Data621 HW1

Shariq Mian

2/28/2022

```
library(ggcorrplot)
```

```
## Warning: package 'ggcorrplot' was built under R version 4.1.2
```

```
## Loading required package: ggplot2
```

```
library(DT)
library(zoo)
```

```
## Warning: package 'zoo' was built under R version 4.1.2
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(heatmaply)
```

```
## Warning: package 'heatmaply' was built under R version 4.1.2
```

```
## Loading required package: plotly
```

```
## Warning: package 'plotly' was built under R version 4.1.2

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

## Loading required package: viridis

## Warning: package 'viridis' was built under R version 4.1.2

## Loading required package: viridisLite

##
## ======================
## Welcome to heatmaply version 1.3.0
##
## Type citation('heatmaply') for how to cite the package.
## Type ?heatmaply for the main documentation.
##
## The github page is: https://github.com/talgalili/heatmaply/
## Please submit your suggestions and bug-reports at: https://github.com/talgalili/heatmaply/issues
## You may ask questions at stackoverflow, use the r and heatmaply tags:
##   https://stackoverflow.com/questions/tagged/heatmaply
## ======================
```

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.1.2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
train_df=read.csv("~/GitHub/SPS/Data 621/HW1/moneyball-training-data.csv")
head(train_df)
```

```
##   INDEX TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## 1     1          39           1445             194              39
## 2     2          70           1339             219              22
```

```
## 3     3        86          1377            232                35
## 4     4        70          1387            209                38
## 5     5        82          1297            186                27
## 6     6        75          1279            200                36
##   TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO TEAM_BASERUN_SB
## 1              13             143             842              NA
## 2             190             685            1075              37
## 3             137             602             917              46
## 4              96             451             922              43
## 5             102             472             920              49
## 6              92             443             973             107
##   TEAM_BASERUN_CS TEAM_BATTING_HBP TEAM_PITCHING_H TEAM_PITCHING_HR
## 1              NA               NA            9364               84
## 2              28               NA            1347              191
## 3              27               NA            1377              137
## 4              30               NA            1396               97
## 5              39               NA            1297              102
## 6              59               NA            1279               92
##   TEAM_PITCHING_BB TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP
## 1              927             5456            1011               NA
## 2              689             1082             193              155
## 3              602              917             175              153
## 4              454              928             164              156
## 5              472              920             138              168
## 6              443              973             123              149
```

```
train_=replace(train_df, TRUE, lapply(train_df, na.aggregate))
summary(train_)
```

```
##      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.: 556.8
##  Median : 47.00   Median :102.00   Median :512.0    Median : 735.6
##  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.: 925.0
##  Max.   :223.00   Max.   :264.00   Max.   :878.0    Max.   :1399.0
##  TEAM_BASERUN_SB  TEAM_BASERUN_CS   TEAM_BATTING_HBP TEAM_PITCHING_H
##  Min.   :  0.0    Min.   :  0.00   Min.   :29.00    Min.   : 1137
##  1st Qu.: 67.0    1st Qu.: 44.00   1st Qu.:59.36    1st Qu.: 1419
##  Median :106.0    Median : 52.80   Median :59.36    Median : 1518
##  Mean   :124.8    Mean   : 52.80   Mean   :59.36    Mean   : 1779
##  3rd Qu.:151.0    3rd Qu.: 54.25   3rd Qu.:59.36    3rd Qu.: 1682
##  Max.   :697.0    Max.   :201.00   Max.   :95.00    Max.   :30132
##  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.: 626.0    1st Qu.: 127.0
##  Median :107.0    Median : 536.5   Median : 817.7    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.:  957.0   3rd Qu.: 249.2
## Max.   :343.0   Max.   :3645.0   Max.   :19278.0   Max.   :1898.0
## TEAM_FIELDING_DP
## Min.   : 52.0
## 1st Qu.:134.0
## Median :146.4
## Mean   :146.4
## 3rd Qu.:161.2
## Max.   :228.0
```

