

DATA621-Week3-Homework1

Muthukumar Srinivasan 7 Rajagopal Srinivasan

March 19, 2017

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(knitr)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.3

require(gridExtra)

## Loading required package: gridExtra

## Warning: package 'gridExtra' was built under R version 3.2.5
```

You can also embed plots, for example:

```
## 'data.frame':    2276 obs. of  17 variables:
## $ INDEX      : int  1 2 3 4 5 6 7 8 11 12 ...
## $ TARGET_WINS : int  39 70 86 70 82 75 80 85 86 76 ...
## $ TEAM_BATTING_H : int  1445 1339 1377 1387 1297 1279 1244 1273 1391 1271 ...
## $ TEAM_BATTING_2B : int  194 219 232 209 186 200 179 171 197 213 ...
## $ TEAM_BATTING_3B : int  39 22 35 38 27 36 54 37 40 18 ...
## $ TEAM_BATTING_HR : int  13 190 137 96 102 92 122 115 114 96 ...
## $ TEAM_BATTING_BB : int  143 685 602 451 472 443 525 456 447 441 ...
## $ TEAM_BATTING_SO : int  842 1075 917 922 920 973 1062 1027 922 827 ...
## $ TEAM_BASERUN_SB : int  NA 37 46 43 49 107 80 40 69 72 ...
## $ TEAM_BASERUN_CS : int  NA 28 27 30 39 59 54 36 27 34 ...
## $ TEAM_BATTING_HBP: int  NA NA NA NA NA NA NA NA NA ...
## $ TEAM_PITCHING_H : int  9364 1347 1377 1396 1297 1279 1244 1281 1391 1271 ...
## $ TEAM_PITCHING_HR: int  84 191 137 97 102 92 122 116 114 96 ...
## $ TEAM_PITCHING_BB: int  927 689 602 454 472 443 525 459 447 441 ...
## $ TEAM_PITCHING_SO: int  5456 1082 917 928 920 973 1062 1033 922 827 ...
## $ TEAM_FIELDING_E : int  1011 193 175 164 138 123 136 112 127 131 ...
## $ TEAM_FIELDING_DP: int  NA 155 153 156 168 149 186 136 169 159 ...

##             TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## TEAM_BATTING_H      1.000       0.563       0.428
## TEAM_BATTING_2B     0.563       1.000      -0.107
## TEAM_BATTING_3B     0.428      -0.107       1.000
## TEAM_BATTING_HR    -0.007       0.435      -0.636
## TEAM_BATTING_BB    -0.072       0.256      -0.287
## TEAM_BATTING_SO      NA         NA         NA
```

## TEAM_BASERUN_SB	NA	NA	NA
## TEAM_BASERUN_CS	NA	NA	NA
## TEAM_BATTING_HBP	NA	NA	NA
## TEAM_PITCHING_H	0.303	0.024	0.195
## TEAM_PITCHING_HR	0.073	0.455	-0.568
## TEAM_PITCHING_BB	0.094	0.178	-0.002
## TEAM_PITCHING_SO	NA	NA	NA
## TEAM_FIELDING_E	0.265	-0.235	0.510
## TEAM_FIELDING_DP	NA	NA	NA
	TEAM_BATTING_HR	TEAM_BATTING_BB	TEAM_BATTING_SO
## TEAM_BATTING_H	-0.007	-0.072	NA
## TEAM_BATTING_2B	0.435	0.256	NA
## TEAM_BATTING_3B	-0.636	-0.287	NA
## TEAM_BATTING_HR	1.000	0.514	NA
## TEAM_BATTING_BB	0.514	1.000	NA
## TEAM_BATTING_SO	NA	NA	1
## TEAM_BASERUN_SB	NA	NA	NA
## TEAM_BASERUN_CS	NA	NA	NA
## TEAM_BATTING_HBP	NA	NA	NA
## TEAM_PITCHING_H	-0.250	-0.450	NA
## TEAM_PITCHING_HR	0.969	0.460	NA
## TEAM_PITCHING_BB	0.137	0.489	NA
## TEAM_PITCHING_SO	NA	NA	NA
## TEAM_FIELDING_E	-0.587	-0.656	NA
## TEAM_FIELDING_DP	NA	NA	NA
	TEAM_BASERUN_SB	TEAM_BASERUN_CS	TEAM_BATTING_HBP
## TEAM_BATTING_H	NA	NA	NA
## TEAM_BATTING_2B	NA	NA	NA
## TEAM_BATTING_3B	NA	NA	NA
## TEAM_BATTING_HR	NA	NA	NA
## TEAM_BATTING_BB	NA	NA	NA
## TEAM_BATTING_SO	NA	NA	NA
## TEAM_BASERUN_SB	1	NA	NA
## TEAM_BASERUN_CS	NA	1	NA
## TEAM_BATTING_HBP	NA	NA	1
## TEAM_PITCHING_H	NA	NA	NA
## TEAM_PITCHING_HR	NA	NA	NA
## TEAM_PITCHING_BB	NA	NA	NA
## TEAM_PITCHING_SO	NA	NA	NA
## TEAM_FIELDING_E	NA	NA	NA
## TEAM_FIELDING_DP	NA	NA	NA
	TEAM_PITCHING_H	TEAM_PITCHING_HR	TEAM_PITCHING_BB
## TEAM_BATTING_H	0.303	0.073	0.094
## TEAM_BATTING_2B	0.024	0.455	0.178
## TEAM_BATTING_3B	0.195	-0.568	-0.002
## TEAM_BATTING_HR	-0.250	0.969	0.137
## TEAM_BATTING_BB	-0.450	0.460	0.489
## TEAM_BATTING_SO	NA	NA	NA
## TEAM_BASERUN_SB	NA	NA	NA
## TEAM_BASERUN_CS	NA	NA	NA
## TEAM_BATTING_HBP	NA	NA	NA
## TEAM_PITCHING_H	1.000	-0.142	0.321
## TEAM_PITCHING_HR	-0.142	1.000	0.222
## TEAM_PITCHING_BB	0.321	0.222	1.000

