       type unique
## 1
                                                 0.00
                  INDEX
                               0
                                     0.00
                                              0
                                                                               2276
                                                           0
                                                                  0 integer
## 2
            TARGET WINS
                                1
                                     0.04
                                                 0.00
                                                                  0 integer
                                                                                108
## 3
        TEAM_BATTING_H
                                0
                                     0.00
                                                 0.00
                                                                                569
                                                           0
                                                                  0 integer
## 4
       TEAM_BATTING_2B
                               0
                                     0.00
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                240
                               2
## 5
       TEAM BATTING 3B
                                     0.09
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                144
## 6
       TEAM BATTING HR
                              15
                                     0.66
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                243
## 7
       TEAM_BATTING_BB
                               1
                                     0.04
                                              0
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                533
## 8
       TEAM BATTING SO
                              20
                                     0.88
                                           102
                                                 4.48
                                                           0
                                                                                822
                                                                  0 integer
## 9
       TEAM_BASERUN_SB
                               2
                                     0.09
                                           131
                                                 5.76
                                                           0
                                                                  0 integer
                                                                                348
## 10
       TEAM_BASERUN_CS
                               1
                                     0.04
                                           772 33.92
                                                           0
                                                                  0 integer
                                                                                128
                                                                                 55
##
  11 TEAM_BATTING_HBP
                               0
                                     0.00 2085 91.61
                                                           0
                                                                  0 integer
       TEAM_PITCHING_H
                               0
                                     0.00
                                              0
                                                 0.00
                                                           0
                                                                                843
  12
                                                                  0 integer
                                                                  0 integer
   13 TEAM_PITCHING_HR
                              15
                                     0.66
                                              0
                                                 0.00
                                                           0
                                                                                256
## 14 TEAM_PITCHING_BB
                               1
                                     0.04
                                              0
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                535
## 15 TEAM_PITCHING_SO
                              20
                                     0.88
                                           102
                                                 4.48
                                                           0
                                                                  0 integer
                                                                                823
       TEAM_FIELDING_E
                                0
                                     0.00
                                              0
                                                 0.00
                                                           0
                                                                                549
                                                                  0 integer
## 17 TEAM_FIELDING_DP
                                     0.00
                                           286 12.57
                                                                                144
                                                                  0 integer
##
               variable q_zeros p_zeros q_na p_na q_inf p_inf
                                                                       type unique
## 11 TEAM BATTING HBP
                               0
                                     0.00 2085 91.61
                                                           0
                                                                                 55
                                                                  0 integer
       TEAM_BASERUN_CS
                                                                                128
                                1
                                     0.04
                                          772 33.92
                                                           0
                                                                  0 integer
  17 TEAM_FIELDING_DP
                               0
                                     0.00
                                           286 12.57
                                                                                144
                                                           0
                                                                  0 integer
## 9
       TEAM BASERUN SB
                               2
                                     0.09
                                           131
                                                 5.76
                                                           0
                                                                  0 integer
                                                                                348
       TEAM_BATTING_SO
                                                                                822
## 8
                              20
                                     0.88
                                           102
                                                 4.48
                                                           0
                                                                  0 integer