```r
corr=round(cor(train_),2)
corr
```

```
##                   INDEX TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B
## INDEX              1.00       -0.02          -0.02            0.01
## TARGET_WINS       -0.02        1.00           0.39            0.29
## TEAM_BATTING_H    -0.02        0.39           1.00            0.56
## TEAM_BATTING_2B    0.01        0.29           0.56            1.00
## TEAM_BATTING_3B   -0.01        0.14           0.43           -0.11
## TEAM_BATTING_HR    0.05        0.18          -0.01            0.44
## TEAM_BATTING_BB   -0.03        0.23          -0.07            0.26
## TEAM_BATTING_SO    0.08       -0.03          -0.45            0.15
## TEAM_BASERUN_SB    0.04        0.12           0.11           -0.19
## TEAM_BASERUN_CS    0.00        0.02           0.01           -0.07
## TEAM_BATTING_HBP   0.02        0.02           0.00            0.01
## TEAM_PITCHING_H    0.02       -0.11           0.30            0.02
## TEAM_PITCHING_HR   0.05        0.19           0.07            0.45
## TEAM_PITCHING_BB  -0.02        0.12           0.09            0.18
## TEAM_PITCHING_SO   0.05       -0.08          -0.25            0.06
## TEAM_FIELDING_E   -0.01       -0.18           0.26           -0.24
## TEAM_FIELDING_DP   0.02       -0.03           0.12            0.26
##                   TEAM_BATTING_3B TEAM_BATTING_HR TEAM_BATTING_BB
## INDEX                       -0.01            0.05           -0.03
## TARGET_WINS                  0.14            0.18            0.23
## TEAM_BATTING_H               0.43           -0.01           -0.07
## TEAM_BATTING_2B             -0.11            0.44            0.26
## TEAM_BATTING_3B              1.00           -0.64           -0.29
## TEAM_BATTING_HR             -0.64            1.00            0.51
## TEAM_BATTING_BB             -0.29            0.51            1.00
## TEAM_BATTING_SO             -0.66            0.70            0.37
## TEAM_BASERUN_SB              0.50           -0.43           -0.08
## TEAM_BASERUN_CS              0.19           -0.29           -0.08
## TEAM_BATTING_HBP            -0.02            0.02            0.01
## TEAM_PITCHING_H              0.19           -0.25           -0.45
## TEAM_PITCHING_HR            -0.57            0.97            0.46
## TEAM_PITCHING_BB             0.00            0.14            0.49
## TEAM_PITCHING_SO            -0.25            0.18           -0.02
## TEAM_FIELDING_E              0.51           -0.59           -0.66
## TEAM_FIELDING_DP            -0.25            0.41            0.34
##                   TEAM_BATTING_SO TEAM_BASERUN_SB TEAM_BASERUN_CS
## INDEX                        0.08            0.04            0.00
## TARGET_WINS                 -0.03            0.12            0.02
## TEAM_BATTING_H              -0.45            0.11            0.01
```