```

## TEAM_PITCHING_SO          NA          NA          NA
## TEAM_FIELDING_E           0.668      -0.493     -0.023
## TEAM_FIELDING_DP          NA          NA          NA
##              TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP
## TEAM_BATTING_H             NA          0.265      NA
## TEAM_BATTING_2B            NA         -0.235      NA
## TEAM_BATTING_3B            NA          0.510      NA
## TEAM_BATTING_HR            NA         -0.587      NA
## TEAM_BATTING_BB            NA         -0.656      NA
## TEAM_BATTING_SO            NA          NA          NA
## TEAM_BASERUN_SB            NA          NA          NA
## TEAM_BASERUN_CS            NA          NA          NA
## TEAM_BATTING_HBP           NA          NA          NA
## TEAM_PITCHING_H             NA          0.668      NA
## TEAM_PITCHING_HR            NA         -0.493      NA
## TEAM_PITCHING_BB            NA         -0.023      NA
## TEAM_PITCHING_SO            1          NA          NA
## TEAM_FIELDING_E             NA         1.000      NA
## TEAM_FIELDING_DP             NA          NA          1

## [1] 144.5912

## [1] 46.80141

## [1] 27.93856

## [1] 60.54687

## [1] 122.6709

## [1] NA

## [1] NA

## [1] NA

## [1] NA

## [1] 1406.843

## [1] 61.29875

## [1] 166.3574

## [1] NA

## [1] 227.771

## [1] NA

```

```

## [1] 1469.27

## [1] 241.2469

## [1] 55.25

## [1] 99.61204

## [1] 501.5589

## [1] NA

## [1] NA

## [1] NA

## [1] NA

## [1] 1779.21

## [1] 105.6986

## [1] 553.0079

## [1] NA

## [1] 246.4807

## [1] NA

##      INDEX      TARGET_WINS      TEAM_BATTING_H TEAM_BATTING_2B
##  Min.   : 1.0   Min.   : 0.00   Min.   : 891   Min.   : 69.0
##  1st Qu.: 630.8 1st Qu.: 71.00  1st Qu.:1383   1st Qu.:208.0
##  Median :1270.5 Median : 82.00  Median :1454   Median :238.0
##  Mean   :1268.5 Mean   : 80.79  Mean   :1469   Mean   :241.2
##  3rd Qu.:1915.5 3rd Qu.: 92.00 3rd Qu.:1537   3rd Qu.:273.0
##  Max.   :2535.0 Max.   :146.00  Max.   :2554   Max.   :458.0
##
##      TEAM_BATTING_3B      TEAM_BATTING_HR      TEAM_BATTING_BB TEAM_BATTING_SO
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.0   Min.   : 0.0
##  1st Qu.: 34.00  1st Qu.: 42.00  1st Qu.:451.0  1st Qu.: 548.0
##  Median : 47.00  Median :102.00  Median :512.0  Median : 750.0
##  Mean   : 55.25  Mean   : 99.61  Mean   :501.6  Mean   : 735.6
##  3rd Qu.: 72.00  3rd Qu.:147.00 3rd Qu.:580.0  3rd Qu.: 930.0
##  Max.   :223.00  Max.   :264.00  Max.   :878.0  Max.   :1399.0
##
##                               NA's   :102
##      TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_BATTING_HBP TEAM_PITCHING_H
##  Min.   : 0.0   Min.   : 0.0   Min.   :29.00   Min.   : 1137
##  1st Qu.: 66.0  1st Qu.: 38.0  1st Qu.:50.50   1st Qu.: 1419
##  Median :101.0  Median : 49.0  Median :58.00   Median : 1518