## 15 TEAM PITCHING SO
                              20
                                     0.88
                                           102
                                                 4.48
                                                                  0 integer
                                                                                823
## 1
                               0
                                     0.00
                                              0
                                                 0.00
                                                                               2276
                  INDEX
                                                           0
                                                                  0 integer
## 2
            TARGET_WINS
                               1
                                     0.04
                                                 0.00
                                                           0
                                                                                108
                                                                  0 integer
## 3
        TEAM_BATTING_H
                                0
                                     0.00
                                              0
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                569
       TEAM BATTING 2B
                                0
                                     0.00
                                                 0.00
                                                                                240
                                                                  0 integer
                                     0.09
                               2
                                                                                144
## 5
       TEAM_BATTING_3B
                                                 0.00
                                                           0
                                              0
                                                                  0 integer
## 6
       TEAM_BATTING_HR
                              15
                                     0.66
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                243
## 7
       TEAM_BATTING_BB
                                     0.04
                                                 0.00
                                                           0
                                                                                533
                               1
                                                                  0 integer
## 12
       TEAM_PITCHING_H
                               0
                                     0.00
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                843
## 13 TEAM_PITCHING_HR
                                     0.66
                                                                                256
                              15
                                              0
                                                 0.00
                                                           0
                                                                  0 integer
## 14 TEAM_PITCHING_BB
                               1
                                     0.04
                                                 0.00
                                                           0
                                                                  0 integer
                                                                                535
       TEAM_FIELDING_E
                               0
                                     0.00
                                                 0.00
                                                           0
                                                                                549
                                                                  0 integer
##
               variable q_zeros p_zeros q_na p_na q_inf p_inf
                                                                      type unique
## 1
                               0
                                              0
                                                   0
                                                          0
                  INDEX
                                     0.00
                                                                0 integer
                                                                              2276
                                                   0
## 2
           TARGET_WINS
                                1
                                     0.04
                                              0
                                                          0
                                                                0 integer
                                                                               108
## 3
        TEAM BATTING H
                                0
                                     0.00
                                                   0
                                                          0
                                                                0 integer
                                                                               569
## 4
       TEAM BATTING 2B
                               0
                                     0.00
                                                   0
                                                          0
                                                                               240
                                              0
                                                                0 integer
## 5
       TEAM BATTING 3B
                                2
                                     0.09
                                              0
                                                   0
                                                          0
                                                                0 integer
                                                                               144
## 6
                              15
                                     0.66
                                              0
                                                   0
                                                          0
                                                                               243
       TEAM_BATTING_HR
                                                                0 integer
## 7
       TEAM_BATTING_BB
                               1
                                     0.04
                                                   0
                                                          0
                                                                0 integer
                                                                               533
       TEAM_BATTING_SO
                              20
                                     0.88
                                                   0
                                                          0
## 8
                                              0
                                                                0 numeric
                                                                               823
## 9
       TEAM_BASERUN_SB
                                2
                                     0.09
                                              0
                                                   0
                                                          0