```
## TEAM_BATTING_2B               0.15           -0.19           -0.07
## TEAM_BATTING_3B              -0.66            0.50            0.19
## TEAM_BATTING_HR               0.70           -0.43           -0.29
## TEAM_BATTING_BB               0.37           -0.08           -0.08
## TEAM_BATTING_SO               1.00           -0.23           -0.16
## TEAM_BASERUN_SB              -0.23            1.00            0.28
## TEAM_BASERUN_CS              -0.16            0.28            1.00
## TEAM_BATTING_HBP              0.03           -0.01           -0.01
## TEAM_PITCHING_H             -0.38            0.06           -0.04
## TEAM_PITCHING_HR             0.64           -0.40           -0.29
## TEAM_PITCHING_BB             0.04            0.12           -0.05
## TEAM_PITCHING_SO             0.42           -0.05           -0.07
## TEAM_FIELDING_E             -0.58            0.37            0.02
## TEAM_FIELDING_DP             0.13           -0.30           -0.14
##                  TEAM_BATTING_HBP TEAM_PITCHING_H TEAM_PITCHING_HR
## INDEX                        0.02            0.02            0.05
## TARGET_WINS                  0.02           -0.11            0.19
## TEAM_BATTING_H               0.00            0.30            0.07
## TEAM_BATTING_2B              0.01            0.02            0.45
## TEAM_BATTING_3B             -0.02            0.19           -0.57
## TEAM_BATTING_HR              0.02           -0.25            0.97
## TEAM_BATTING_BB              0.01           -0.45            0.46
## TEAM_BATTING_SO              0.03           -0.38            0.64
## TEAM_BASERUN_SB             -0.01            0.06           -0.40
## TEAM_BASERUN_CS             -0.01           -0.04           -0.29
## TEAM_BATTING_HBP             1.00            0.00            0.02
## TEAM_PITCHING_H              0.00            1.00           -0.14
## TEAM_PITCHING_HR             0.02           -0.14            1.00
## TEAM_PITCHING_BB             0.01            0.32            0.22
## TEAM_PITCHING_SO             0.01            0.27            0.20
## TEAM_FIELDING_E              0.00            0.67           -0.49
## TEAM_FIELDING_DP            -0.01           -0.06            0.40
##                  TEAM_PITCHING_BB TEAM_PITCHING_SO TEAM_FIELDING_E
## INDEX                       -0.02            0.05           -0.01
## TARGET_WINS                  0.12           -0.08           -0.18
## TEAM_BATTING_H               0.09           -0.25            0.26
## TEAM_BATTING_2B              0.18            0.06           -0.24
## TEAM_BATTING_3B              0.00           -0.25            0.51
## TEAM_BATTING_HR              0.14            0.18           -0.59
## TEAM_BATTING_BB              0.49           -0.02           -0.66
## TEAM_BATTING_SO              0.04            0.42           -0.58
## TEAM_BASERUN_SB              0.12           -0.05            0.37
## TEAM_BASERUN_CS             -0.05           -0.07            0.02
## TEAM_BATTING_HBP             0.01            0.01            0.00
## TEAM_PITCHING_H              0.32            0.27            0.67
## TEAM_PITCHING_HR             0.22            0.20           -0.49
## TEAM_PITCHING_BB             1.00            0.48           -0.02
## TEAM_PITCHING_SO             0.48            1.00           -0.02
## TEAM_FIELDING_E             -0.02           -0.02            1.00
## TEAM_FIELDING_DP             0.19            0.01           -0.25
##                  TEAM_FIELDING_DP
## INDEX                        0.02
## TARGET_WINS                 -0.03
## TEAM_BATTING_H               0.12
```

```
## TEAM_BATTING_2B              0.26
## TEAM_BATTING_3B             -0.25
## TEAM_BATTING_HR              0.41
## TEAM_BATTING_BB              0.34
## TEAM_BATTING_SO              0.13
## TEAM_BASERUN_SB             -0.30
## TEAM_BASERUN_CS             -0.14
## TEAM_BATTING_HBP            -0.01
## TEAM_PITCHING_H             -0.06
## TEAM_PITCHING_HR             0.40
## TEAM_PITCHING_BB             0.19
## TEAM_PITCHING_SO             0.01
## TEAM_FIELDING_E             -0.25
## TEAM_FIELDING_DP             1.00
```

