621 - Homework 1

2024-02-16

```
##      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
```

Part 1: Data Exploration

The money training dataset contains 2,276 variables and 17 elements, including the INDEX or identifying variable, TARGET_WINS, the response variable, and 15 explanatory variables. The variable types are

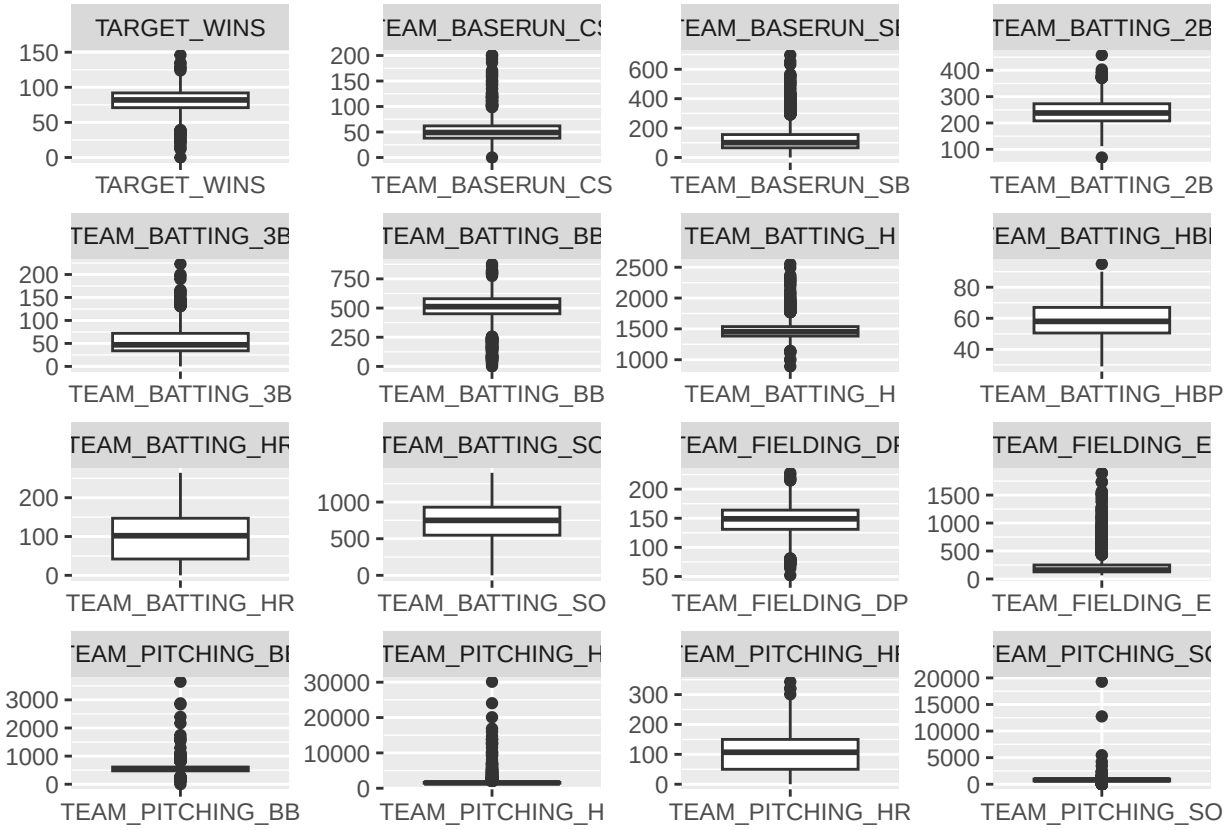
numeric values represented as integers. Six variables have missing values and will be imputed in part two. The variable with the most missing variables is TEAM_BATTING_HBP (batters hit by pitch).

```
## [1] 2276   17
```

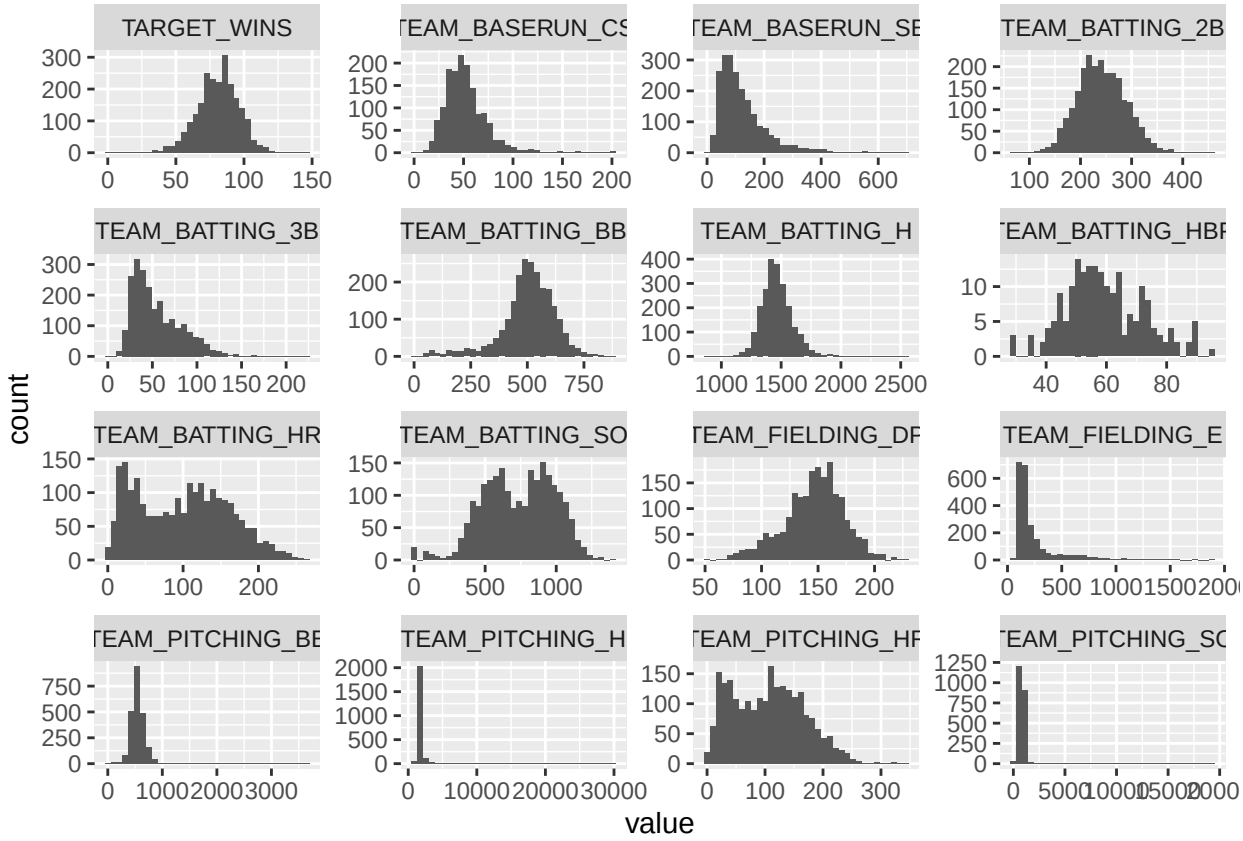
```
##             INDEX      TARGET_WINS   TEAM_BATTING_H  TEAM_BATTING_2B
##                 0                0                0                0
##  TEAM_BATTING_3B  TEAM_BATTING_HR  TEAM_BATTING_BB  TEAM_BATTING_SO
##                 0                0                0              102
##  TEAM_BASERUN_SB  TEAM_BASERUN_CS TEAM_BATTING_HBP  TEAM_PITCHING_H
##               131              772             2085                0
## TEAM_PITCHING_HR TEAM_PITCHING_BB TEAM_PITCHING_SO  TEAM_FIELDING_E
##                 0                0              102                0
## TEAM_FIELDING_DP
##              286
```

The summary data also states the mean and median for reference, many variables have similar values, and when the mean and median are very different it suggests the dataset is skewed. The Box plots below visualize these values and show the spread of the data. The boxplots also show many outliers in most of the variables. Additionally, histogram charts can be used to view the skewness of the variables. Both these visuals confirm that the response variable seems to have a normal distribution. While, many variables seem to have a right skew, and some have a bimodal distribution.

```
## Warning: Removed 3478 rows containing non-finite values (`stat_boxplot()`).
```



```
## Warning: Removed 3478 rows containing non-finite values (‘stat_bin()‘).
```

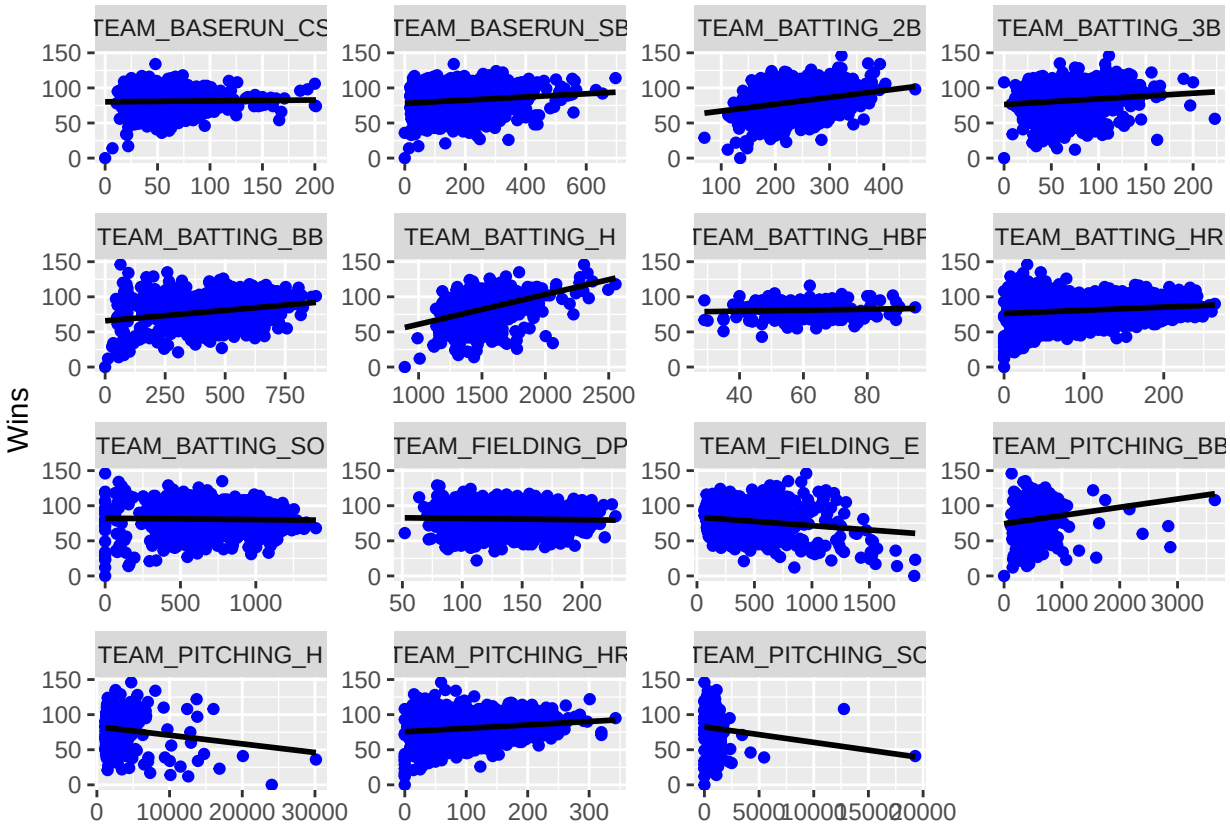


The target variable is the response variable ‘TARGET_WINS’. The scatter plots below show the visual of all other variables, the spread of their data and the correlation line to observe the relationship to the target variable. The sorted_correlation lists the correlation metrics from highest to lowest, indicating these variables are the most positively correlated with results of .46 or above TEAM_PITCHING_H, TEAM_BATTING_H, TEAM_BATTING_BB, and TEAM_PITCHING_BB. To assess if variables are correlated to other variables in the dataset a correlation matrix, with the coefficients added, is also shown below. This will come in handy when trying to avoid collinearity.

