

baseball4.R

Andrew

2021-10-19

```
library(readxl)
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.0.5
```

```
#mydata <- read_excel("~/NU Q4/Datasets/MoneyBall.xlsx")
allmydata <- read_excel("~/NU Q4/Datasets/MoneyBall.xlsx")
mydata <- subset(allmydata, select = -c(INDEX, TEAM_BATTING_HBP, TEAM_BASERUN_CS))
mydata <- mydata[complete.cases(mydata),]
head(mydata)
```

```
## # A tibble: 6 x 14
##   TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B TEAM_BATTING_HR
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1         70         1339         219           22          190
## 2         86         1377         232           35          137
## 3         70         1387         209           38           96
## 4         82         1297         186           27          102
## 5         75         1279         200           36           92
## 6         80         1244         179           54          122
## # ... with 9 more variables: TEAM_BATTING_BB <dbl>, TEAM_BATTING_SO <dbl>,
## #   TEAM_BASERUN_SB <dbl>, TEAM_PITCHING_H <dbl>, TEAM_PITCHING_HR <dbl>,
## #   TEAM_PITCHING_BB <dbl>, TEAM_PITCHING_SO <dbl>, TEAM_FIELDING_E <dbl>,
## #   TEAM_FIELDING_DP <dbl>
```

```
str(mydata)
```

```
## tibble [1,835 x 14] (S3: tbl_df/tbl/data.frame)
##  $ TARGET_WINS      : num [1:1835] 70 86 70 82 75 80 85 86 76 78 ...
##  $ TEAM_BATTING_H    : num [1:1835] 1339 1377 1387 1297 1279 ...
##  $ TEAM_BATTING_2B   : num [1:1835] 219 232 209 186 200 179 171 197 213 179 ...
##  $ TEAM_BATTING_3B   : num [1:1835] 22 35 38 27 36 54 37 40 18 27 ...
##  $ TEAM_BATTING_HR   : num [1:1835] 190 137 96 102 92 122 115 114 96 82 ...
##  $ TEAM_BATTING_BB   : num [1:1835] 685 602 451 472 443 525 456 447 441 374 ...
##  $ TEAM_BATTING_SO   : num [1:1835] 1075 917 922 920 973 ...
```

```
## $ TEAM_BASERUN_SB : num [1:1835] 37 46 43 49 107 80 40 69 72 60 ...
## $ TEAM_PITCHING_H : num [1:1835] 1347 1377 1396 1297 1279 ...
## $ TEAM_PITCHING_HR: num [1:1835] 191 137 97 102 92 122 116 114 96 86 ...
## $ TEAM_PITCHING_BB: num [1:1835] 689 602 454 472 443 525 459 447 441 391 ...
## $ TEAM_PITCHING_SO: num [1:1835] 1082 917 928 920 973 ...
## $ TEAM_FIELDING_E : num [1:1835] 193 175 164 138 123 136 112 127 131 119 ...
## $ TEAM_FIELDING_DP: num [1:1835] 155 153 156 168 149 186 136 169 159 141 ...
```

```
cor(mydata)
```

```
##          TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## TARGET_WINS      1.00000000      0.35220269      0.21333408      0.12271778
## TEAM_BATTING_H    0.35220269      1.00000000      0.66299786      0.36894288
## TEAM_BATTING_2B   0.21333408      0.66299786      1.00000000     -0.05099146
## TEAM_BATTING_3B   0.12271778      0.36894288     -0.05099146      1.00000000
## TEAM_BATTING_HR   0.22013777      0.11173161      0.37534805     -0.63558893
## TEAM_BATTING_BB   0.30158847      0.10120528      0.22955389     -0.23770483
## TEAM_BATTING_SO  -0.05947441     -0.35068211      0.12196187     -0.70107395
## TEAM_BASERUN_SB   0.12135320     -0.05678718     -0.11392914      0.26024771
## TEAM_PITCHING_H   0.22123595      0.70590736      0.39861718      0.40396409
## TEAM_PITCHING_HR  0.21973105      0.12702351      0.37873992     -0.61378279
## TEAM_PITCHING_BB  0.27172463      0.12836002      0.19621520     -0.12146732
## TEAM_PITCHING_SO  -0.06555941     -0.33325343      0.11651495     -0.65505668
## TEAM_FIELDING_E   -0.18482328      0.08436548     -0.28145551      0.67497666
## TEAM_FIELDING_DP  -0.03664517      0.15747196      0.17745164     -0.21630359
##          TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO
## TARGET_WINS      0.22013777      0.301588472     -0.05947441
## TEAM_BATTING_H    0.11173161      0.101205279     -0.35068211
## TEAM_BATTING_2B   0.37534805      0.229553888      0.12196187
## TEAM_BATTING_3B   -0.63558893     -0.237704830     -0.70107395
## TEAM_BATTING_HR   1.00000000      0.391503214      0.69703956
## TEAM_BATTING_BB   0.39150321      1.000000000      0.18080074
## TEAM_BATTING_SO   0.69703956      0.180800745      1.00000000
## TEAM_BASERUN_SB  -0.30794334     -0.153218448     -0.01308392
## TEAM_PITCHING_H   -0.09909496      0.002068846     -0.41457430
## TEAM_PITCHING_HR   0.98124381      0.392183897      0.66865912
## TEAM_PITCHING_BB   0.25857882      0.886916177      0.05211811
## TEAM_PITCHING_SO   0.64983923      0.167252900      0.94844880
## TEAM_FIELDING_E   -0.70072212     -0.339455367     -0.62971400
## TEAM_FIELDING_DP   0.28776499      0.256928398      0.05934322
##          TEAM_BASERUN_SB TEAM_PITCHING_H TEAM_PITCHING_HR
## TARGET_WINS      0.12135320      0.221235948      0.21973105
## TEAM_BATTING_H    -0.05678718      0.705907360      0.12702351
## TEAM_BATTING_2B   -0.11392914      0.398617185      0.37873992
## TEAM_BATTING_3B   0.26024771      0.403964092     -0.61378279
## TEAM_BATTING_HR   -0.30794334     -0.099094956      0.98124381
## TEAM_BATTING_BB   -0.15321845      0.002068846      0.39218390
## TEAM_BATTING_SO   -0.01308392     -0.414574297      0.66865912
## TEAM_BASERUN_SB   1.00000000      0.063500728     -0.30028587
## TEAM_PITCHING_H    0.06350073      1.000000000      0.03559024
## TEAM_PITCHING_HR  -0.30028587      0.035590239      1.00000000
## TEAM_PITCHING_BB  -0.08776779      0.350098202      0.34222423
## TEAM_PITCHING_SO   0.01793924     -0.182083863      0.68070952
## TEAM_FIELDING_E    0.34608860      0.254751858     -0.68192930
```

```
## TEAM_FIELDING_DP      -0.39900458      0.032537334      0.28860627
##                      TEAM_PITCHING_BB TEAM_PITCHING_SO TEAM_FIELDING_E
## TARGET_WINS           0.27172463      -0.06555941      -0.18482328
## TEAM_BATTING_H        0.12836002      -0.33325343      0.08436548
## TEAM_BATTING_2B       0.19621520      0.11651495      -0.28145551
## TEAM_BATTING_3B      -0.12146732      -0.65505668      0.67497666
## TEAM_BATTING_HR       0.25857882      0.64983923      -0.70072212
## TEAM_BATTING_BB       0.88691618      0.16725290      -0.33945537
## TEAM_BATTING_SO       0.05211811      0.94844880      -0.62971400
## TEAM_BASERUN_SB      -0.08776779      0.01793924      0.34608860
## TEAM_PITCHING_H       0.35009820      -0.18208386      0.25475186
## TEAM_PITCHING_HR      0.34222423      0.68070952      -0.68192930
## TEAM_PITCHING_BB      1.00000000      0.18441026      -0.20232668
## TEAM_PITCHING_SO      0.18441026      1.00000000      -0.57550044
## TEAM_FIELDING_E      -0.20232668      -0.57550044      1.00000000
## TEAM_FIELDING_DP      0.20245071      0.04305787      -0.36359184
##                      TEAM_FIELDING_DP
## TARGET_WINS          -0.03664517
## TEAM_BATTING_H        0.15747196
## TEAM_BATTING_2B       0.17745164
## TEAM_BATTING_3B      -0.21630359
## TEAM_BATTING_HR       0.28776499
## TEAM_BATTING_BB       0.25692840
## TEAM_BATTING_SO       0.05934322
## TEAM_BASERUN_SB      -0.39900458
## TEAM_PITCHING_H       0.03253733
## TEAM_PITCHING_HR      0.28860627
## TEAM_PITCHING_BB      0.20245071
## TEAM_PITCHING_SO      0.04305787
## TEAM_FIELDING_E      -0.36359184
## TEAM_FIELDING_DP      1.00000000
```

```
colSums(is.na(mydata))
```

```
##      TARGET_WINS  TEAM_BATTING_H  TEAM_BATTING_2B  TEAM_BATTING_3B
##           0           0           0           0
## TEAM_BATTING_HR  TEAM_BATTING_BB  TEAM_BATTING_SO  TEAM_BASERUN_SB
##           0           0           0           0
## TEAM_PITCHING_H  TEAM_PITCHING_HR  TEAM_PITCHING_BB  TEAM_PITCHING_SO
##           0           0           0           0
## TEAM_FIELDING_E  TEAM_FIELDING_DP
##           0           0
```

```
corrplot(cor(mydata[complete.cases(mydata),]))
```