```
p.train=cor_pmat(train_)
head(p.train)
```

```
##                     INDEX   TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B
## INDEX          0.00000000 3.153265e-01    3.928126e-01    5.938698e-01
## TARGET_WINS    0.31532651 0.000000e+00    5.238956e-83    4.585585e-45
## TEAM_BATTING_H 0.39281260 5.238956e-83    0.000000e+00   2.258301e-190
## TEAM_BATTING_2B 0.59386984 4.585585e-45  2.258301e-190    0.000000e+00
## TEAM_BATTING_3B 0.78158537 8.217437e-12  6.818764e-102    2.877545e-07
## TEAM_BATTING_HR 0.01403725 2.551007e-17   7.549934e-01   6.206460e-106
##                TEAM_BATTING_3B TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO
## INDEX             7.815854e-01    1.403725e-02    2.051616e-01    1.418041e-04
## TARGET_WINS       8.217437e-12    2.551007e-17    2.500423e-29    1.434314e-01
## TEAM_BATTING_H   6.818764e-102    7.549934e-01    5.407324e-04   3.102524e-114
## TEAM_BATTING_2B   2.877545e-07   6.206460e-106    2.610613e-35    1.063625e-13
## TEAM_BATTING_3B   0.000000e+00   8.324035e-258    1.753410e-44   9.458864e-282
## TEAM_BATTING_HR  8.324035e-258    0.000000e+00   1.598288e-153    0.000000e+00
##                TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_BATTING_HBP
## INDEX             6.359527e-02    9.827609e-01        0.2689154
## TARGET_WINS       3.961236e-09    4.579803e-01        0.4359570
## TEAM_BATTING_H    4.894146e-08    5.788551e-01        0.8326810
## TEAM_BATTING_2B   6.578135e-20    4.151409e-04        0.7209012