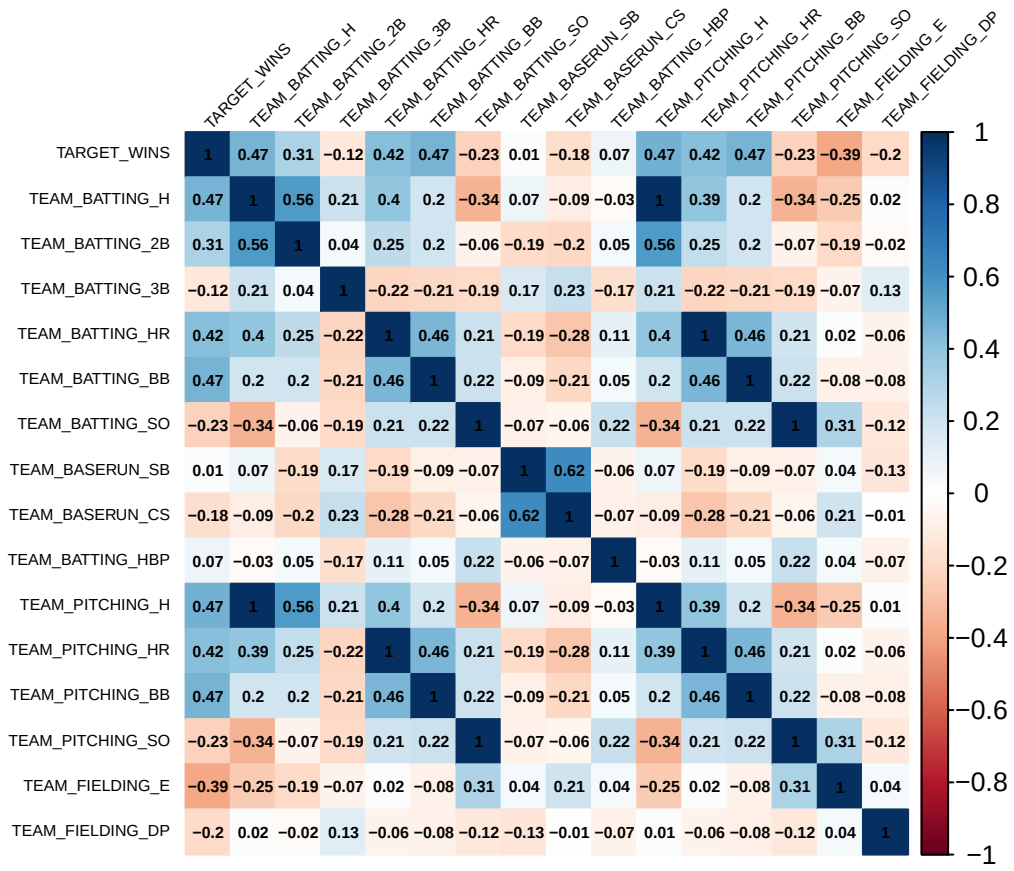
```
## ‘geom_smooth()‘ using formula = ’y ~ x’
```

```
## Warning: Removed 3478 rows containing non-finite values (‘stat_smooth()‘).
```

```
## Warning: Removed 3478 rows containing missing values (‘geom_point()‘).
```




```
##      TARGET_WINS  TEAM_PITCHING_H   TEAM_BATTING_H  TEAM_BATTING_BB
##        1.00000000       0.47123431       0.46994665       0.46868793
## TEAM_PITCHING_BB TEAM_PITCHING_HR  TEAM_BATTING_HR  TEAM_BATTING_2B
##        0.46839882       0.42246683       0.42241683       0.31298400
## TEAM_BATTING_HBP  TEAM_BASERUN_SB  TEAM_BATTING_3B  TEAM_BASERUN_CS
##        0.07350424       0.01483639      -0.12434586      -0.17875598
## TEAM_FIELDING_DP  TEAM_BATTING_SO TEAM_PITCHING_SO  TEAM_FIELDING_E
##       -0.19586601      -0.22889273      -0.22936481      -0.38668800
```