baseball4_files/figure-latex/unnamed-chunk-1-1.pdf

```
summary(allmydata)
```

```
##      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
```

```
summary(mydata)
```

```
##      TARGET_WINS      TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## Min.   : 38.00    Min.   :1137    Min.   :130.0    Min.   : 11.00
## 1st Qu.: 72.00    1st Qu.:1386    1st Qu.:216.0    1st Qu.: 32.00
## Median : 82.00    Median :1448    Median :245.0    Median : 42.00
## Mean   : 80.99    Mean   :1457    Mean   :247.8    Mean   : 47.84
## 3rd Qu.: 90.00    3rd Qu.:1524    3rd Qu.:277.0    3rd Qu.: 60.00
## Max.   :120.00    Max.   :1876    Max.   :392.0    Max.   :138.00
```

```
## TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO TEAM_BASERUN_SB
## Min. : 4.0 Min. :273.0 Min. : 324.0 Min. : 18.0
## 1st Qu.: 77.0 1st Qu.:473.0 1st Qu.: 602.5 1st Qu.: 62.0
## Median :119.0 Median :525.0 Median : 816.0 Median : 91.0
## Mean :116.5 Mean :530.3 Mean : 786.8 Mean :100.8
## 3rd Qu.:156.0 3rd Qu.:586.0 3rd Qu.: 956.5 3rd Qu.:131.0
## Max. :264.0 Max. :878.0 Max. :1399.0 Max. :367.0
## TEAM_PITCHING_H TEAM_PITCHING_HR TEAM_PITCHING_BB TEAM_PITCHING_SO
## Min. :1137 Min. : 4.0 Min. : 325 Min. : 345.0
## 1st Qu.:1409 1st Qu.: 80.0 1st Qu.: 488 1st Qu.: 641.0
## Median :1493 Median :123.0 Median : 542 Median : 829.0
## Mean :1525 Mean :120.7 Mean : 554 Mean : 818.0
## 3rd Qu.:1602 3rd Qu.:160.0 3rd Qu.: 609 3rd Qu.: 972.5
## Max. :2394 Max. :343.0 Max. :1090 Max. :1781.0
## TEAM_FIELDING_E TEAM_FIELDING_DP
## Min. : 65.0 Min. : 72.0
## 1st Qu.:122.0 1st Qu.:136.0
## Median :143.0 Median :151.0
## Mean :159.9 Mean :150.4
## 3rd Qu.:181.5 3rd Qu.:165.0
## Max. :515.0 Max. :228.0
```

```
summary(mydata$TARGET_WINS)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 38.00   72.00   82.00   80.99   90.00  120.00
```

```
hist(mydata$TARGET_WINS)
```

baseball4_files/figure-latex/unnamed-chunk-1-2.pdf

```
boxplot(mydata$TARGET_WINS)
```

baseball4_files/figure-latex/unnamed-chunk-1-3.pdf

```
summary(lm(TARGET_WINS~TEAM_BATTING_H, data=mydata))
```

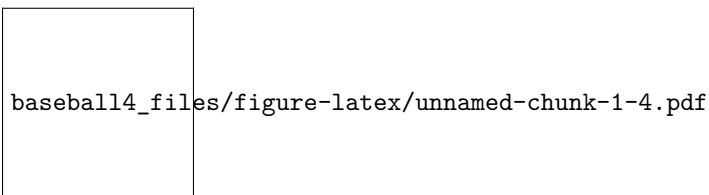
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H, data = mydata)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.955  -8.237   0.426   8.798  36.783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.514395   3.888385   4.761 2.07e-06 ***
## TEAM_BATTING_H  0.042879   0.002661  16.111 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.32 on 1833 degrees of freedom
## Multiple R-squared:  0.124, Adjusted R-squared:  0.1236
## F-statistic: 259.6 on 1 and 1833 DF, p-value: < 2.2e-16
```

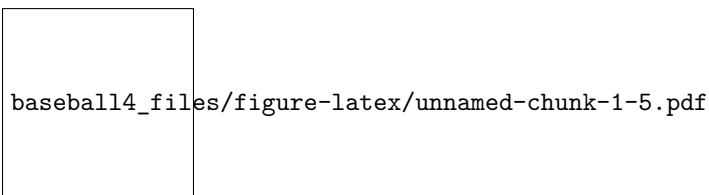
```
summary(mydata$TEAM_BATTING_H)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      1137    1386    1448    1457    1524    1876
```