## TEAM_BATTING_3B  6.146188e-145    7.149797e-21        0.4369921
## TEAM_BATTING_HR  3.132324e-102    1.405082e-45        0.4334372
##                TEAM_PITCHING_H TEAM_PITCHING_HR TEAM_PITCHING_BB
## INDEX             4.147525e-01     1.498867e-02     4.660200e-01
## TARGET_WINS       1.457270e-07     9.502088e-20     2.784686e-09
## TEAM_BATTING_H    1.945133e-49     5.045119e-04     6.755492e-06
## TEAM_BATTING_2B   2.585473e-01    1.894857e-116     1.145276e-17
## TEAM_BATTING_3B   6.462227e-21    1.809562e-194     9.155425e-01
## TEAM_BATTING_HR   8.232022e-34     0.000000e+00     5.388223e-11
##                TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP
## INDEX              9.087418e-03    6.597515e-01     3.731054e-01
## TARGET_WINS        2.959322e-04    2.219881e-17     1.689848e-01
## TEAM_BATTING_H     1.407565e-32    7.430104e-38     3.292899e-08
## TEAM_BATTING_2B    3.228077e-03    5.757433e-30     2.962447e-37
## TEAM_BATTING_3B    7.542961e-35   8.242651e-151     1.175411e-32
## TEAM_BATTING_HR    1.820672e-17   3.510100e-211     3.950938e-91
```

```
ggcorrplot(corr, hc.order = TRUE, lab=TRUE, p.mat =p.train, insig = "blank" )
```

```
# Check correlations (as scatterplots), distribution and print corrleation coefficient
ggpairs(train_df, title='correlogram with ggpairs')
```

```
ggplot(train_df, aes(x=TARGET_WINS, y=TEAM_PITCHING_SO)) +
  geom_point(size=2)
```

Warning: Removed 102 rows containing missing values (geom_point).
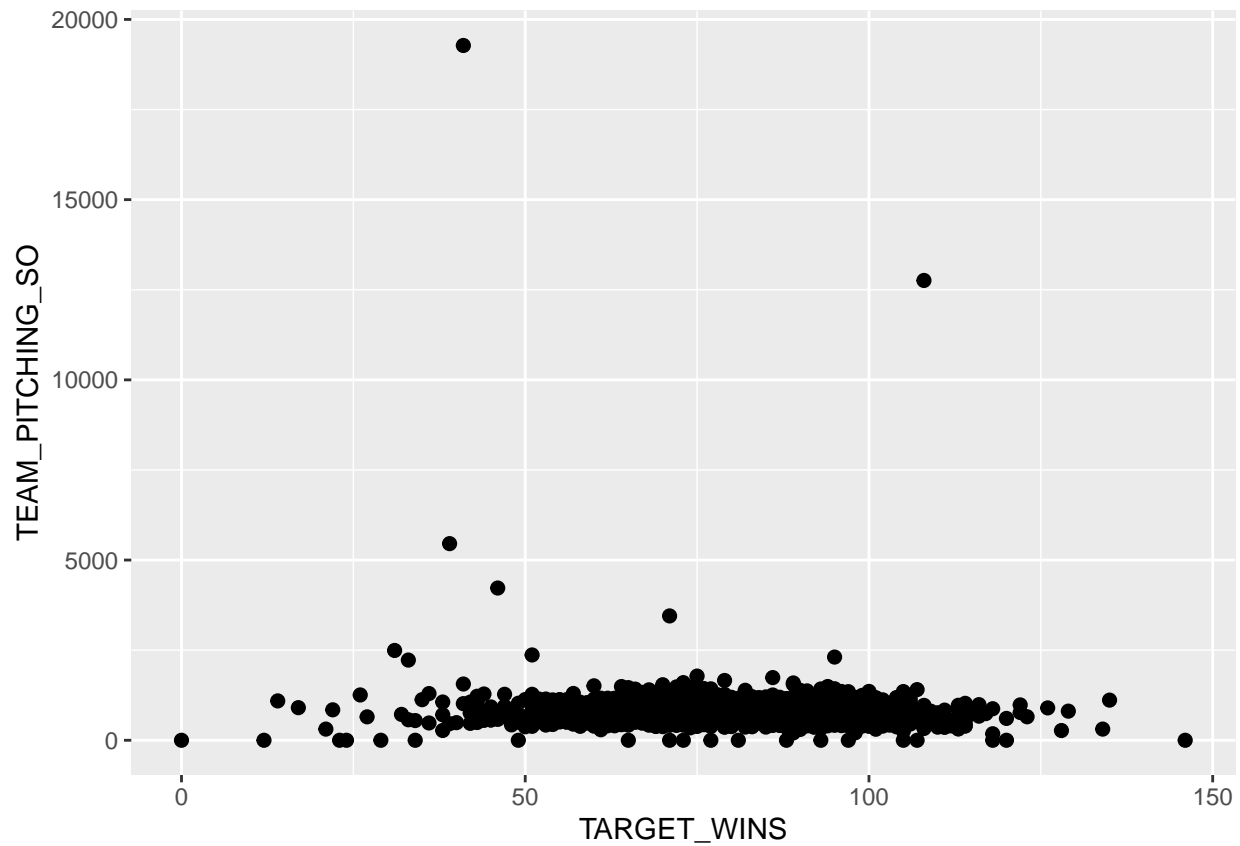