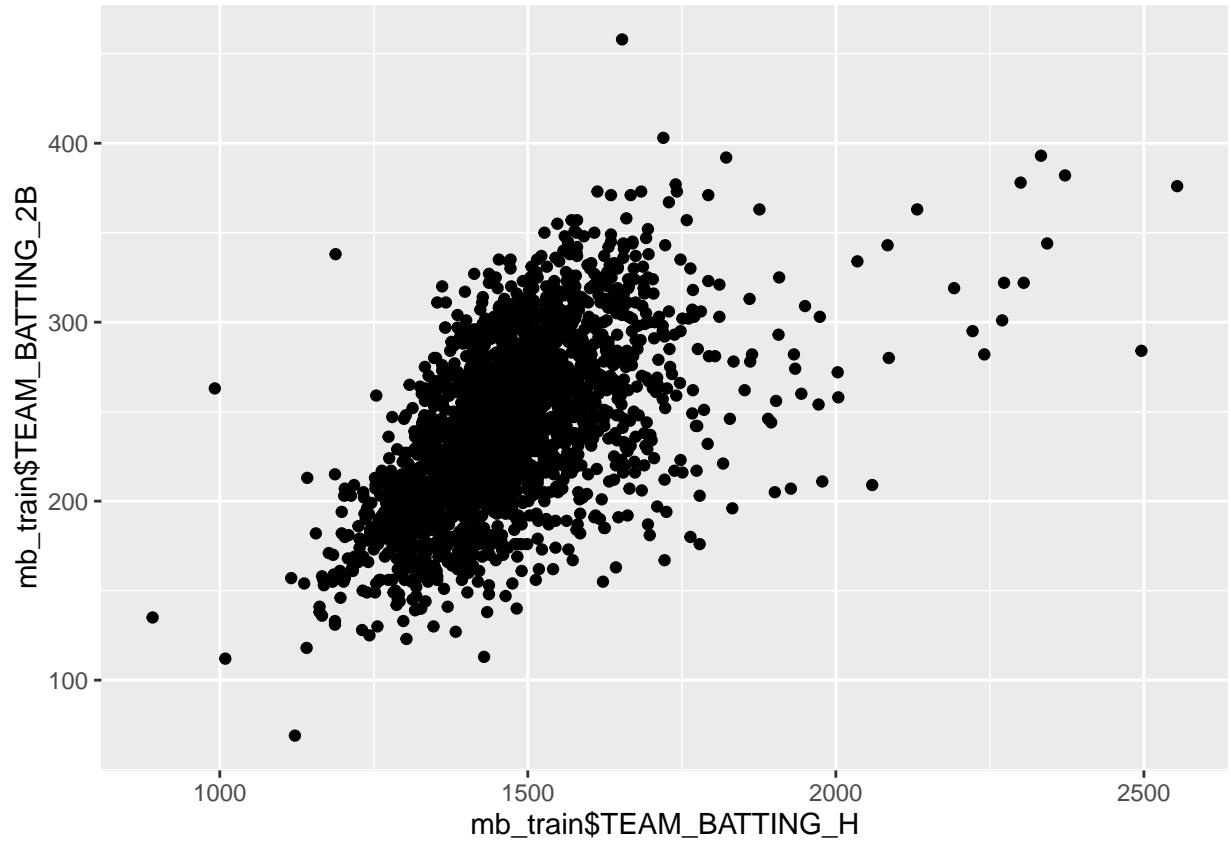
```

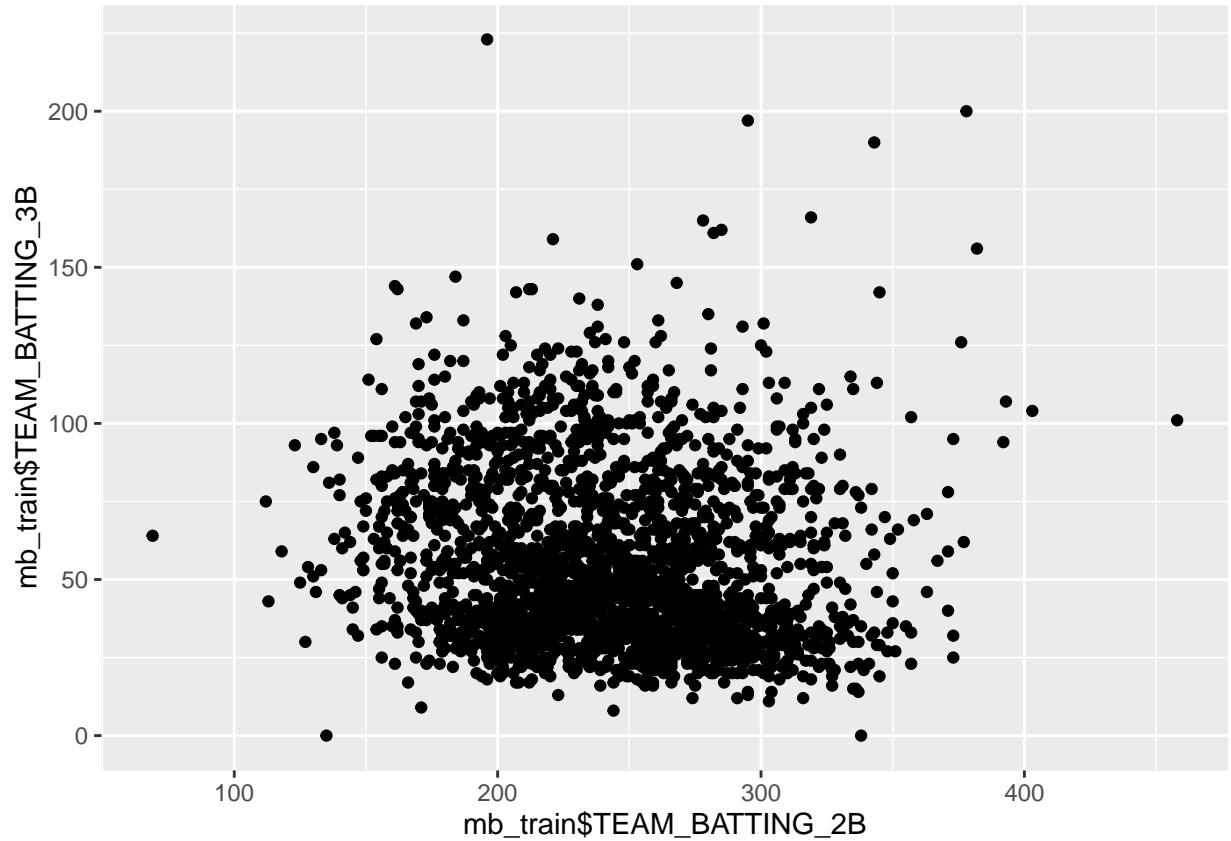
```

##  Mean    :124.8   Mean    : 52.8   Mean    :59.36   Mean    : 1779
## 3rd Qu.:156.0   3rd Qu.: 62.0   3rd Qu.:67.00   3rd Qu.: 1682
## Max.    :697.0   Max.    :201.0   Max.    :95.00   Max.    :30132
## NA's    :131     NA's    :772     NA's    :2085
## TEAM_PITCHING_HR TEAM_PITCHING_BB TEAM_PITCHING_SO  TEAM_FIELDING_E
## Min.    : 0.0    Min.    : 0.0    Min.    : 0.0    Min.    : 65.0
## 1st Qu.: 50.0   1st Qu.: 476.0  1st Qu.: 615.0  1st Qu.: 127.0
## Median  :107.0   Median  : 536.5  Median  : 813.5  Median  : 159.0
## Mean    :105.7   Mean    : 553.0  Mean    : 817.7  Mean    : 246.5
## 3rd Qu.:150.0   3rd Qu.: 611.0  3rd Qu.: 968.0  3rd Qu.: 249.2
## Max.    :343.0   Max.    :3645.0  Max.    :19278.0 Max.    :1898.0
##                                     NA's    :102
## TEAM_FIELDING_DP
## Min.    : 52.0
## 1st Qu.:131.0
## Median  :149.0
## Mean    :146.4
## 3rd Qu.:164.0
## Max.    :228.0
## NA's    :286