Part 2: Median Imputations for missing variables

To compensate for missingness, we will perform a simple median imputation. We will use the median instead of the mean as some variables in this dataset are heavily skewed.

```
##   TARGET_WINS      TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
##  Min.   :  0.00   Min.   : 891   Min.   : 69.0   Min.   :  0.00
##  1st Qu.: 71.00   1st Qu.:1383   1st Qu.:208.0   1st Qu.: 34.00
##  Median : 82.00   Median :1454   Median :238.0   Median : 47.00
##  Mean   : 80.79   Mean   :1469   Mean   :241.2   Mean   : 55.25
##  3rd Qu.: 92.00   3rd Qu.:1537   3rd Qu.:273.0   3rd Qu.: 72.00
##  Max.   :146.00   Max.   :2554   Max.   :458.0   Max.   :223.00
##  TEAM_BATTING_HR  TEAM_BATTING_BB TEAM_BATTING_SO  TEAM_BASERUN_SB
##  Min.   :  0.00   Min.   :  0.0   Min.   :   0.0   Min.   :  0.0
##  1st Qu.: 42.00   1st Qu.:451.0   1st Qu.: 556.8   1st Qu.: 67.0
##  Median :102.00   Median :512.0   Median : 750.0   Median :101.0
##  Mean   : 99.61   Mean   :501.6   Mean   : 736.3   Mean   :123.4
##  3rd Qu.:147.00   3rd Qu.:580.0   3rd Qu.: 925.0   3rd Qu.:151.0
##  Max.   :264.00   Max.   :878.0   Max.   :1399.0   Max.   :697.0
##  TEAM_BASERUN_CS  TEAM_BATTING_HBP TEAM_PITCHING_H TEAM_PITCHING_HR
##  Min.   :  0.00   Min.   :29.00    Min.   : 1137   Min.   :  0.0
##  1st Qu.: 44.00   1st Qu.:58.00    1st Qu.: 1419   1st Qu.: 50.0
##  Median : 49.00   Median :58.00    Median : 1518   Median :107.0
##  Mean   : 51.51   Mean   :58.11    Mean   : 1779   Mean   :105.7
##  3rd Qu.: 54.25   3rd Qu.:58.00    3rd Qu.: 1682   3rd Qu.:150.0
```

```
##  Max.   :201.00   Max.   :95.00    Max.   :30132   Max.   :343.0
##  TEAM_PITCHING_BB TEAM_PITCHING_SO  TEAM_FIELDING_E  TEAM_FIELDING_DP
##  Min.   :   0.0   Min.   :    0.0   Min.   :  65.0   Min.   : 52.0
##  1st Qu.: 476.0   1st Qu.:  626.0   1st Qu.: 127.0   1st Qu.:134.0
##  Median : 536.5   Median :  813.5   Median : 159.0   Median :149.0
##  Mean   : 553.0   Mean   :  817.5   Mean   : 246.5   Mean   :146.7
##  3rd Qu.: 611.0   3rd Qu.:  957.0   3rd Qu.: 249.2   3rd Qu.:161.2
##  Max.   :3645.0   Max.   :19278.0   Max.   :1898.0   Max.   :228.0
```

```
##      money_train_comp
## 1                 842
## 2                1075
## 3                 917
## 4                 922
## 5                 920
## 6                 973
## 7                1062
## 8                1027
## 9                 922
## 10                827
## 11                888
## 12                801
## 13                816
## 14                812
## 15                880
## 16                682
## 17                843
## 18                900
## 19                841
## 20                760
## 21                835
## 22                928
## 23                902
## 24                860
## 25                926
## 26                819
## 27               1011
## 28               1000
## 29                928
## 30                882
## 31                930
## 32               1007
## 33                993
## 34                980
## 35                953
## 36               1028
## 37               1022
## 38               1024
## 39               1001
## 40                805
## 41                838
## 42                942
## 43                848
## 44               1239
```

```
## 45              1045
## 46               975
## 47              1052
## 48              1016
## 49              1006
## 50              1022
## 51              1094
## 52               965
## 53                99
## 54               227
## 55               327
## 56               428
## 57               426
## 58               471
## 59               699
## 60               963
## 61               755
## 62               744
## 63               525
## 64               633
## 65               570
## 66               627
## 67               632
## 68               367
## 69               320
## 70               292
## 71               339
## 72               322
## 73               329
## 74               287
## 75               326
## 76               569
## 77                NA
## 78                NA
## 79                NA
## 80                NA
## 81                NA
## 82                NA
## 83                NA
## 84               572
## 85               619
## 86               731
## 87               687
## 88               653
## 89               661
## 90               622
## 91               572
## 92               561
## 93               520
## 94               498
## 95               478
## 96               425
## 97               478
## 98               402
```

```
## 99              371
## 100             382
## 101             399
## 102             454
## 103             418
## 104             452
## 105             522
## 106             450
## 107             472
## 108             462
## 109             612
## 110             754
## 111             530
## 112             619
## 113             555
## 114             645
## 115             535
## 116             544
## 117             496
## 118             526
## 119             568
## 120             690
## 121             648
## 122             649
## 123             753
## 124             651
## 125             773
## 126             767
## 127             794
## 128             834
## 129             926
## 130             975
## 131             954
## 132             825
## 133             976
## 134             913
## 135             947
## 136             782
## 137             736
## 138             810
## 139             875
## 140             772
## 141             764
## 142             811
## 143             876
## 144             874
## 145             828
## 146             905
## 147             825
## 148             869
## 149             847
## 150             896
## 151             849
## 152             910
```

```
## 153              839
## 154              859
## 155             1010
## 156              906
## 157              924
## 158              946
## 159              949
## 160             1050
## 161             1032
## 162             1160
## 163             1062
## 164              962
## 165             1010
## 166             1039
## 167              933
## 168             1158
## 169             1084
## 170             1169
## 171              454
## 172              390
## 173              628
## 174              649
## 175               NA
## 176               NA
## 177               NA
## 178               NA
## 179               NA
## 180               NA
## 181               NA
## 182              814
## 183              914
## 184              805
## 185              673
## 186              568
## 187              451
## 188              516
## 189              359
## 190              428
## 191              401
## 192              451
## 193              372
## 194              489
## 195              445
## 196              504
## 197              579
## 198              610
## 199              555
## 200              597
## 201              673
## 202              598
## 203              668
## 204              536
## 205              637
## 206              675
```

```
## 207                581
## 208                689
## 209                635
## 210                595
## 211                750
## 212                698
## 213                606
## 214                736
## 215                783
## 216                729
## 217                757
## 218                667
## 219                781
## 220                763
## 221                745
## 222                774
## 223                726
## 224                843
## 225                902
## 226                931
## 227                940
## 228               1019
## 229                907
## 230                938
## 231               1008
## 232               1019
## 233                806
## 234                952
## 235                865
## 236                984
## 237                752
## 238                770
## 239                850
## 240                883
## 241                951
## 242                869
## 243                863
## 244                766
## 245                700
## 246                796
## 247                800
## 248                884
## 249                914
## 250                862
## 251                939
## 252                874
## 253                957
## 254                968
## 255                974
## 256                827
## 257                930
## 258                947
## 259                903
## 260                915
```

```
## 261  952
## 262  903
## 263  890
## 264  900
## 265  995
## 266  993
## 267  949
## 268  902
## 269  450
## 270  816
## 271  626
## 272   84
## 273   72
## 274  477
## 275  559
## 276  833
## 277  786
## 278  746
## 279  401
## 280  566
## 281  643
## 282  724
## 283  664
## 284  529
## 285  403
## 286  253
## 287  303
## 288  252
## 289  319
## 290  344
## 291  419
## 292  620
## 293  674
## 294 1006
## 295  103
## 296   90
## 297   66
## 298    0
## 299    0
## 300  336
## 301  443
## 302  659
## 303  600
## 304   NA
## 305   NA
## 306   NA
## 307   NA
## 308   NA
## 309   NA
## 310   NA
## 311   NA
## 312  577
## 313  581
## 314  511
```

```
## 315  507
## 316  504
## 317  417
## 318  486
## 319  454
## 320  362
## 321  512
## 322  450
## 323  476
## 324  480
## 325  542
## 326  520
## 327  581
## 328  567
## 329  504
## 330  570
## 331  498
## 332  489
## 333  594
## 334  503
## 335  542
## 336  628
## 337  630
## 338  531
## 339  562
## 340  621
## 341  577
## 342  536
## 343  612
## 344  625
## 345  777
## 346  636
## 347  694
## 348  771
## 349  777
## 350  863
## 351  852
## 352  839
## 353  847
## 354  960
## 355  917
## 356  964
## 357 1020
## 358 1020
## 359  974
## 360  923
## 361  871
## 362  799
## 363  811
## 364  750
## 365  832
## 366  830
## 367  717
## 368  729
```

```
## 369  736
## 370  758
## 371  816
## 372  825
## 373  755
## 374  795
## 375  820
## 376  865
## 377  871
## 378 1018
## 379 1038
## 380 1020
## 381 1044
## 382 1049
## 383  928
## 384 1019
## 385 1138
## 386  944
## 387  943
## 388 1189
## 389 1044
## 390 1056
## 391  105
## 392  129
## 393    0
## 394 1170
## 395  746
## 396  546
## 397  700
## 398  697
## 399  652
## 400  646
## 401  527
## 402  440
## 403  571
## 404  668
## 405  550
## 406  453
## 407  685
## 408  606
## 409  423
## 410  457
## 411  432
## 412  516
## 413  625
## 414  833
## 415    0
## 416  140
## 417   96
## 418  110
## 419  305
## 420  424
## 421  603
## 422  419
```

```
## 423              432
## 424              505
## 425              660
## 426              678
## 427              621
## 428              670
## 429              536
## 430              676
## 431              633
## 432              548
## 433              535
## 434              366
## 435              367
## 436              389
## 437              426
## 438              374
## 439              443
## 440              620
## 441              668
## 442               NA
## 443               NA
## 444               NA
## 445               NA
## 446               NA
## 447               NA
## 448               NA
## 449              527
## 450              649
## 451              664
## 452              671
## 453              607
## 454              701
## 455              431
## 456              415
## 457              396
## 458              470
## 459              510
## 460              552
## 461              494
## 462              470
## 463              521
## 464              544
## 465              604
## 466              668
## 467              674
## 468              541
## 469              676
## 470              495
## 471              522
## 472              507
## 473              582
## 474              595
## 475              705
## 476              639
```

```
## 477              553
## 478              548
## 479              486
## 480              634
## 481              608
## 482              608
## 483              603
## 484              812
## 485              681
## 486              749
## 487              785
## 488              729
## 489              853
## 490              816
## 491             1040
## 492              897
## 493              958
## 494              944
## 495             1080
## 496             1044
## 497             1049
## 498              948
## 499              998
## 500              918
## 501              928
## 502              844
## 503              772
## 504              852
## 505              860
## 506              857
## 507              802
## 508              834
## 509              796
## 510              746
## 511              762
## 512              961
## 513              869
## 514              868
## 515              973
## 516              943
## 517              978
## 518             1071
## 519              910
## 520              921
## 521              869
## 522              890
## 523              923
## 524             1075
## 525             1072
## 526             1090
## 527             1003
## 528             1215
## 529             1170
## 530             1120
```

```
## 531              1077
## 532              1269
## 533              1158
## 534              1080
## 535               920
## 536               928
## 537               679
## 538               629
## 539               485
## 540               461
## 541               635
## 542                NA
## 543                NA
## 544                NA
## 545                NA
## 546                NA
## 547                NA
## 548               586
## 549               641
## 550               605
## 551               622
## 552               504
## 553               468
## 554               371
## 555               499
## 556               487
## 557               482
## 558               443
## 559               426
## 560               403
## 561               412
## 562               513
## 563               465
## 564               504
## 565               471
## 566               414
## 567               449
## 568               558
## 569               432
## 570               447
## 571               470
## 572               535
## 573               528
## 574               599
## 575               501
## 576               467
## 577               611
## 578               471
## 579               508
## 580               631
## 581               554
## 582               563
## 583               595
## 584               551
```

```
## 585  548
## 586  558
## 587  564
## 588  626
## 589  694
## 590  784
## 591  704
## 592  667
## 593  682
## 594  612
## 595  674
## 596  896
## 597  902
## 598  916
## 599  872
## 600  849
## 601  840
## 602  844
## 603  872
## 604  870
## 605 1042
## 606  952
## 607  869
## 608  805
## 609  744
## 610  666
## 611  629
## 612  676
## 613  678
## 614  792
## 615  866
## 616  888
## 617  883
## 618  843
## 619  940
## 620  971
## 621  914
## 622  878
## 623  903
## 624  896
## 625  784
## 626  834
## 627  814
## 628  863
## 629  927
## 630  916
## 631  815
## 632  960
## 633  998
## 634  952
## 635  916
## 636 1056
## 637  413
## 638  431
```

```
## 639              600
## 640              608
## 641              743
## 642              439
## 643              671
## 644              596
## 645              463
## 646              519
## 647              512
## 648              324
## 649              314
## 650              310
## 651              268
## 652              320
## 653              319
## 654              476
## 655               NA
## 656               NA
## 657               NA
## 658               NA
## 659               NA
## 660               NA
## 661              542
## 662              521
## 663              613
## 664              660
## 665              539
## 666              607
## 667              502
## 668              383
## 669              426
## 670              389
## 671              326
## 672              401
## 673              386
## 674              354
## 675              346
## 676              352
## 677              352
## 678              365
## 679              514
## 680              487
## 681              459
## 682              377
## 683              571
## 684              579
## 685              614
## 686              616
## 687              559
## 688              566
## 689              533
## 690              450
## 691              585
## 692              411
```