```
## Ewmocving Team Pitching >5000 as max is 1800.
train_df %>%
  filter(`TEAM_PITCHING_SO` > 3000)
```

```
##   INDEX TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## 1     1          39           1445             194              39
## 2   309          71           1325             162              76
## 3  1494         108           1188             338               0
## 4  2051          46           1254             154             127
## 5  2380          41            992             263              20
##   TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO TEAM_BASERUN_SB
## 1              13             143             842              NA
## 2              10             596             724             481
```

7

```
## 3                0              270              945               NA
## 4               27              204             1095               NA
## 5                0              142              952               NA
##    TEAM_BASERUN_CS TEAM_BATTING_HBP TEAM_PITCHING_H TEAM_PITCHING_HR
## 1               NA               NA            9364               84
## 2               NA               NA            6313               48
## 3               NA               NA           16038                0
## 4               NA               NA            4837              104
## 5               NA               NA           20088                0
##    TEAM_PITCHING_BB TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP
## 1               927             5456            1011               NA
## 2              2840             3450             519               NA
## 3              3645            12758             716               NA
## 4               787             4224             918               NA
## 5              2876            19278             952               NA
```

```r
ggplot(train_df, aes(x=TARGET_WINS, y=TEAM_PITCHING_H)) +
  geom_point(size=2)
```
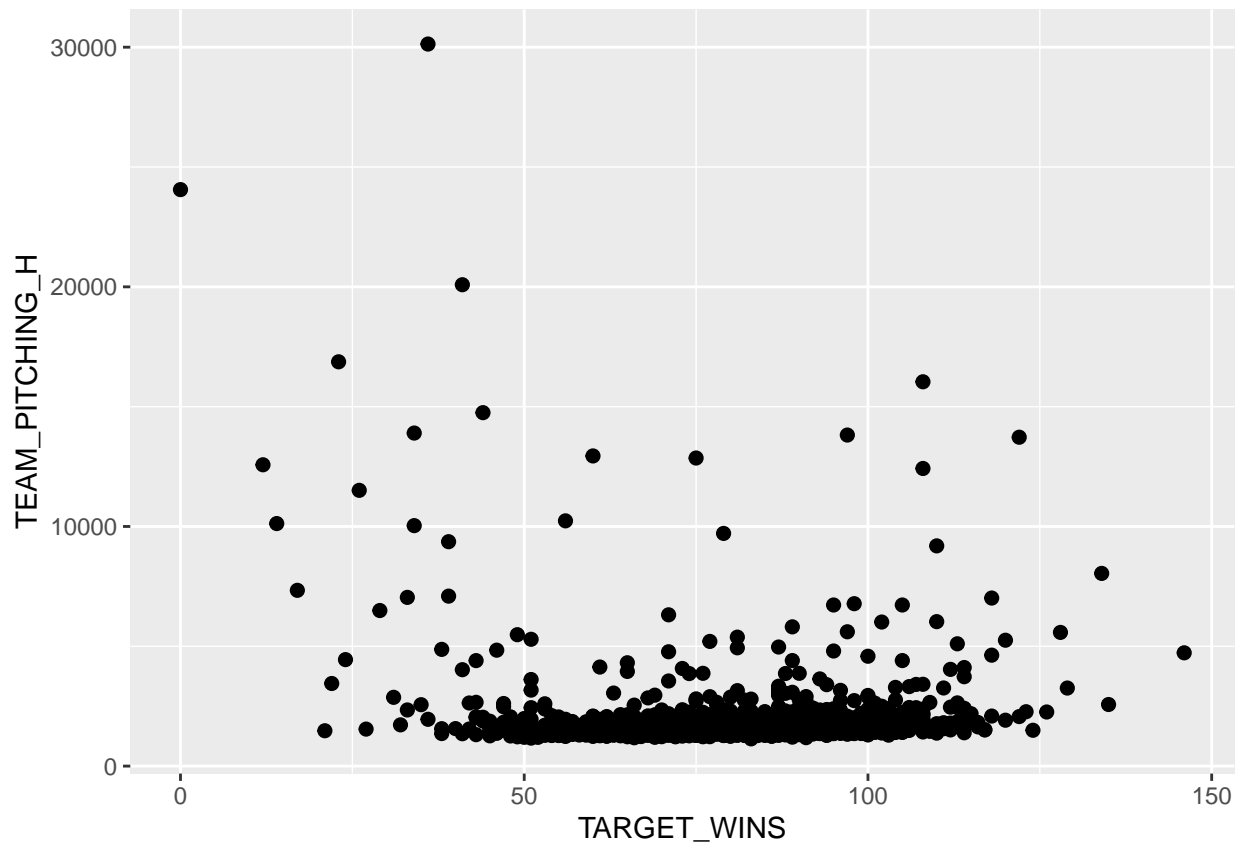


```r
train_df %>%
  filter(`TEAM_PITCHING_H` > 5000)
```

```
##   INDEX TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## 1     1          39           1445             194              39
## 2    59         113           1890             246             100
```

```
## 3     298    105   2241     282   105
## 4     299     98   2343     344   113
## 5     309     71   1325     162    76
## 6     322    108   2300     378   200
## 7     323    134   2333     393   107
## 8     324    118   2554     376   126
## 9     325    120   2270     301   132
## 10    433     39   1620     201    44
## 11    434     51   1764     180    81
## 12    459     23   1458     220    35
## 13    460     56   1832     196   223
## 14    461     44   2003     272    44
## 15    462    128   2273     322    79
## 16    952     77   1895     244     8
## 17   1089     17   1313     145    34
## 18   1106     49   1794     281    58
## 19   1202     60   1518     162    68
## 20   1346     34   2059     209    54
## 21   1347      0    891     135     0
## 22   1492     95   1494     261    68
## 23   1494    108   1188     338     0
## 24   1499     79   1978     211   103
## 25   1500    102   2004     258    39
## 26   1501     89   1901     205   125
## 27   1769     36   1674     216    54
## 28   2034    122   2372     382   156
## 29   2035    110   2496     284    85
## 30   2037     97   1903     256    50
## 31   2047    110   1972     254    61
## 32   2048     81   1927     207   142
## 33   2050     14   1437     148    56
## 34   2053     26   1776     285   162
## 35   2253     34   1177     171     9
## 36   2380     41    992     263    20
## 37   2470     75   2222     295   197
## 38   2471     97   2132     363    71
## 39   2485     33   1695     187   104
## 40   2486     12   1009     112    75
## 41   2493     29   1122      69    64
##      TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO TEAM_BASERUN_SB
## 1                 13             143             842              NA
## 2                 11             176             327              NA
## 3                 42              81              84             105
## 4                 35             116              72              61
## 5                 10             596             724             481
## 6                 16             324             103             394
## 7                 24              97              90             162
## 8                 36             170              66             107
## 9                 42              74               0              NA
## 10                 0              83             105              74
## 11                18             159             129              54
## 12                 0              93               0              NA
## 13                39             145             140             101
## 14                 0             125              96              88
```