                                                                0 numeric
                                                                               349
## 10
       TEAM_BASERUN_CS
                               1
                                     0.04
                                                   0
                                                                0 numeric
                                                                               129
      TEAM_BATTING_HBP
                               0
                                     0.00
                                              0
                                                   0
                                                          0
                                                                0 numeric
                                                                                56
  11
       TEAM_PITCHING_H
                               0
                                     0.00
                                              0
                                                   0
                                                          0
                                                                0 integer
                                                                               843
                                     0.66
## 13 TEAM_PITCHING_HR
                              15
                                              0
                                                   0
                                                          0
                                                                               256
                                                                0 integer
## 14 TEAM_PITCHING_BB
                               1
                                     0.04
                                                   0
                                                          0
                                                                0 integer
                                                                               535
## 15 TEAM_PITCHING_SO
                              20
                                     0.88
                                              0
                                                   0
                                                          0
                                                                0 numeric
                                                                               824
       TEAM FIELDING E
                               0
                                     0.00
                                                   0
                                                                0 integer
```

##	17	TEAM_FIELDING_DP	0	0.00	0	0	0	0 numeric	145
##	18	HBP_missing	191	8.39	0	0	0	0 numeric	2
##	19	CS_missing	1504	66.08	0	0	0	0 numeric	2
##	20	DP missing	1990	87.43	0	0	0	0 numeric	2

#### **Build Model**

#### Model 1

Model 1 is based on reviewing the p value from a model with all the predictors fitted. After reviewing the predictors that were statistically significant, a decission was made to fit this model with three statically significant predictors: TEAM\_BATTING\_H, TEAM\_BASERUN\_SB, and TEAM\_FIELDING\_E. The coefficients in this model tells us that there are possitive coefficients for TEAM\_BATTING\_H (Base Hits by batters) and TEAM\_BASERUN\_SB (Stolen bases) which makes sense since both of ther predictors will cause target wins to go up. The predictor with negative coefficient is TEAM\_FIELDING\_E(errors) which will indeed cause your teams wins to go down. This model has an R^2 of 0.27 and the F-statistics is 284

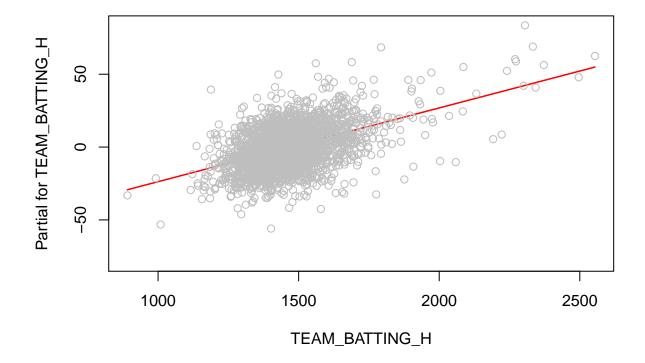
```
##
## Call:
  lm(formula = TARGET_WINS ~ (TEAM_BATTING_H + TEAM_BASERUN_SB +
##
       TEAM_FIELDING_E), data = train_data)
##
##
  Residuals:
##
       Min
                1Q
                    Median
                                3Q
                                       Max
   -52.536
           -8.953
                     0.136
                             8.776
                                    53.676
##
##
##
  Coefficients:
##
                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                    8.010983
                               2.928195
                                          2.736
                                                 0.00627 **
## TEAM_BATTING_H
                    0.050627
                               0.002022
                                         25.043
                                                  < 2e-16 ***
## TEAM_BASERUN_SB
                    0.038590
                               0.003558
                                         10.847
                                                  < 2e-16 ***
## TEAM FIELDING E -0.026042
                               0.001372 -18.987
                                                 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 13.44 on 2272 degrees of freedom
## Multiple R-squared: 0.2728, Adjusted R-squared: 0.2718
## F-statistic: 284.1 on 3 and 2272 DF, p-value: < 2.2e-16
```

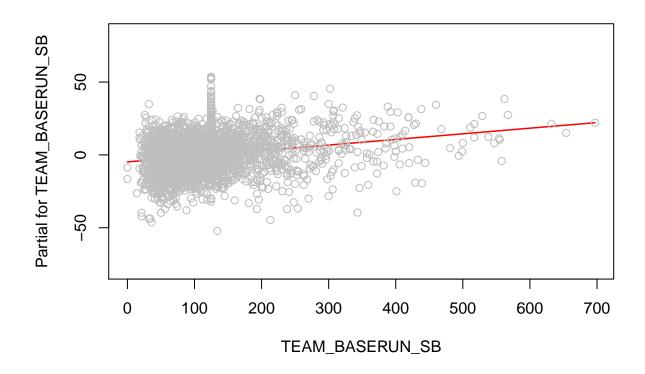
	Estimate	Std. Error	t value	$\Pr(> t )$
(Intercept)	8.011	2.928	2.736	0.006271
TEAM_BATTING_H	0.05063	0.002022	25.04	1.979e-122
TEAM_BASERUN_SB	0.03859	0.003558	10.85	9.252e-27
${f TEAM\_FIELDING\_E}$	-0.02604	0.001372	-18.99	9.866e-75

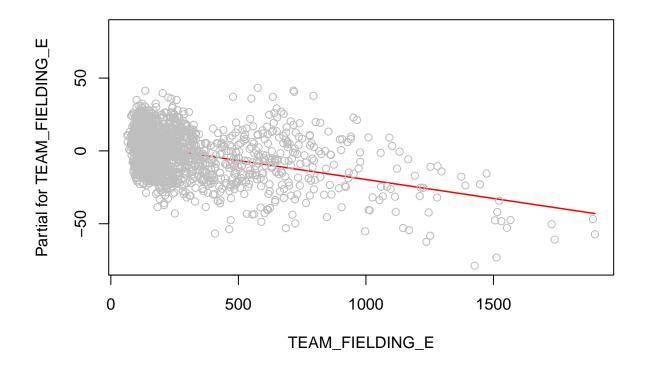
```
Table 2: Fitting linear model: TARGET_WINS \sim (TEAM_BATTING_H + TEAM_BASERUN_SB + TEAM_FIELDING_E)
```

Observations	Residual Std. Error	$R^2$	Adjusted $\mathbb{R}^2$
2276	13.44	0.2728	0.2718

The termplot below gives us a visual of the coefficients from model 1 agains the slopes of each of the predictors.







### R Code

```
# Data Preparation
eva_data <- read.csv(file = "moneyball-evaluation-data.csv",</pre>
    header = TRUE, sep = ",")
train_data <- read.csv(file = "moneyball-training-data.csv",</pre>
    header = TRUE, sep = ",")
eva_data_status <- df_status(eva_data)</pre>
eva_data_status[order(-eva_data_status$p_zeros), ]
train_data_status <- df_status(train_data)</pre>
train_data_status[order(-train_data_status$p_na), ]
train_data$HBP_missing <- ifelse(is.na(train_data$TEAM_BATTING_HBP),</pre>
train_data$TEAM_BATTING_HBP[is.na(train_data$TEAM_BATTING_HBP)] <- mean(train_data$TEAM_BATTING_HBP,
    na.rm = TRUE)
train_data$CS_missing <- ifelse(is.na(train_data$TEAM_BASERUN_CS),</pre>
train_data$TEAM_BASERUN_CS[is.na(train_data$TEAM_BASERUN_CS)] <- mean(train_data$TEAM_BASERUN_CS,
    na.rm = TRUE)
train_data$DP_missing <- ifelse(is.na(train_data$TEAM_FIELDING_DP),</pre>
```

```
1, 0)
train_data$TEAM_FIELDING_DP[is.na(train_data$TEAM_FIELDING_DP)] <- mean(train_data$TEAM_FIELDING_DP,
    na.rm = TRUE)
train_data$TEAM_BASERUN_SB[is.na(train_data$TEAM_BASERUN_SB)] <- mean(train_data$TEAM_BASERUN_SB,
    na.rm = TRUE)
train_data$TEAM_BATTING_SO[is.na(train_data$TEAM_BATTING_SO)] <- mean(train_data$TEAM_BATTING_SO,
    na.rm = TRUE)
train_data$TEAM_PITCHING_SO[is.na(train_data$TEAM_PITCHING_SO)] <- mean(train_data$TEAM_PITCHING_SO,
    na.rm = TRUE)
df_status(train_data)
# Build Model
# This model contains all the predictor variables from our
# training dataset. All significant and non-significan
# predictors are displayed in the regression model below:
# remove the index column from the train data.
train_data <- select(train_data, -c(INDEX))</pre>
model_all_predictors <- lm(TARGET_WINS ~ ., data = train_data)</pre>
summary(model_all_predictors)
## Model 1
model1 <- lm(TARGET_WINS ~ (TEAM_BATTING_H + TEAM_BASERUN_SB +</pre>
    TEAM_FIELDING_E), data = train_data)
summary(model1)
pander(summary(model1))
termplot(model1)
```