```
## 693              560
## 694              635
## 695              558
## 696              620
## 697              588
## 698              526
## 699              607
## 700              746
## 701              737
## 702              691
## 703              791
## 704              805
## 705              803
## 706              903
## 707              801
## 708              903
## 709              960
## 710              974
## 711             1003
## 712              888
## 713              969
## 714              984
## 715              907
## 716              947
## 717              940
## 718              916
## 719              902
## 720              911
## 721              905
## 722              908
## 723              852
## 724              830
## 725              817
## 726             1006
## 727              978
## 728              861
## 729              920
## 730              928
## 731              928
## 732             1028
## 733              913
## 734              888
## 735             1025
## 736             1049
## 737             1064
## 738             1134
## 739             1113
## 740             1107
## 741             1118
## 742              995
## 743             1172
## 744             1188
## 745             1326
## 746             1335
```

```
## 747             1303
## 748             1192
## 749              624
## 750              388
## 751              424
## 752              689
## 753              766
## 754               NA
## 755               NA
## 756               NA
## 757               NA
## 758               NA
## 759               NA
## 760              594
## 761              725
## 762              726
## 763              636
## 764              627
## 765              492
## 766              428
## 767              399
## 768              396
## 769              348
## 770              407
## 771              393
## 772              399
## 773              349
## 774              388
## 775              448
## 776              387
## 777              485
## 778              455
## 779              484
## 780              457
## 781              580
## 782              645
## 783              604
## 784              628
## 785              636
## 786              572
## 787              552
## 788              646
## 789              733
## 790              641
## 791              601
## 792              562
## 793              656
## 794              665
## 795              718
## 796              703
## 797              752
## 798              804
## 799              832
## 800              867
```

```
## 801              758
## 802              724
## 803              939
## 804             1102
## 805             1063
## 806              914
## 807              984
## 808              863
## 809              912
## 810              909
## 811              868
## 812              791
## 813              793
## 814              756
## 815              680
## 816              643
## 817              711
## 818              791
## 819              633
## 820              596
## 821              691
## 822              815
## 823              817
## 824              944
## 825              977
## 826              866
## 827              934
## 828              836
## 829              888
## 830              885
## 831              843
## 832              902
## 833              862
## 834              849
## 835              961
## 836             1099
## 837             1057
## 838             1076
## 839             1000
## 840             1062
## 841             1009
## 842             1093
## 843             1204
## 844              430
## 845              715
## 846              673
## 847              627
## 848              573
## 849              686
## 850              582
## 851              541
## 852              583
## 853              290
## 854              378
```

```
## 855  450
## 856  400
## 857  425
## 858  333
## 859  295
## 860    0
## 861    0
## 862  339
## 863  313
## 864  381
## 865  419
## 866  541
## 867  944
## 868 1054
## 869 1061
## 870 1108
## 871 1060
## 872  949
## 873  863
## 874  907
## 875 1027
## 876 1043
## 877 1134
## 878 1181
## 879 1103
## 880 1108
## 881  744
## 882  899
## 883  415
## 884  627
## 885  677
## 886   NA
## 887   NA
## 888   NA
## 889   NA
## 890   NA
## 891  530
## 892  569
## 893  556
## 894  504
## 895  489
## 896  494
## 897  411
## 898  398
## 899  405
## 900  421
## 901  406
## 902  461
## 903  522
## 904  534
## 905  492
## 906  561
## 907  555
## 908  489
```

```
## 909              486
## 910              748
## 911              611
## 912              623
## 913              585
## 914              614
## 915              582
## 916              526
## 917              564
## 918              648
## 919              594
## 920              530
## 921              528
## 922              505
## 923              552
## 924              636
## 925              634
## 926              613
## 927              650
## 928              676
## 929              713
## 930              766
## 931              867
## 932              900
## 933              908
## 934              912
## 935              952
## 936              987
## 937              994
## 938              964
## 939              922
## 940              825
## 941              854
## 942              824
## 943              722
## 944              784
## 945              888
## 946              735
## 947              764
## 948              695
## 949              819
## 950              844
## 951              743
## 952              807
## 953              831
## 954              941
## 955              932
## 956              885
## 957              913
## 958              841
## 959              899
## 960              952
## 961             1185
## 962             1055
```

```
## 963              1122
## 964              1264
## 965              1110
## 966              1164
## 967              1070
## 968              1056
## 969               982
## 970               972
## 971              1041
## 972              1099
## 973              1144
## 974              1038
## 975              1133
## 976               601
## 977               625
## 978               677
## 979               561
## 980               337
## 981               490
## 982               162
## 983              1054
## 984              1051
## 985              1038
## 986              1122
## 987              1074
## 988              1120
## 989              1145
## 990              1191
## 991              1130
## 992               978
## 993               968
## 994               918
## 995              1249
## 996               186
## 997               271
## 998                 0
## 999                 0
## 1000              816
## 1001              938
## 1002              872
## 1003              877
## 1004              885
## 1005              934
## 1006              988
## 1007              972
## 1008              911
## 1009              888
## 1010              960
## 1011              962
## 1012              864
## 1013              767
## 1014              719
## 1015              839
## 1016              743
```

```
## 1017              745
## 1018              750
## 1019              719
## 1020              830
## 1021              869
## 1022              837
## 1023              873
## 1024              916
## 1025              936
## 1026              840
## 1027              860
## 1028              997
## 1029             1027
## 1030             1025
## 1031              911
## 1032             1011
## 1033             1116
## 1034             1057
## 1035             1085
## 1036             1122
## 1037             1138
## 1038             1129
## 1039             1119
## 1040             1120
## 1041             1021
## 1042             1037
## 1043             1076
## 1044              532
## 1045              848
## 1046              487
## 1047              590
## 1048              540
## 1049              741
## 1050              740
## 1051              666
## 1052              533
## 1053              901
## 1054              958
## 1055              824
## 1056              748
## 1057              696
## 1058              768
## 1059              650
## 1060              687
## 1061              675
## 1062              659
## 1063              722
## 1064              840
## 1065              919
## 1066             1034
## 1067              950
## 1068              897
## 1069              969
## 1070              741
```

```
## 1071              936
## 1072              983
## 1073              955
## 1074              949
## 1075             1068
## 1076              990
## 1077              840
## 1078              898
## 1079              926
## 1080             1057
## 1081             1040
## 1082             1085
## 1083               77
## 1084              650
## 1085              469
## 1086              618
## 1087              441
## 1088              474
## 1089              453
## 1090              514
## 1091              532
## 1092              375
## 1093              364
## 1094              393
## 1095              333
## 1096              313
## 1097              351
## 1098              324
## 1099              535
## 1100              574
## 1101               NA
## 1102               NA
## 1103               NA
## 1104               NA
## 1105              738
## 1106              618
## 1107              603
## 1108              588
## 1109              529
## 1110              579
## 1111              565
## 1112              419
## 1113              469
## 1114              411
## 1115              426
## 1116              402
## 1117              376
## 1118              491
## 1119              540
## 1120              546
## 1121              604
## 1122              480
## 1123              551
## 1124              482
```

```
## 1125              617
## 1126              669
## 1127              677
## 1128              604
## 1129              563
## 1130              447
## 1131              474
## 1132              457
## 1133              597
## 1134              590
## 1135              720
## 1136              600
## 1137              665
## 1138              670
## 1139              740
## 1140              722
## 1141              657
## 1142              760
## 1143              776
## 1144              892
## 1145              894
## 1146              925
## 1147              880
## 1148              837
## 1149              870
## 1150              867
## 1151              893
## 1152              891
## 1153              830
## 1154              881
## 1155              980
## 1156              823
## 1157              846
## 1158              755
## 1159              821
## 1160              800
## 1161              820
## 1162              825
## 1163              744
## 1164              896
## 1165              818
## 1166              834
## 1167              841
## 1168              810
## 1169              804
## 1170              925
## 1171              829
## 1172              846
## 1173              923
## 1174              953
## 1175              896
## 1176              952
## 1177              957
## 1178              899
```

```
## 1179              937
## 1180              976
## 1181             1190
## 1182             1079
## 1183             1056
## 1184             1030
## 1185             1083
## 1186             1062
## 1187              940
## 1188              985
## 1189             1094
## 1190              959
## 1191              241
## 1192              378
## 1193              509
## 1194              612
## 1195              648
## 1196              665
## 1197              424
## 1198              725
## 1199              612
## 1200              565
## 1201              559
## 1202              541
## 1203              397
## 1204              454
## 1205              399
## 1206              528
## 1207              565
## 1208              460
## 1209              400
## 1210               81
## 1211                0
## 1212             1015
## 1213              985
## 1214              930
## 1215              901
## 1216              977
## 1217              922
## 1218              915
## 1219              862
## 1220              805
## 1221              750
## 1222              745
## 1223              685
## 1224              714
## 1225              665
## 1226              751
## 1227              992
## 1228             1040
## 1229              911
## 1230              791
## 1231              821
## 1232              802
```

```
## 1233              958
## 1234              900
## 1235              986
## 1236              973
## 1237             1039
## 1238             1072
## 1239             1245
## 1240             1399
## 1241             1125
## 1242             1221
## 1243             1320
## 1244             1162
## 1245             1233
## 1246              414
## 1247              353
## 1248              547
## 1249               NA
## 1250               NA
## 1251               NA
## 1252               NA
## 1253               NA
## 1254               NA
## 1255               NA
## 1256              626
## 1257              673
## 1258              573
## 1259              632
## 1260              608
## 1261              457
## 1262              591
## 1263              579
## 1264              500
## 1265              465
## 1266              474
## 1267              412
## 1268              458
## 1269              399
## 1270              378
## 1271              410
## 1272              426
## 1273              461
## 1274              483
## 1275              465
## 1276              421
## 1277              476
## 1278              430
## 1279              533
## 1280              407
## 1281              490
## 1282              530
## 1283              513
## 1284              575
## 1285              613
## 1286              502
```

```
## 1287              514
## 1288              562
## 1289              606
## 1290              521
## 1291              637
## 1292              542
## 1293              644
## 1294              756
## 1295              688
## 1296              923
## 1297              790
## 1298              927
## 1299              929
## 1300              851
## 1301              823
## 1302              918
## 1303             1019
## 1304              844
## 1305              976
## 1306              966
## 1307              906
## 1308              905
## 1309              857
## 1310              952
## 1311              954
## 1312              791
## 1313              760
## 1314              714
## 1315              759
## 1316              707
## 1317              739
## 1318              887
## 1319              802
## 1320              735
## 1321              898
## 1322              832
## 1323              743
## 1324              749
## 1325              747
## 1326              834
## 1327              850
## 1328              910
## 1329             1031
## 1330              958
## 1331             1121
## 1332              915
## 1333             1021
## 1334             1083
## 1335             1096
## 1336             1027
## 1337              982
## 1338              978
## 1339              872
## 1340              513
```

```
## 1341              578
## 1342              945
## 1343              679
## 1344              586
## 1345                0
## 1346               74
## 1347              156
## 1348               67
## 1349                0
## 1350                0
## 1351              450
## 1352             1003
## 1353             1078
## 1354             1129
## 1355              998
## 1356              981
## 1357             1203
## 1358             1089
## 1359             1062
## 1360             1028
## 1361              810
## 1362              735
## 1363              805
## 1364              797
## 1365              887
## 1366              829
## 1367              817
## 1368              840
## 1369              948
## 1370             1005
## 1371             1001
## 1372              872
## 1373              968
## 1374             1012
## 1375              853
## 1376              934
## 1377              851
## 1378              794
## 1379              956
## 1380              879
## 1381             1157
## 1382             1118
## 1383             1069
## 1384             1029
## 1385             1049
## 1386              988
## 1387             1037
## 1388             1062
## 1389             1050
## 1390             1041
## 1391             1159
## 1392             1075
## 1393              437
## 1394              477
```