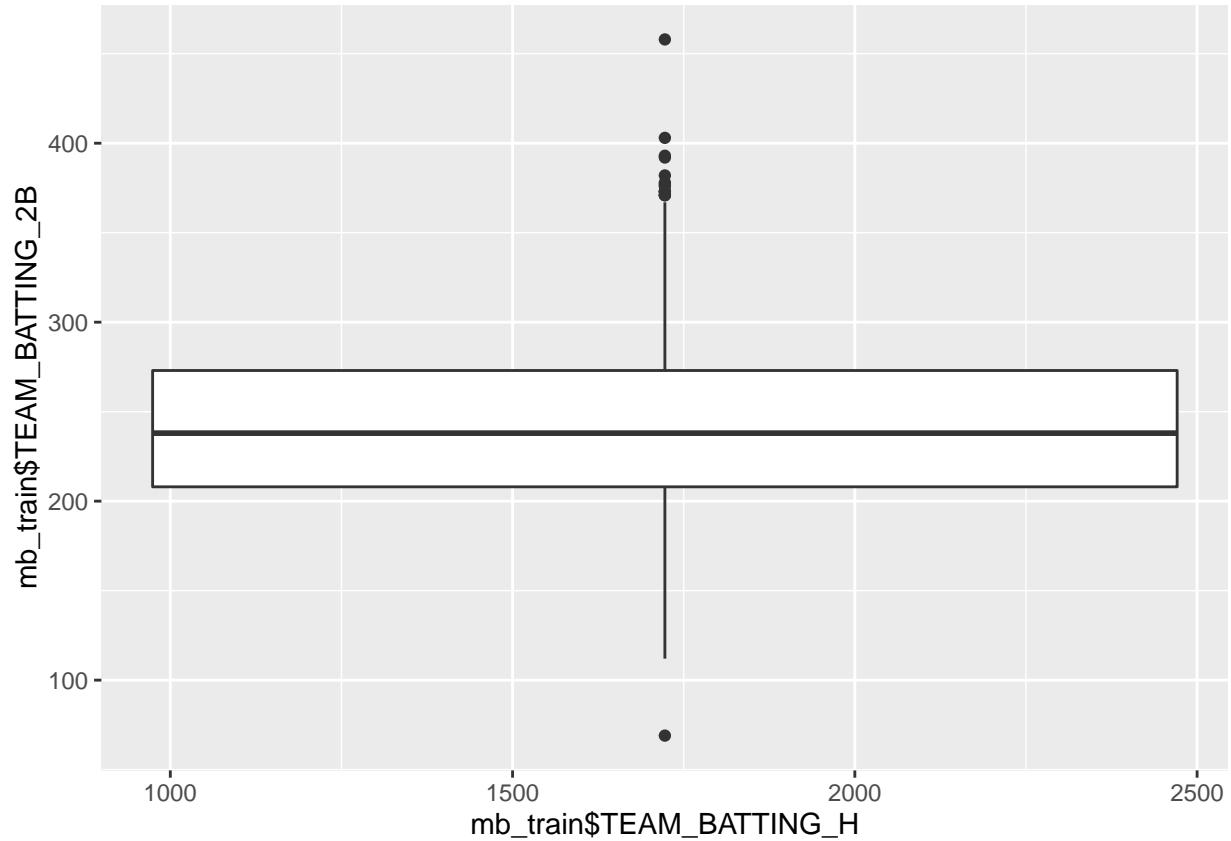
## 'data.frame': 2276 obs. of 17 variables:
## $ INDEX      : int  1 2 3 4 5 6 7 8 11 12 ...
## $ TARGET_WINS : int  39 70 86 70 82 75 80 85 86 76 ...
## $ TEAM_BATTING_H : int  1445 1339 1377 1387 1297 1279 1244 1273 1391 1271 ...
## $ TEAM_BATTING_2B : int  194 219 232 209 186 200 179 171 197 213 ...
## $ TEAM_BATTING_3B : int  39 22 35 38 27 36 54 37 40 18 ...
## $ TEAM_BATTING_HR : int  13 190 137 96 102 92 122 115 114 96 ...
## $ TEAM_BATTING_BB : int  143 685 602 451 472 443 525 456 447 441 ...
## $ TEAM_BATTING_SO : int  842 1075 917 922 920 973 1062 1027 922 827 ...
## $ TEAM_BASERUN_SB : int  NA 37 46 43 49 107 80 40 69 72 ...