```
hist(mydata$TEAM_BATTING_H)
```



```
boxplot(mydata$TEAM_BATTING_H)
```



```
summary(lm(TARGET_WINS ~ TEAM_BATTING_2B, data=mydata))
```

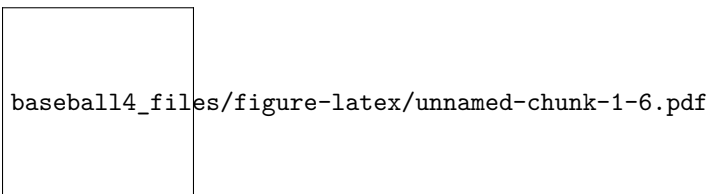
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_2B, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40.936  -8.936   0.502   9.175  40.829
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  64.792891   1.758366  36.848 <2e-16 ***
```

```
## TEAM_BATTING_2B 0.065355 0.006991 9.349 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.86 on 1833 degrees of freedom
## Multiple R-squared: 0.04551, Adjusted R-squared: 0.04499
## F-statistic: 87.4 on 1 and 1833 DF, p-value: < 2.2e-16
```

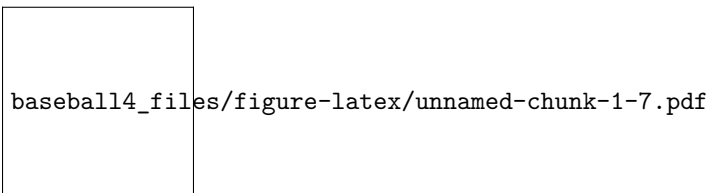
```
summary(mydata$TEAM_BATTING_2B)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    130.0  216.0   245.0   247.8   277.0   392.0
```

```
hist(mydata$TEAM_BATTING_2B)
```



```
boxplot(mydata$TEAM_BATTING_2B)
```



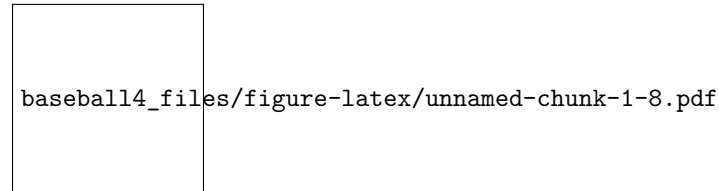
```
summary(lm(TARGET_WINS ~ TEAM_BATTING_3B, data=mydata))
```

```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_3B, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -44.564  -8.630   0.684   9.279  36.518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    77.43189    0.73808  104.910 < 2e-16 ***
## TEAM_BATTING_3B  0.07438    0.01405   5.294 1.34e-07 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.06 on 1833 degrees of freedom
## Multiple R-squared: 0.01506, Adjusted R-squared: 0.01452
## F-statistic: 28.03 on 1 and 1833 DF, p-value: 1.341e-07
```

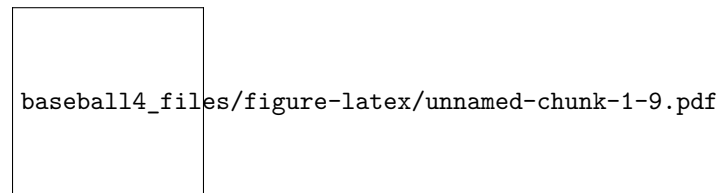
```
summary(mydata$TEAM_BATTING_3B)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    11.00   32.00   42.00   47.84   60.00  138.00
```

```
hist(mydata$TEAM_BATTING_3B)
```



```
boxplot(mydata$TEAM_BATTING_3B)
```



```
summary(lm(TARGET_WINS ~ TEAM_BATTING_HR, data=mydata))
```

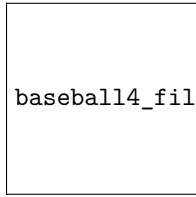
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_HR, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.293  -8.957   0.513   9.144  44.083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    74.80070     0.70726  105.762  <2e-16 ***
## TEAM_BATTING_HR  0.05314     0.00550   9.662  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.84 on 1833 degrees of freedom
## Multiple R-squared:  0.04846,    Adjusted R-squared:  0.04794
## F-statistic: 93.35 on 1 and 1833 DF,  p-value: < 2.2e-16
```

```
summary(mydata$TEAM_BATTING_HR)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       4.0    77.0   119.0   116.5   156.0   264.0
```

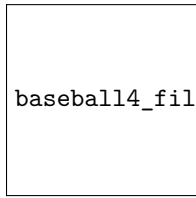


```
hist(mydata$TEAM_BATTING_HR)
```



baseball4_files/figure-latex/unnamed-chunk-1-10.pdf

```
boxplot(mydata$TEAM_BATTING_HR)
```



baseball4_files/figure-latex/unnamed-chunk-1-11.pdf