```
## 1395              694
## 1396              564
## 1397              101
## 1398              504
## 1399              562
## 1400              588
## 1401               NA
## 1402               NA
## 1403               NA
## 1404               NA
## 1405               NA
## 1406               NA
## 1407               NA
## 1408              662
## 1409              748
## 1410              713
## 1411              665
## 1412              566
## 1413              487
## 1414              558
## 1415              659
## 1416              602
## 1417              560
## 1418              550
## 1419              448
## 1420              507
## 1421              610
## 1422              636
## 1423              572
## 1424              545
## 1425              554
## 1426              546
## 1427              628
## 1428              657
## 1429              583
## 1430              637
## 1431              594
## 1432              585
## 1433              591
## 1434              660
## 1435              604
## 1436              743
## 1437              611
## 1438              567
## 1439              487
## 1440              575
## 1441              686
## 1442              691
## 1443              665
## 1444              692
## 1445              794
## 1446              865
## 1447              871
## 1448              860
```

```
## 1449              785
## 1450              842
## 1451              813
## 1452              976
## 1453              951
## 1454              832
## 1455             1043
## 1456              958
## 1457              845
## 1458              808
## 1459              717
## 1460              720
## 1461              680
## 1462              690
## 1463              719
## 1464              628
## 1465              681
## 1466              691
## 1467              739
## 1468              657
## 1469              719
## 1470              686
## 1471              673
## 1472              776
## 1473              911
## 1474              949
## 1475              941
## 1476              836
## 1477              861
## 1478              910
## 1479              946
## 1480              957
## 1481              909
## 1482              954
## 1483             1025
## 1484              978
## 1485             1013
## 1486             1048
## 1487             1178
## 1488             1042
## 1489              982
## 1490              989
## 1491             1053
## 1492              410
## 1493              349
## 1494              616
## 1495              649
## 1496               NA
## 1497               NA
## 1498               NA
## 1499               NA
## 1500               NA
## 1501               NA
## 1502               NA
```

```
## 1503              551
## 1504              676
## 1505              668
## 1506              550
## 1507              614
## 1508              654
## 1509              624
## 1510              598
## 1511              622
## 1512              551
## 1513              514
## 1514              460
## 1515              468
## 1516              475
## 1517              559
## 1518              579
## 1519              663
## 1520              663
## 1521              631
## 1522              655
## 1523              625
## 1524              598
## 1525              629
## 1526              690
## 1527              619
## 1528              515
## 1529              489
## 1530              515
## 1531              500
## 1532              625
## 1533              550
## 1534              519
## 1535              519
## 1536              594
## 1537              633
## 1538              712
## 1539              763
## 1540              805
## 1541              786
## 1542              821
## 1543              777
## 1544              803
## 1545              829
## 1546             1104
## 1547             1020
## 1548              994
## 1549             1025
## 1550             1022
## 1551              977
## 1552             1024
## 1553              926
## 1554              919
## 1555              876
## 1556              846
```

```
## 1557             916
## 1558             800
## 1559             751
## 1560             824
## 1561             962
## 1562             948
## 1563             872
## 1564             871
## 1565             861
## 1566             983
## 1567             855
## 1568             992
## 1569             981
## 1570             831
## 1571            1048
## 1572             975
## 1573            1025
## 1574            1114
## 1575            1181
## 1576            1122
## 1577            1129
## 1578            1166
## 1579            1021
## 1580             898
## 1581            1061
## 1582             819
## 1583             976
## 1584              72
## 1585             475
## 1586             621
## 1587             594
## 1588             354
## 1589             443
## 1590             643
## 1591             593
## 1592             836
## 1593             576
## 1594             663
## 1595             741
## 1596             733
## 1597             456
## 1598             604
## 1599             450
## 1600             498
## 1601             487
## 1602             545
## 1603             421
## 1604             310
## 1605             370
## 1606             367
## 1607             415
## 1608             363
## 1609             439
## 1610             635
```

```
## 1611              569
## 1612               NA
## 1613               NA
## 1614               NA
## 1615               NA
## 1616               NA
## 1617              592
## 1618              627
## 1619              655
## 1620              620
## 1621              600
## 1622              639
## 1623              605
## 1624              568
## 1625              527
## 1626              555
## 1627              647
## 1628              647
## 1629              585
## 1630              485
## 1631              574
## 1632              514
## 1633              507
## 1634              544
## 1635              498
## 1636              483
## 1637              518
## 1638              575
## 1639              700
## 1640              616
## 1641              678
## 1642              548
## 1643              521
## 1644              558
## 1645              627
## 1646              524
## 1647              585
## 1648              529
## 1649              621
## 1650              625
## 1651              705
## 1652              599
## 1653              552
## 1654              562
## 1655              628
## 1656              652
## 1657              797
## 1658              903
## 1659             1109
## 1660              976
## 1661              929
## 1662              955
## 1663              924
## 1664             1033
```

```
## 1665            1003
## 1666            1130
## 1667            1031
## 1668             966
## 1669             979
## 1670             822
## 1671             960
## 1672             793
## 1673             806
## 1674             866
## 1675             764
## 1676             708
## 1677            1084
## 1678            1095
## 1679            1161
## 1680            1109
## 1681             987
## 1682             926
## 1683             915
## 1684            1026
## 1685            1059
## 1686            1049
## 1687            1002
## 1688            1092
## 1689            1080
## 1690            1081
## 1691            1117
## 1692            1125
## 1693            1102
## 1694            1155
## 1695            1133
## 1696            1083
## 1697            1203
## 1698             979
## 1699             397
## 1700             639
## 1701             380
## 1702             570
## 1703             617
## 1704             784
## 1705             843
## 1706             498
## 1707             705
## 1708             546
## 1709             480
## 1710             343
## 1711             367
## 1712             359
## 1713             413
## 1714             375
## 1715             374
## 1716             575
## 1717             520
## 1718              NA
```

```
## 1719              NA
## 1720              NA
## 1721              NA
## 1722              NA
## 1723              NA
## 1724              NA
## 1725             555
## 1726             551
## 1727             593
## 1728             640
## 1729             690
## 1730             650
## 1731             610
## 1732             369
## 1733             444
## 1734             426
## 1735             393
## 1736             343
## 1737             381
## 1738             419
## 1739             384
## 1740             371
## 1741             373
## 1742             375
## 1743             355
## 1744             472
## 1745             478
## 1746             405
## 1747             351
## 1748             430
## 1749             463
## 1750             528
## 1751             505
## 1752             442
## 1753             445
## 1754             520
## 1755             543
## 1756             591
## 1757             595
## 1758             652
## 1759             505
## 1760             584
## 1761             723
## 1762             608
## 1763             583
## 1764             734
## 1765             647
## 1766             752
## 1767             775
## 1768             791
## 1769             771
## 1770             792
## 1771             752
## 1772             758
```

```
## 1773              841
## 1774              940
## 1775              970
## 1776             1008
## 1777             1011
## 1778              914
## 1779              953
## 1780              944
## 1781              919
## 1782              910
## 1783              842
## 1784              828
## 1785              837
## 1786              807
## 1787              879
## 1788              855
## 1789              760
## 1790              785
## 1791              862
## 1792              873
## 1793              847
## 1794              929
## 1795              914
## 1796              914
## 1797              914
## 1798              901
## 1799              872
## 1800              972
## 1801             1030
## 1802              989
## 1803             1161
## 1804             1204
## 1805             1106
## 1806             1116
## 1807             1049
## 1808             1092
## 1809             1200
## 1810              133
## 1811              173
## 1812                0
## 1813                0
## 1814              589
## 1815              332
## 1816              359
## 1817              413
## 1818              492
## 1819              511
## 1820              678
## 1821              633
## 1822              119
## 1823                0
## 1824                0
## 1825              155
## 1826             1095
```

```
## 1827              692
## 1828              194
## 1829              159
## 1830              397
## 1831             1143
## 1832             1164
## 1833              972
## 1834             1033
## 1835              966
## 1836              900
## 1837              754
## 1838              716
## 1839             1057
## 1840              848
## 1841              791
## 1842              773
## 1843              877
## 1844              822
## 1845              810
## 1846              809
## 1847              992
## 1848              898
## 1849             1013
## 1850              902
## 1851             1069
## 1852              864
## 1853             1046
## 1854             1055
## 1855              981
## 1856             1014
## 1857             1129
## 1858             1072
## 1859             1169
## 1860             1273
## 1861             1062
## 1862             1073
## 1863              910
## 1864             1104
## 1865              769
## 1866              711
## 1867              725
## 1868              727
## 1869              822
## 1870              806
## 1871              840
## 1872              871
## 1873              942
## 1874             1148
## 1875              863
## 1876              792
## 1877              838
## 1878              749
## 1879              811
## 1880              841
```

```
## 1881  901
## 1882  943
## 1883  973
## 1884 1059
## 1885 1110
## 1886 1088
## 1887 1095
## 1888 1073
## 1889  989
## 1890 1003
## 1891  989
## 1892 1058
## 1893  986
## 1894  974
## 1895  501
## 1896  451
## 1897  558
## 1898  429
## 1899  564
## 1900  496
## 1901  592
## 1902  484
## 1903  503
## 1904  342
## 1905  266
## 1906  361
## 1907  335
## 1908  404
## 1909  402
## 1910  389
## 1911  403
## 1912  680
## 1913  631
## 1914   NA
## 1915   NA
## 1916   NA
## 1917   NA
## 1918   NA
## 1919   NA
## 1920  514
## 1921  536
## 1922  533
## 1923  534
## 1924  504
## 1925  583
## 1926  595
## 1927  561
## 1928  477
## 1929  471
## 1930  573
## 1931  443
## 1932  430
## 1933  507
## 1934  527
```

```
## 1935              451
## 1936              486
## 1937              396
## 1938              435
## 1939              402
## 1940              421
## 1941              411
## 1942              508
## 1943              557
## 1944              507
## 1945              475
## 1946              524
## 1947              570
## 1948              535
## 1949              509
## 1950              548
## 1951              545
## 1952              498
## 1953              505
## 1954              487
## 1955              574
## 1956              598
## 1957              682
## 1958              550
## 1959              644
## 1960              707
## 1961              640
## 1962              590
## 1963              611
## 1964              693
## 1965              704
## 1966              859
## 1967              920
## 1968              890
## 1969              804
## 1970              889
## 1971              900
## 1972              844
## 1973              865
## 1974              978
## 1975              904
## 1976             1054
## 1977             1005
## 1978             1042
## 1979             1008
## 1980              913
## 1981              869
## 1982              780
## 1983              778
## 1984              814
## 1985              925
## 1986              792
## 1987              915
## 1988              990
```

```
## 1989  962
## 1990 1087
## 1991 1094
## 1992 1023
## 1993 1071
## 1994  973
## 1995 1067
## 1996  930
## 1997 1013
## 1998 1193
## 1999 1189
## 2000 1120
## 2001 1034
## 2002 1028
## 2003 1032
## 2004 1090
## 2005  967
## 2006  986
## 2007  874
## 2008  901
## 2009  897
## 2010  710
## 2011  528
## 2012  777
## 2013  618
## 2014  871
## 2015    0
## 2016    0
## 2017  458
## 2018  397
## 2019  513
## 2020  408
## 2021  495
## 2022  408
## 2023  625
## 2024  572
## 2025  584
## 2026  524
## 2027  530
## 2028  308
## 2029  345
## 2030  374
## 2031  388
## 2032  434
## 2033  281
## 2034  368
## 2035  625
## 2036  530
## 2037   NA
## 2038   NA
## 2039   NA
## 2040   NA
## 2041   NA
## 2042   NA
```

```
## 2043              NA
## 2044             615
## 2045             707
## 2046             656
## 2047             619
## 2048             654
## 2049             697
## 2050             689
## 2051             695
## 2052             494
## 2053             509
## 2054             479
## 2055             447
## 2056             472
## 2057             440
## 2058             438
## 2059             541
## 2060             461
## 2061             485
## 2062             522
## 2063             500
## 2064             541
## 2065             559
## 2066             566
## 2067             548
## 2068             607
## 2069             599
## 2070             528
## 2071             599
## 2072             646
## 2073             575
## 2074             533
## 2075             461
## 2076             558
## 2077             538
## 2078             548
## 2079             507
## 2080             640
## 2081             518
## 2082             504
## 2083             649
## 2084             616
## 2085             628
## 2086             654
## 2087             707
## 2088             670
## 2089             786
## 2090             833
## 2091             784
## 2092             846
## 2093             915
## 2094             925
## 2095             887
## 2096             977
```