## $ TEAM_BASERUN_CS : int  NA 28 27 30 39 59 54 36 27 34 ...
## $ TEAM_BATTING_HBP: int  NA NA NA NA NA NA NA NA NA ...
## $ TEAM_PITCHING_H : int  9364 1347 1377 1396 1297 1279 1244 1281 1391 1271 ...
## $ TEAM_PITCHING_HR: int  84 191 137 97 102 92 122 116 114 96 ...
## $ TEAM_PITCHING_BB: int  927 689 602 454 472 443 525 459 447 441 ...
## $ TEAM_PITCHING_SO: int  5456 1082 917 928 920 973 1062 1033 922 827 ...
## $ TEAM_FIELDING_E : int  1011 193 175 164 138 123 136 112 127 131 ...
## $ TEAM_FIELDING_DP: int  NA 155 153 156 168 149 186 136 169 159 ...

```

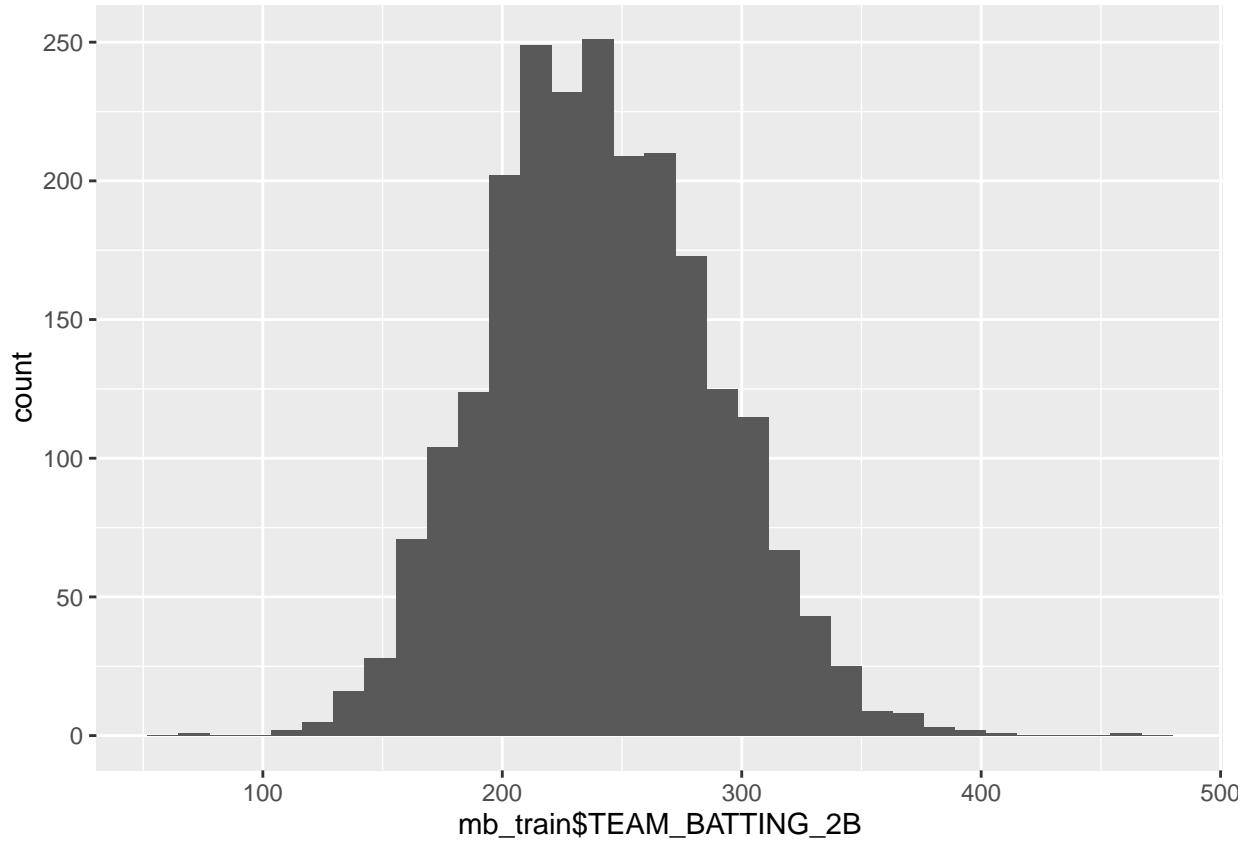


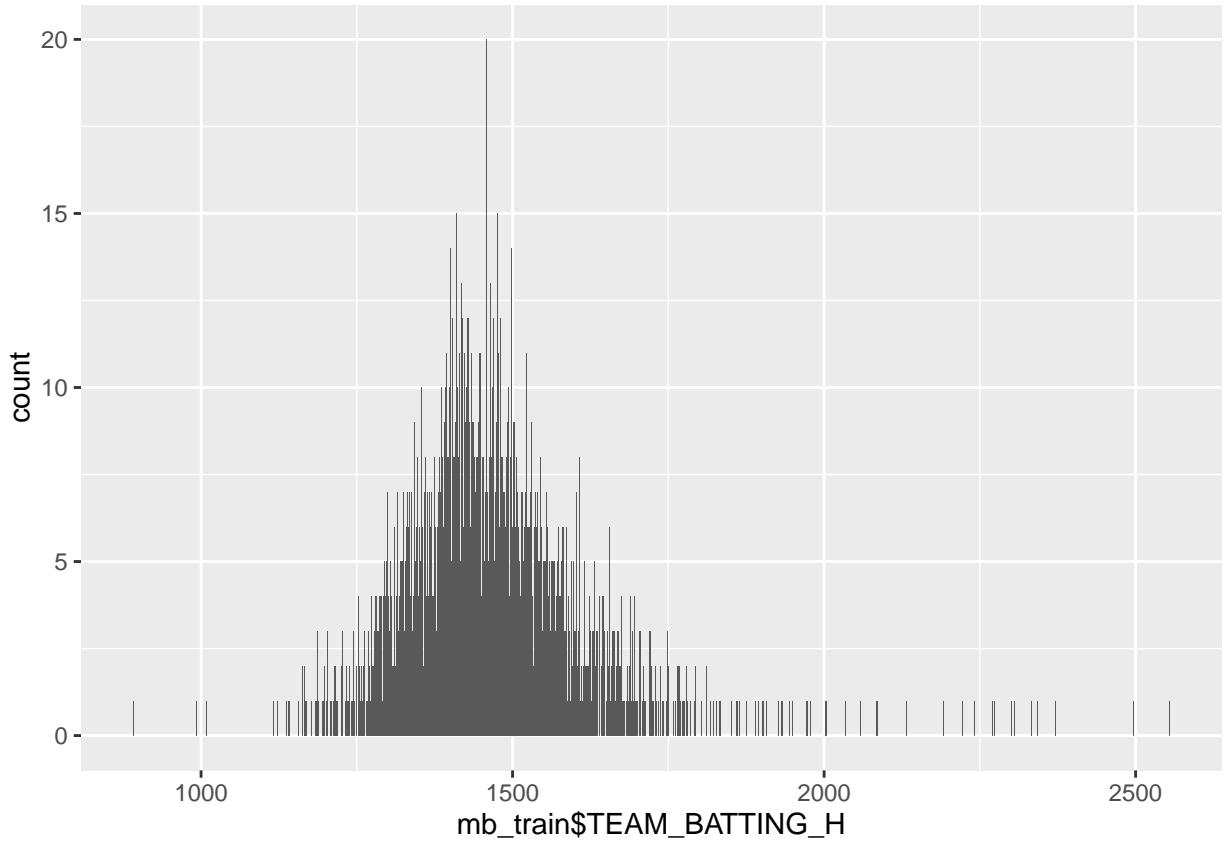