```
## 15            20         172         110          NA
## 16             8          93           0          NA
## 17             0          78         162          45
## 18             6          79           0          NA
## 19            17         281          77         136
## 20             7          34          81          34
## 21             0           0           0           0
## 22            59         482         513         212
## 23             0         270         945          NA
## 24             5         162          74         226
## 25            12         165         156         168
## 26            15         128          67          46
## 27             0          72          72           0
## 28            52         266         133         324
## 29            15         254         173         214
## 30            18          71           0          NA
## 31            24         190         119         134
## 32             8          78           0          NA
## 33             0          56         155          14
## 34            19         246         194         343
## 35             0         119           0          NA
## 36             0         142         952          NA
## 37            35         284         110         359
## 38            32          45          91          58
## 39            12          79         137          21
## 40             0          12           0          NA
## 41             0          29           0          NA
##    TEAM_BASERUN_CS TEAM_BATTING_HBP TEAM_PITCHING_H TEAM_PITCHING_HR
## 1               NA               NA            9364              84
## 2               NA               NA            5103              30
## 3               45               NA            6723             126
## 4               29               NA            6778             101
## 5               NA               NA            6313              48
## 6               NA               NA           12420              86
## 7               48               NA            8041              83
## 8               74               NA            7013              99
## 9               NA               NA            5253              97
## 10              61               NA            7093               0
## 11              27               NA            5292              54
## 12              NA               NA           16871               0
## 13              NA               NA           10234             218
## 14              22               NA           14749               0
## 15              NA               NA            5579              49
## 16              NA               NA            5203              22
## 17              22               NA            7335               0
## 18              NA               NA            5484              18
## 19              NA               NA           12943             145
## 20              20               NA           13898              47
## 21               0               NA           24057               0
## 22              NA               NA            6723             266
## 23              NA               NA           16038               0
## 24              NA               NA            9710              25
## 25              63               NA            6012              36
## 26              15               NA            5811              46
```

```
## 27           54           NA        30132            0
## 28           NA           NA        13724          301
## 29          118           NA         9190           55
## 30           NA           NA         5605           53
## 31           43           NA         6028           73
## 32           NA           NA         5382           22
## 33            7           NA        10121            0
## 34           NA           NA        11508          123
## 35           NA           NA        10035            0
## 36           NA           NA        20088            0
## 37           NA           NA        12856          203
## 38           45           NA        13815          207
## 39           21           NA         7041           50
## 40           NA           NA        12574            0
## 41           NA           NA         6492            0
##    TEAM_PITCHING_BB TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP
## 1               927             5456            1011               NA
## 2               475              883             780               97
## 3               243              252            1296               NA
## 4               336              208            1059               NA
## 5              2840             3450             519               NA
## 6              1750              556            1215               NA
## 7               334              310             907               NA
## 8               467              181            1192               NA
## 9               171                0            1058               NA
## 10              363              460            1567               NA
## 11              477              387            1515               NA
## 12             1076                0            1898               NA
## 13              810              782            1246               NA
## 14              920              707            1237               NA
## 15              422              270             692               81
## 16              255                0            1225               NA
## 17              436              905            1553               NA
## 18              241                0            1531               NA
## 19             2396              657            1390               NA
## 20              230              547            1512               NA
## 21                0                0            1890               NA
## 22             2169             2309             522               NA
## 23             3645            12758             716               NA
## 24              795              363            1114               NA
## 25              495              468             978               NA
## 26              391              205            1281               NA
## 27             1296             1296            1728               NA
## 28             1539              770            1122               NA
## 29              935              637            1097               NA
## 30              209                0            1166               NA
## 31              581              364            1103               NA
## 32              218                0            1447               NA
## 33              394             1092            1740               NA
## 34             1594             1257            1426               NA
## 35             1015                0            1279               NA
## 36             2876            19278             952               NA
## 37             1643              636            1146               NA
## 38              292              590             933               NA
```

```
## 39               328             569            997            NA
## 40               150               0            847            NA
## 41               168               0           1522            NA
```

```
train_df$index_cat<-cut(train_df$INDEX, c(0,250,500,750,1000,1250,1500,1750,2000,2250,2500,2750,3000))
```

#train_df$group <- as.factor(ifelse(train_df$TEAM_BATTING_HBP==59.9), 'A','B')

```
ggplot(train_df, aes(x=TARGET_WINS, y=TEAM_PITCHING_H, color=index_cat)) +
  geom_point(size=2)
```