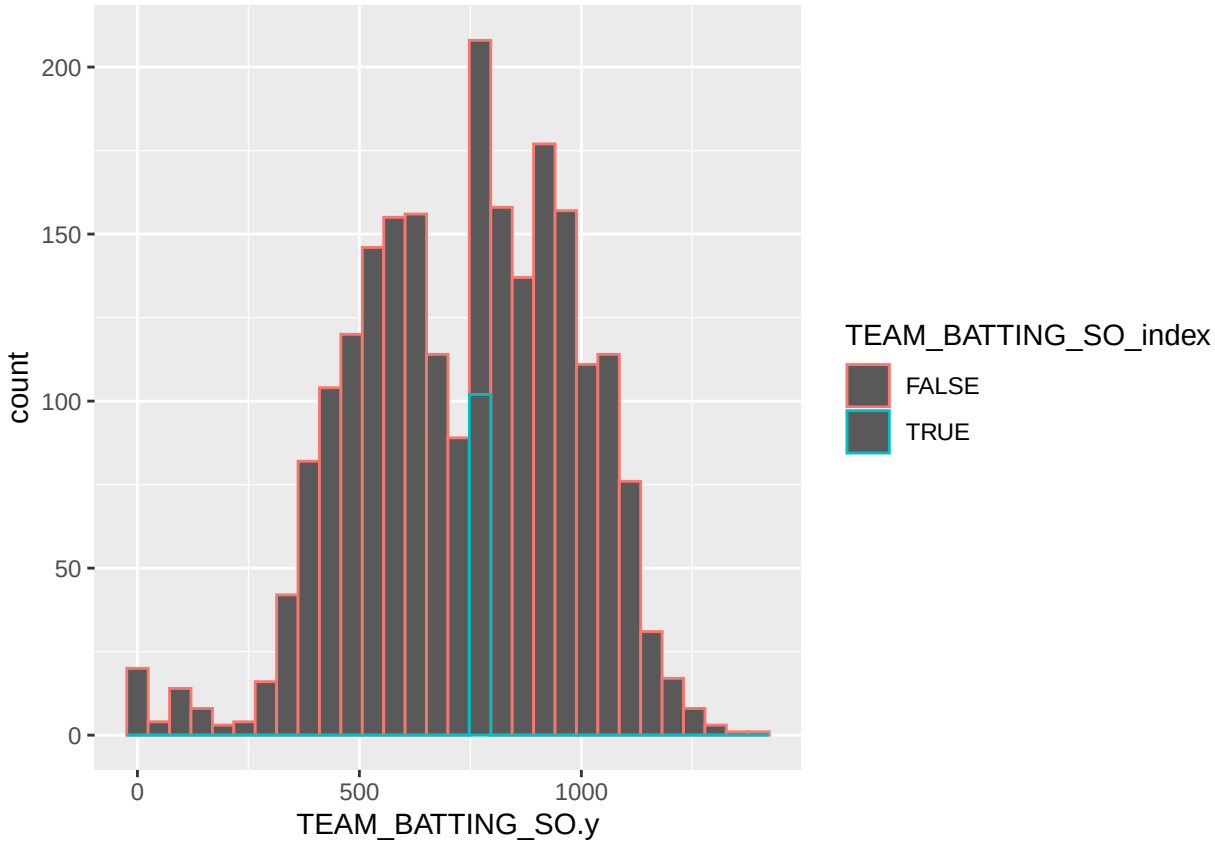
```
## 2097  925
## 2098  897
## 2099  876
## 2100  961
## 2101  757
## 2102  824
## 2103  796
## 2104  757
## 2105  649
## 2106  860
## 2107  823
## 2108  713
## 2109  838
## 2110  786
## 2111  805
## 2112  879
## 2113  924
## 2114  853
## 2115  911
## 2116  933
## 2117  827
## 2118  848
## 2119  844
## 2120  857
## 2121  996
## 2122  882
## 2123  975
## 2124 1042
## 2125 1089
## 2126 1191
## 2127 1179
## 2128 1209
## 2129 1253
## 2130 1089
## 2131  927
## 2132  952
## 2133 1085
## 2134  947
## 2135  928
## 2136  952
## 2137  551
## 2138  615
## 2139 1107
## 2140 1042
## 2141 1028
## 2142 1116
## 2143 1122
## 2144 1030
## 2145  950
## 2146  990
## 2147 1106
## 2148  923
## 2149  794
## 2150  963
```

```
## 2151            1124
## 2152            1125
## 2153            1089
## 2154            1043
## 2155             966
## 2156             900
## 2157             989
## 2158             974
## 2159             791
## 2160             719
## 2161             863
## 2162             809
## 2163             904
## 2164             779
## 2165             607
## 2166             611
## 2167             750
## 2168             767
## 2169             824
## 2170            1088
## 2171            1081
## 2172            1028
## 2173             989
## 2174            1054
## 2175            1039
## 2176            1036
## 2177            1037
## 2178             987
## 2179            1041
## 2180            1116
## 2181            1045
## 2182             937
## 2183             922
## 2184            1093
## 2185            1055
## 2186            1052
## 2187            1099
## 2188            1112
## 2189            1061
## 2190             685
## 2191             843
## 2192             824
## 2193             663
## 2194             813
## 2195             850
## 2196             749
## 2197             810
## 2198             816
## 2199             812
## 2200             848
## 2201             970
## 2202             935
## 2203             923
## 2204             970
```