```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



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



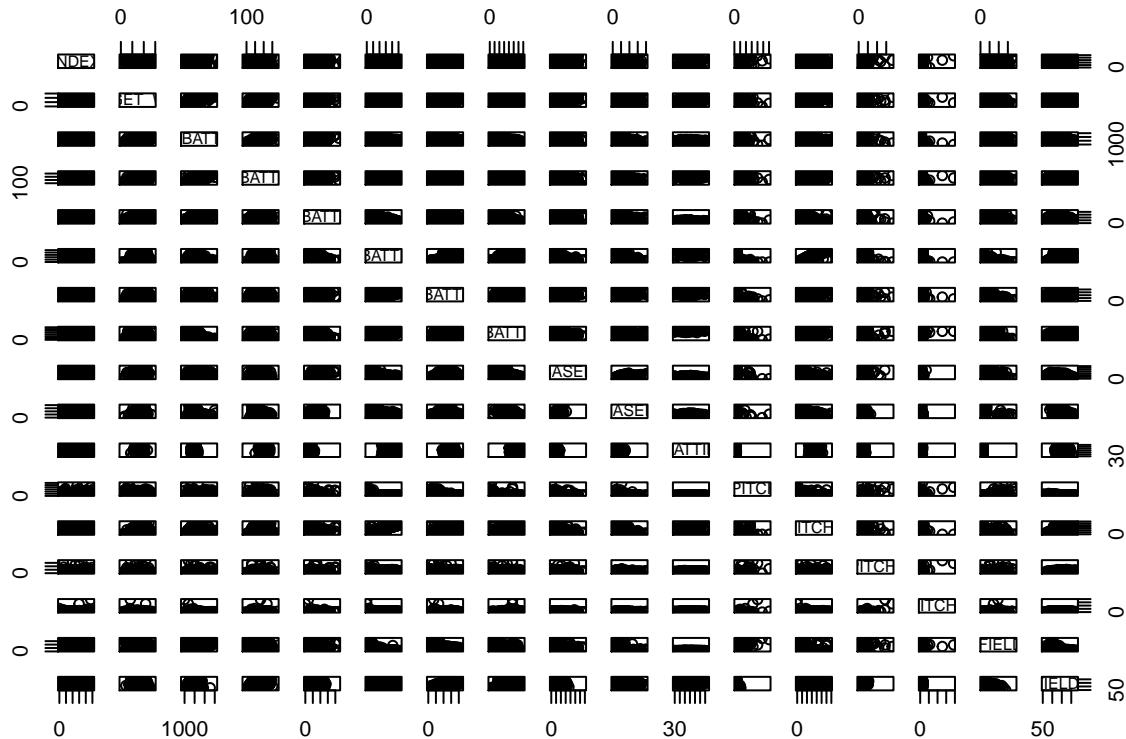


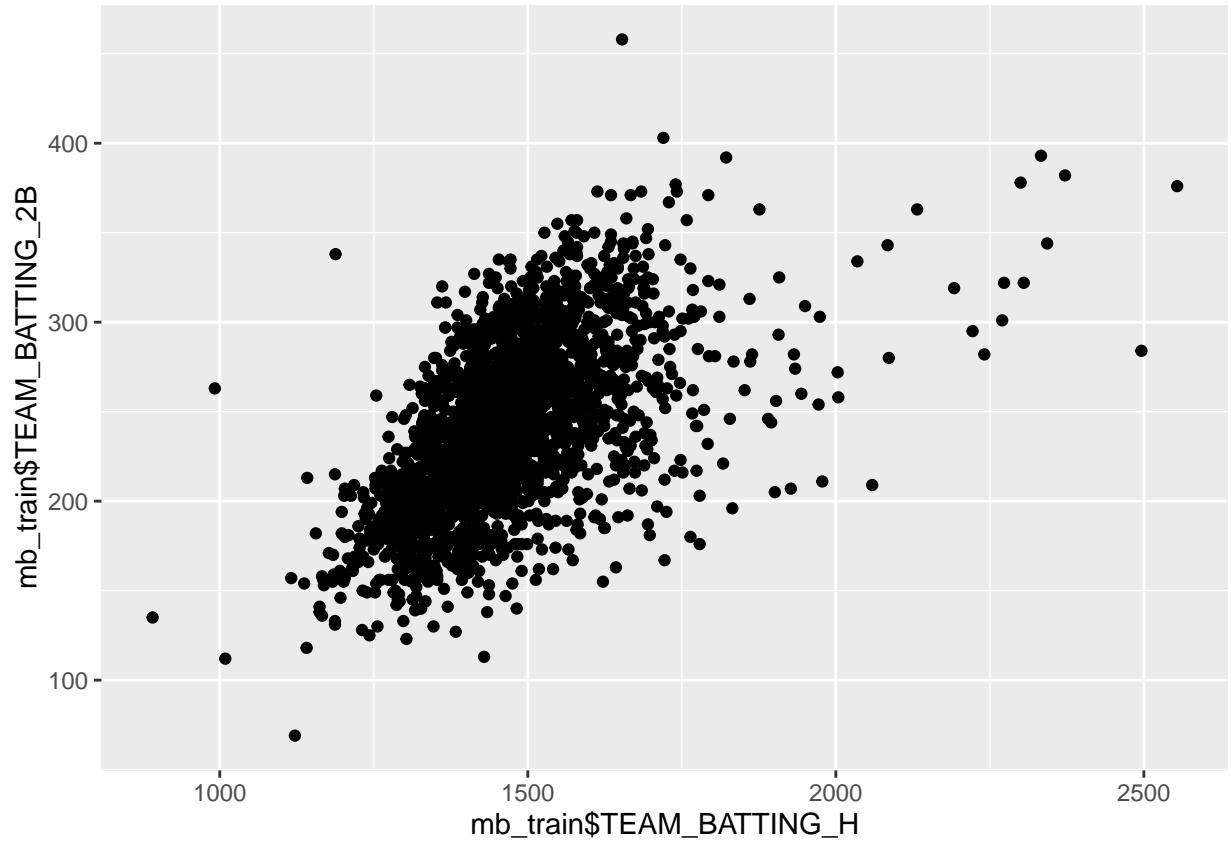
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_2B +
##     TEAM_BATTING_3B + TEAM_BATTING_HR + TEAM_BATTING_BB + TEAM_BATTING_SO +
##     TEAM_BASERUN_SB + TEAM_BASERUN_CS + TEAM_BATTING_HBP + TEAM_PITCHING_H +
##     TEAM_PITCHING_HR + TEAM_PITCHING_BB + TEAM_PITCHING_SO +
##     TEAM_FIELDING_E + TEAM_FIELDING_DP, data = mb_train)
##
## Residuals:
##      Min        1Q     Median        3Q       Max
## -19.8708  -5.6564  -0.0599  5.2545  22.9274
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.28826  19.67842   3.064  0.00253 **
## TEAM_BATTING_H  1.91348  2.76139   0.693  0.48927
## TEAM_BATTING_2B  0.02639  0.03029   0.871  0.38484
## TEAM_BATTING_3B -0.10118  0.07751  -1.305  0.19348
## TEAM_BATTING_HR -4.84371 10.50851  -0.461  0.64542
## TEAM_BATTING_BB -4.45969  3.63624  -1.226  0.22167
## TEAM_BATTING_SO  0.34196  2.59876   0.132  0.89546
## TEAM_BASERUN_SB  0.03304  0.02867   1.152  0.25071
## TEAM_BASERUN_CS -0.01104  0.07143  -0.155  0.87730
## TEAM_BATTING_HBP  0.08247  0.04960   1.663  0.09815 .
## TEAM_PITCHING_H -1.89096  2.76095  -0.685  0.49432
## TEAM_PITCHING_HR  4.93043 10.50664   0.469  0.63946
```

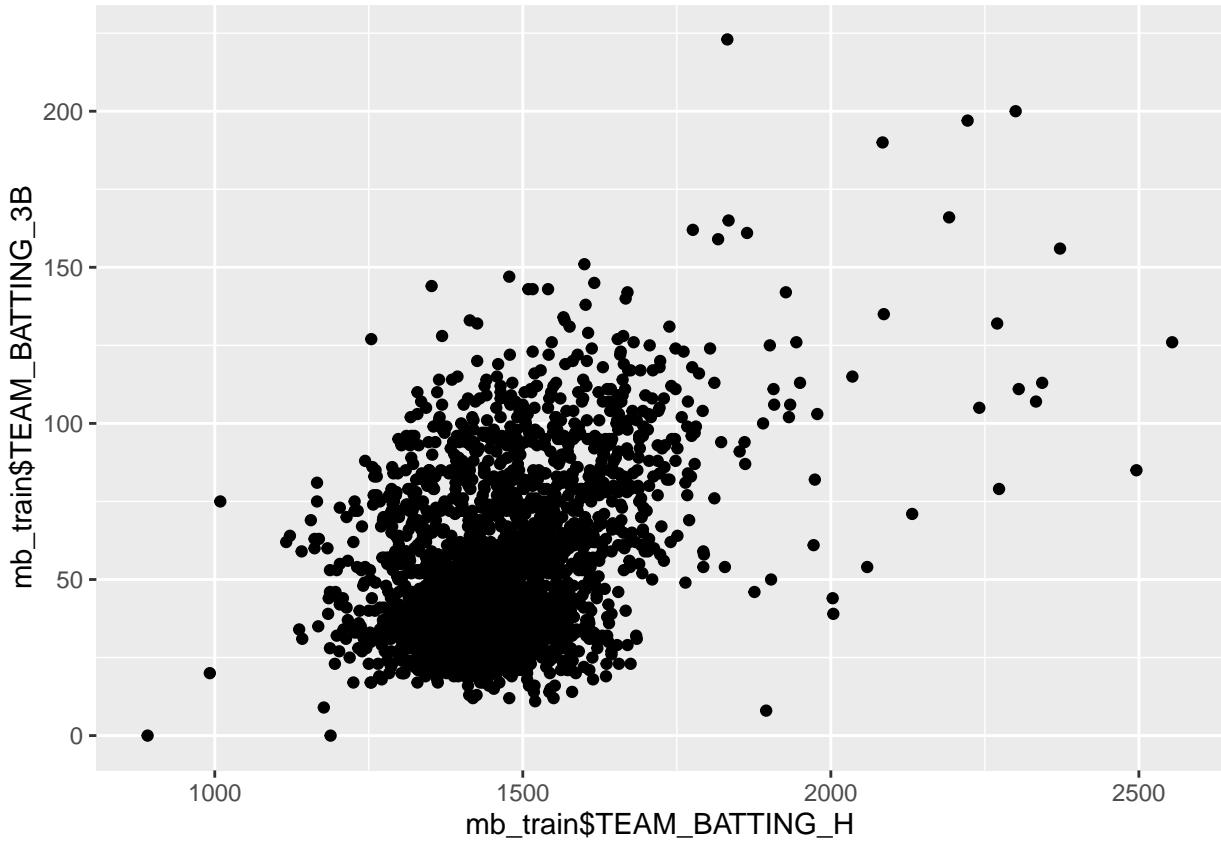
```

## TEAM_PITCHING_BB 4.51089    3.63372   1.241  0.21612
## TEAM_PITCHING_SO -0.37364   2.59705   -0.144  0.88577
## TEAM_FIELDING_E -0.17204   0.04140   -4.155 5.08e-05 ***
## TEAM_FIELDING_DP -0.10819   0.03654   -2.961  0.00349 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.467 on 175 degrees of freedom
##   (2085 observations deleted due to missingness)
## Multiple R-squared:  0.5501, Adjusted R-squared:  0.5116
## F-statistic: 14.27 on 15 and 175 DF,  p-value: < 2.2e-16

```