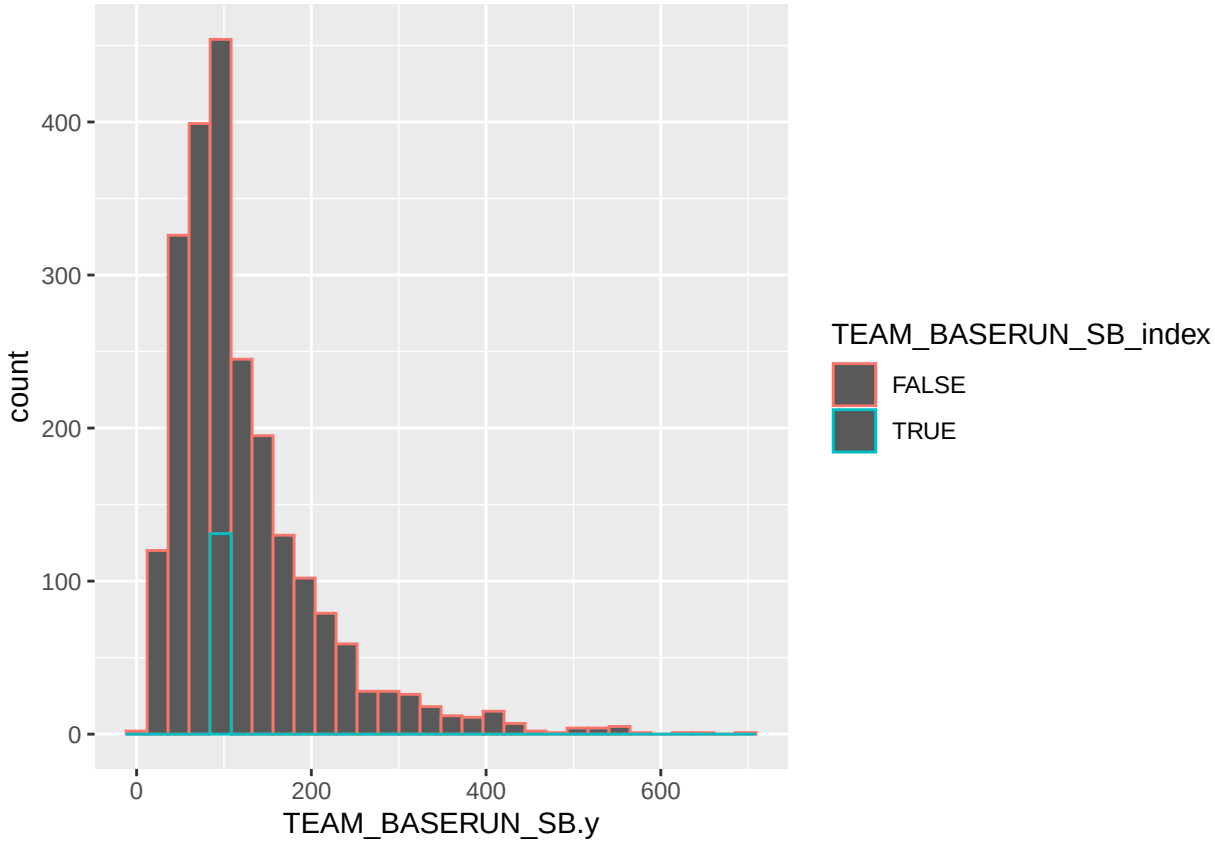
```
## 2205 1043
## 2206  933
## 2207  861
## 2208  973
## 2209 1019
## 2210 1105
## 2211 1138
## 2212 1132
## 2213 1077
## 2214 1026
## 2215 1081
## 2216 1090
## 2217  955
## 2218  906
## 2219  110
## 2220   91
## 2221  507
## 2222  463
## 2223  582
## 2224  582
## 2225  593
## 2226  298
## 2227  460
## 2228  501
## 2229  427
## 2230  411
## 2231  363
## 2232  137
## 2233    0
## 2234  807
## 2235  786
## 2236  450
## 2237  603
## 2238  596
## 2239    0
## 2240  543
## 2241  334
## 2242  584
## 2243  962
## 2244  972
## 2245  805
## 2246  817
## 2247  954
## 2248  841
## 2249  877
## 2250  881
## 2251  901
## 2252  865
## 2253  747
## 2254  816
## 2255  733
## 2256  787
## 2257  885
## 2258 1022
```

```
## 2259                 918
## 2260                1053
## 2261                 958
## 2262                1024
## 2263                1063
## 2264                 951
## 2265                1014
## 2266                1077
## 2267                1058
## 2268                 939
## 2269                1048
## 2270                1071
## 2271                1104
## 2272                 990
## 2273                 925
## 2274                1090
## 2275                1156
## 2276                 969
```

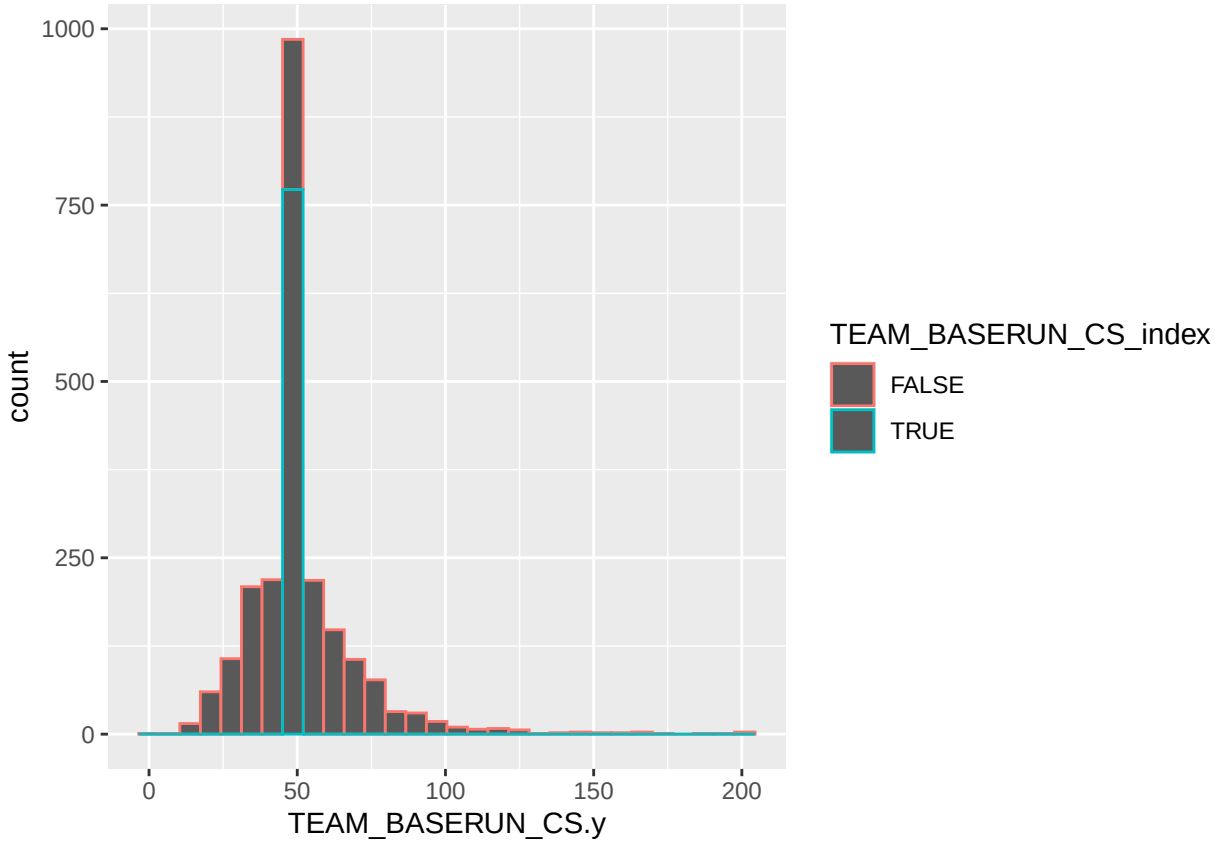
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