```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_2B +
##     TEAM_BATTING_3B + TEAM_BATTING_HR + TEAM_BATTING_BB + TEAM_BATTING_SO +
##     TEAM_BASERUN_SB + TEAM_BATTING_HBP + TEAM_PITCHING_H + TEAM_PITCHING_HR +
##     TEAM_PITCHING_BB + TEAM_PITCHING_SO + TEAM_FIELDING_E + TEAM_FIELDING_DP,
##     data = mb_train)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -19.8636  -5.6235  -0.1226   5.2900  22.8482
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 59.92880  19.48633  3.075  0.00244 **
## TEAM_BATTING_H  1.94556  2.74594  0.709  0.47956
## TEAM_BATTING_2B  0.02626  0.03019  0.870  0.38569
## TEAM_BATTING_3B -0.10297  0.07642 -1.348  0.17954
## TEAM_BATTING_HR -4.88545 10.47587 -0.466  0.64154
## TEAM_BATTING_BB -4.49878  3.61737 -1.244  0.21528
## TEAM_BATTING_SO  0.32395  2.58894  0.125  0.90057
## TEAM_BASERUN_SB  0.03042  0.02304  1.320  0.18844
## TEAM_BATTING_HBP  0.08248  0.04946  1.668  0.09718 .
## TEAM_PITCHING_H -1.92283  2.74559 -0.700  0.48464
## TEAM_PITCHING_HR  4.97242 10.47397  0.475  0.63556
## TEAM_PITCHING_BB  4.55005  3.61482  1.259  0.20980
```

```

## TEAM_PITCHING_SO -0.35558    2.58722  -0.137  0.89084
## TEAM_FIELDING_E  -0.17343    0.04030  -4.303  2.79e-05 ***
## TEAM_FIELDING_DP -0.10849    0.03639  -2.981  0.00328 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.444 on 176 degrees of freedom
##   (2085 observations deleted due to missingness)
## Multiple R-squared:  0.5501, Adjusted R-squared:  0.5143
## F-statistic: 15.37 on 14 and 176 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_2B +
##     TEAM_BATTING_3B + TEAM_BATTING_HR + TEAM_BATTING_BB + TEAM_BATTING_SO +
##     TEAM_BASERUN_SB + TEAM_BATTING_HBP + TEAM_PITCHING_H + TEAM_PITCHING_BB +
##     TEAM_PITCHING_SO + TEAM_FIELDING_E + TEAM_FIELDING_DP, data = mb_train)
##
## Residuals:
##      Min        1Q        Median       3Q        Max
## -19.8294  -5.7340   0.0122   5.4517  22.9175
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 59.97819  19.44337  3.085 0.002364 ** 
## TEAM_BATTING_H  1.24194  2.30655  0.538 0.590950  
## TEAM_BATTING_2B  0.02688  0.03010  0.893 0.372982  
## TEAM_BATTING_3B -0.10586  0.07601 -1.393 0.165418  
## TEAM_BATTING_HR  0.08786  0.02588  3.396 0.000846 *** 
## TEAM_BATTING_BB -4.36781  3.59894 -1.214 0.226502  
## TEAM_BATTING_SO  0.39728  2.57866  0.154 0.877733  
## TEAM_BASERUN_SB  0.03257  0.02254  1.445 0.150285  
## TEAM_BATTING_HBP 0.08317  0.04933  1.686 0.093559 .  
## TEAM_PITCHING_H -1.21947  2.30648 -0.529 0.597666  
## TEAM_PITCHING_BB 4.41869  3.59632  1.229 0.220826  
## TEAM_PITCHING_SO -0.42892  2.57694 -0.166 0.867998  
## TEAM_FIELDING_E -0.17404  0.04019 -4.330 2.49e-05 *** 
## TEAM_FIELDING_DP -0.10782  0.03628 -2.972 0.003374 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.425 on 177 degrees of freedom
##   (2085 observations deleted due to missingness)
## Multiple R-squared:  0.5495, Adjusted R-squared:  0.5164
## F-statistic: 16.61 on 13 and 177 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_2B +
##     TEAM_BATTING_3B + TEAM_BATTING_HR + TEAM_BATTING_BB + TEAM_BATTING_SO +
##     TEAM_BASERUN_SB + TEAM_PITCHING_H + TEAM_PITCHING_BB + TEAM_PITCHING_SO +
##     TEAM_FIELDING_E + TEAM_FIELDING_DP, data = mb_train)
##
## Residuals:
```

```

##      Min     1Q   Median     3Q    Max
## -32.168 -7.256    0.142   6.945  29.896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 59.054832  6.025125  9.801 < 2e-16 ***
## TEAM_BATTING_H -0.033844  0.015767 -2.146  0.03197 *
## TEAM_BATTING_2B -0.049268  0.008874 -5.552 3.24e-08 ***
## TEAM_BATTING_3B  0.183497  0.018984  9.666 < 2e-16 ***
## TEAM_BATTING_HR  0.100263  0.009155 10.952 < 2e-16 ***
## TEAM_BATTING_BB  0.118363  0.041385  2.860  0.00428 **
## TEAM_BATTING_SO  0.033316  0.017529  1.901  0.05751 .
## TEAM_BASERUN_SB  0.069465  0.005534 12.552 < 2e-16 ***
## TEAM_PITCHING_H  0.059813  0.014323  4.176 3.11e-05 ***
## TEAM_PITCHING_BB -0.080309  0.039308 -2.043  0.04119 *
## TEAM_PITCHING_SO -0.053523  0.016559 -3.232  0.00125 **
## TEAM_FIELDING_E -0.118864  0.007122 -16.690 < 2e-16 ***
## TEAM_FIELDING_DP -0.112317  0.012271 -9.153 < 2e-16 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.18 on 1822 degrees of freedom
##   (441 observations deleted due to missingness)
## Multiple R-squared:  0.4058, Adjusted R-squared:  0.4019
## F-statistic: 103.7 on 12 and 1822 DF,  p-value: < 2.2e-16

```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.