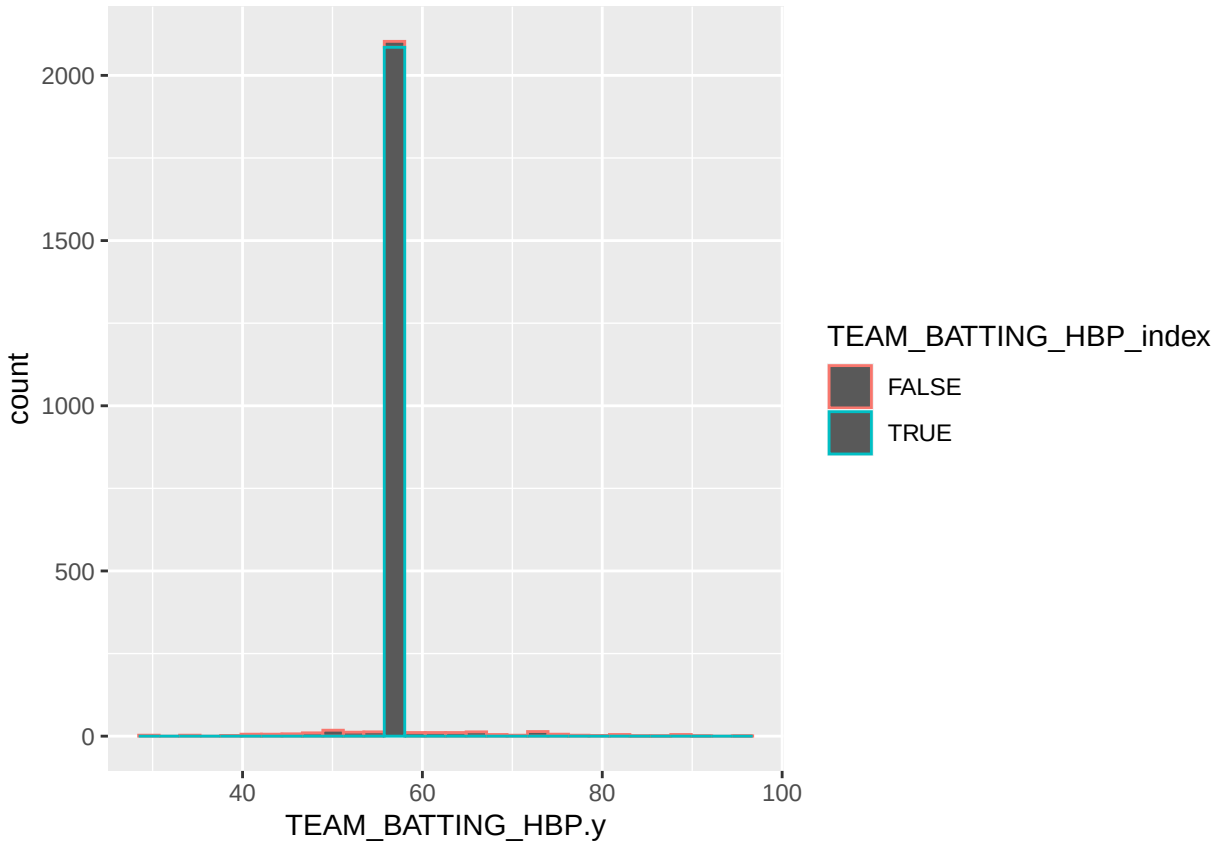

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



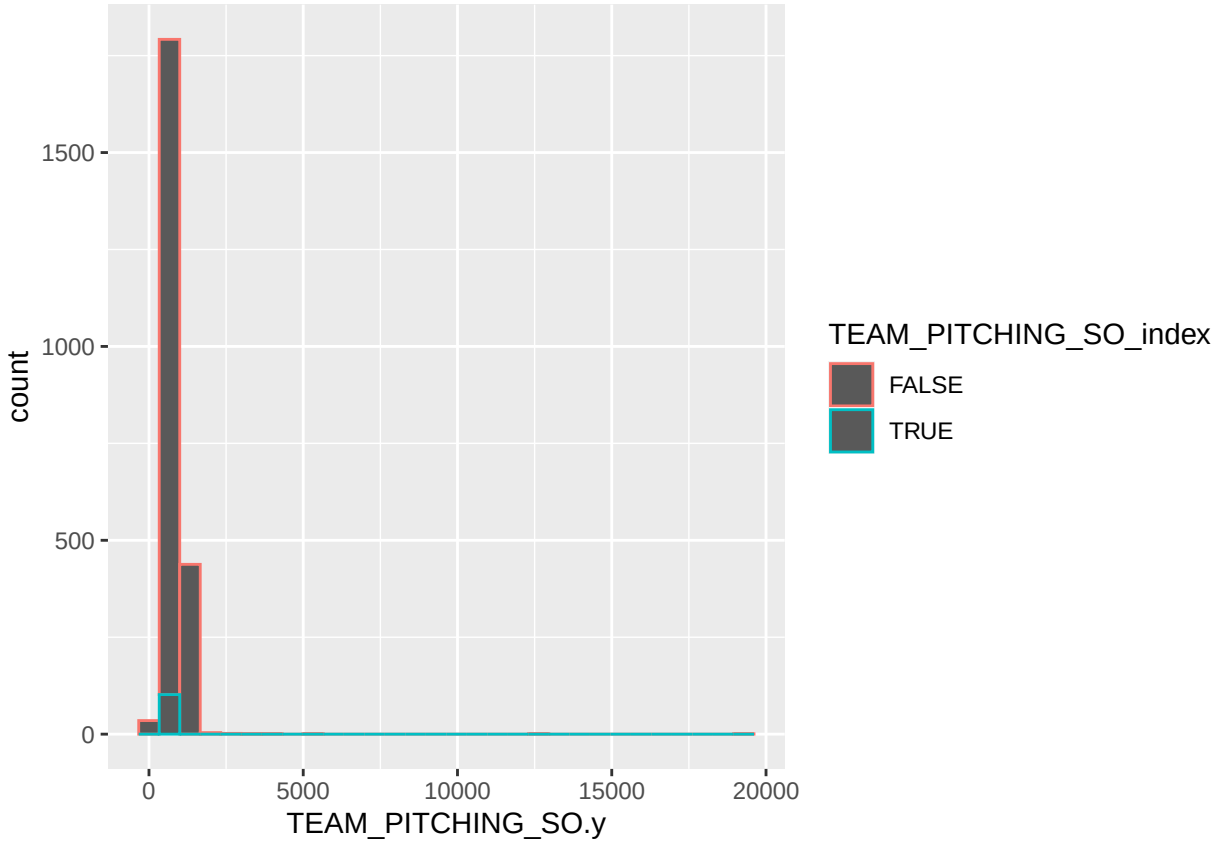

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



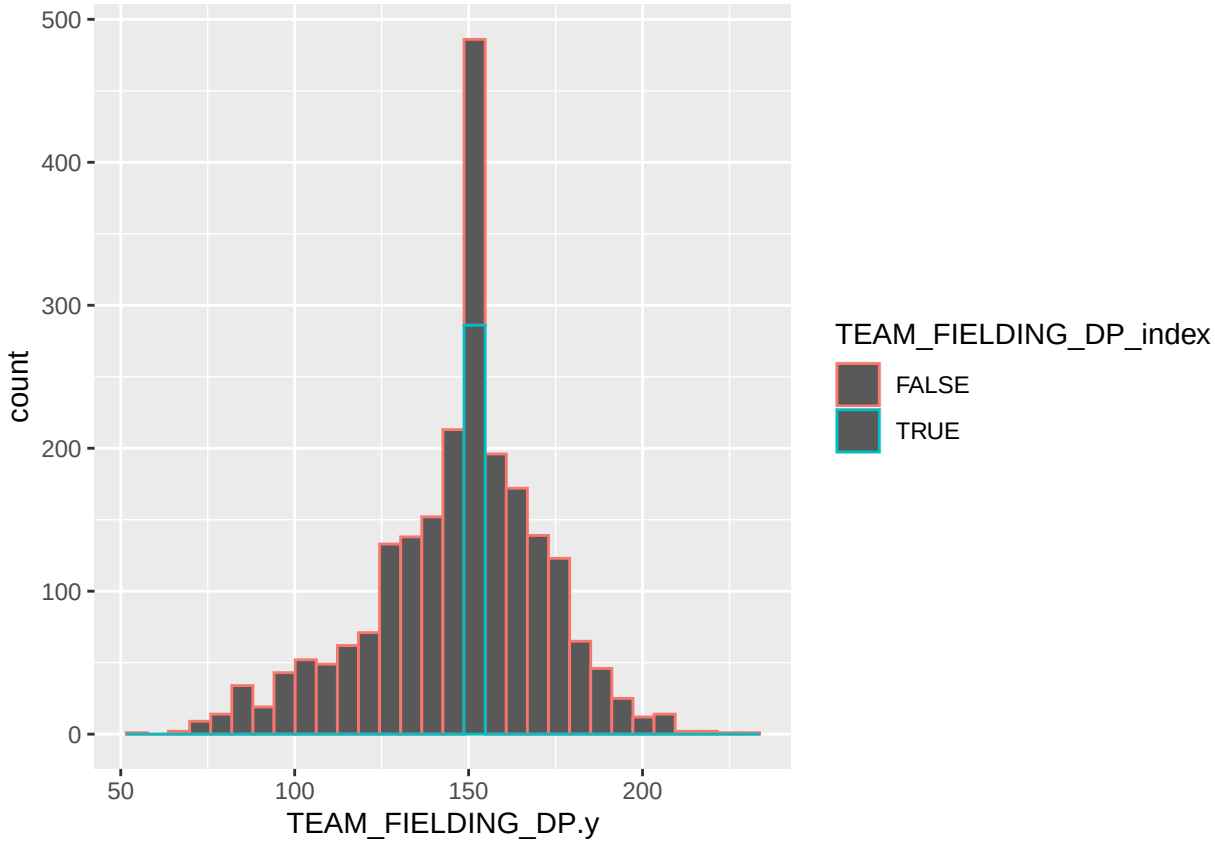

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

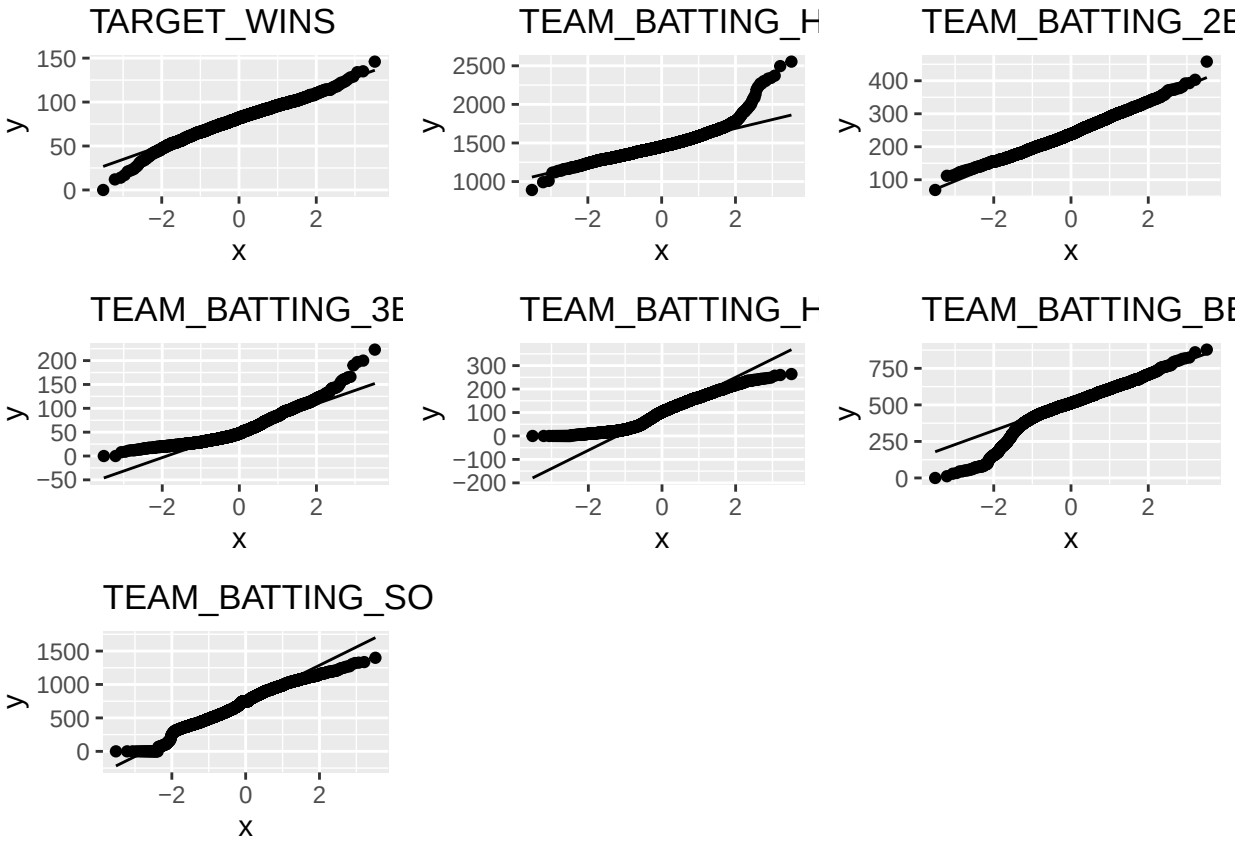


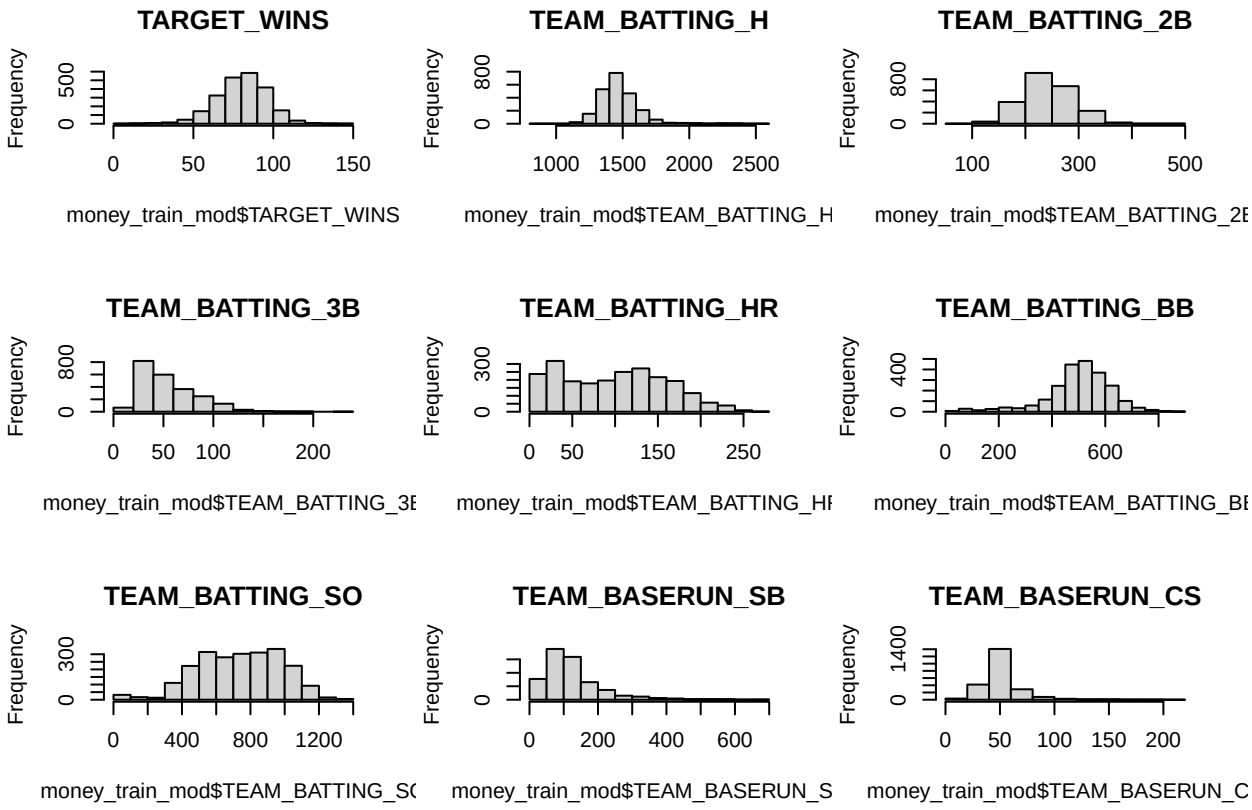

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

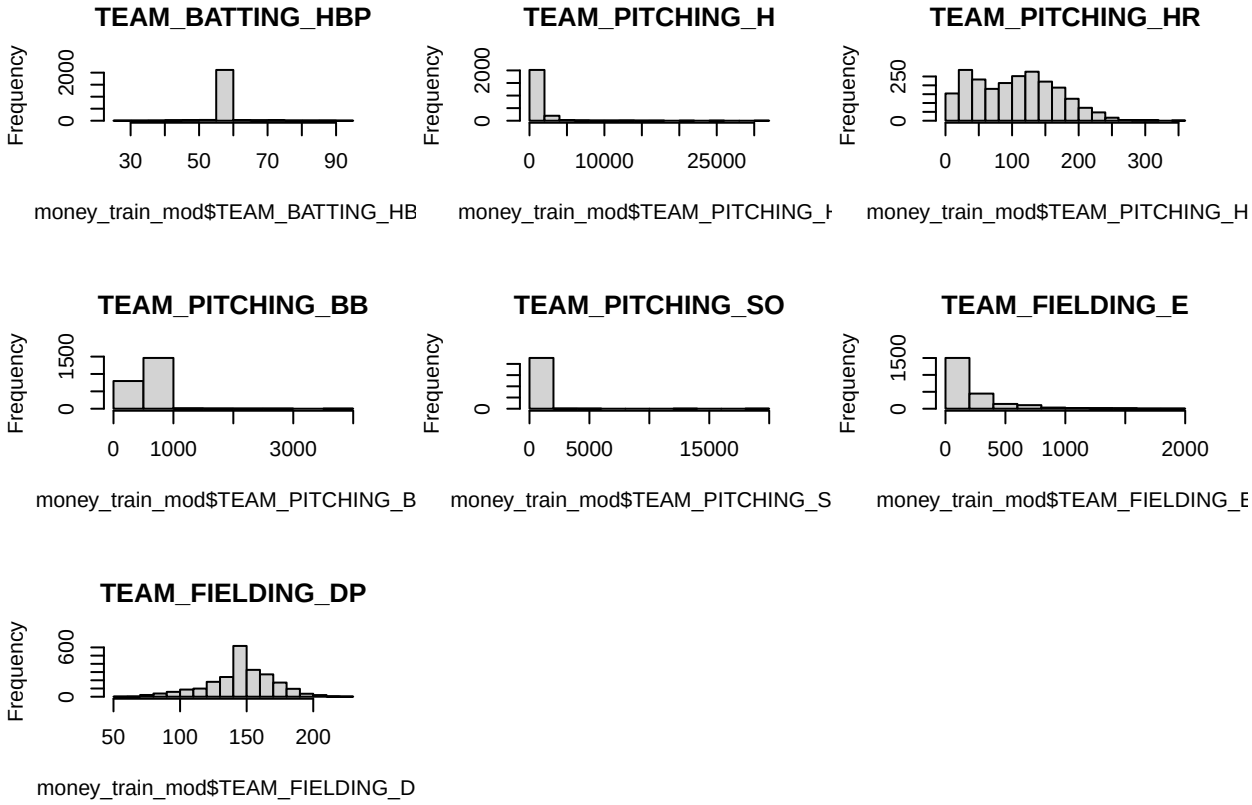


The imputation was a success for the most part, as most variables were not heavily affected in terms of their distributions. However, the TEAM_BATTING_HBP variable was heavily modified following the imputation, as ~88% of its observations were missing. As a result, we will not use this variable in our upcoming predictive tests

Investigate whether the assumptions for linear regression have been met for each variable




TARGET_WINS
Frequency
0
500
0
50
100
150
money_train_mod$TARGET_WINS
TEAM_BATTING_H
Frequency
0
800
1000
1500
2000
2500
money_train_mod$TEAM_BATTING_H
TEAM_BATTING_2B
Frequency
0
800
100
300
500
money_train_mod$TEAM_BATTING_2B
TEAM_BATTING_3B
Frequency
0
800
0
50
100
200
money_train_mod$TEAM_BATTING_3B
TEAM_BATTING_HR
Frequency
0
300
0
50
150
250
money_train_mod$TEAM_BATTING_HR
TEAM_BATTING_BB
Frequency
0
400
0
200
600
money_train_mod$TEAM_BATTING_BB
TEAM_BATTING_SO
Frequency
0
300
0
400
800
1200
money_train_mod$TEAM_BATTING_SO
TEAM_BASERUN_SB
Frequency
0
0
200
400
600
money_train_mod$TEAM_BASERUN_SB
TEAM_BASERUN_CS
Frequency
0
1400
0
50
100
150
200
money_train_mod$TEAM_BASERUN_CS



Following our imputation, we can see that some variables are either lightly or heavily skewed. In order to correct for this, we will perform log transformations in an attempt to normalize the variables.

Correlation plots following imputation to test for collinearity among variables

```
##                  TARGET_WINS TEAM_BATTING_H TEAM_BATTING_2B TEAM_BATTING_3B
## TARGET_WINS                1             NA              NA             NA
## TEAM_BATTING_H            NA      1.0000000       0.5628497             NA
## TEAM_BATTING_2B           NA      0.5628497       1.0000000             NA
## TEAM_BATTING_3B           NA             NA              NA      1.0000000
## TEAM_BATTING_HR           NA             NA              NA     -0.6355669
## TEAM_BATTING_BB           NA             NA              NA             NA
## TEAM_BATTING_SO           NA             NA              NA     -0.6557096
## TEAM_BASERUN_SB           NA             NA              NA             NA
## TEAM_BASERUN_CS           NA             NA              NA             NA
## TEAM_BATTING_HBP          NA             NA              NA             NA
## TEAM_PITCHING_H           NA             NA              NA             NA
## TEAM_PITCHING_HR          NA             NA              NA     -0.5678367
## TEAM_PITCHING_BB          NA             NA              NA             NA
## TEAM_PITCHING_SO          NA             NA              NA             NA
## TEAM_FIELDING_E           NA             NA              NA      0.5097784
## TEAM_FIELDING_DP          NA             NA              NA             NA
## INDEX                     NA             NA              NA             NA
##                  TEAM_BATTING_HR TEAM_BATTING_BB TEAM_BATTING_SO
## TARGET_WINS                   NA              NA              NA
## TEAM_BATTING_H                NA              NA              NA
```

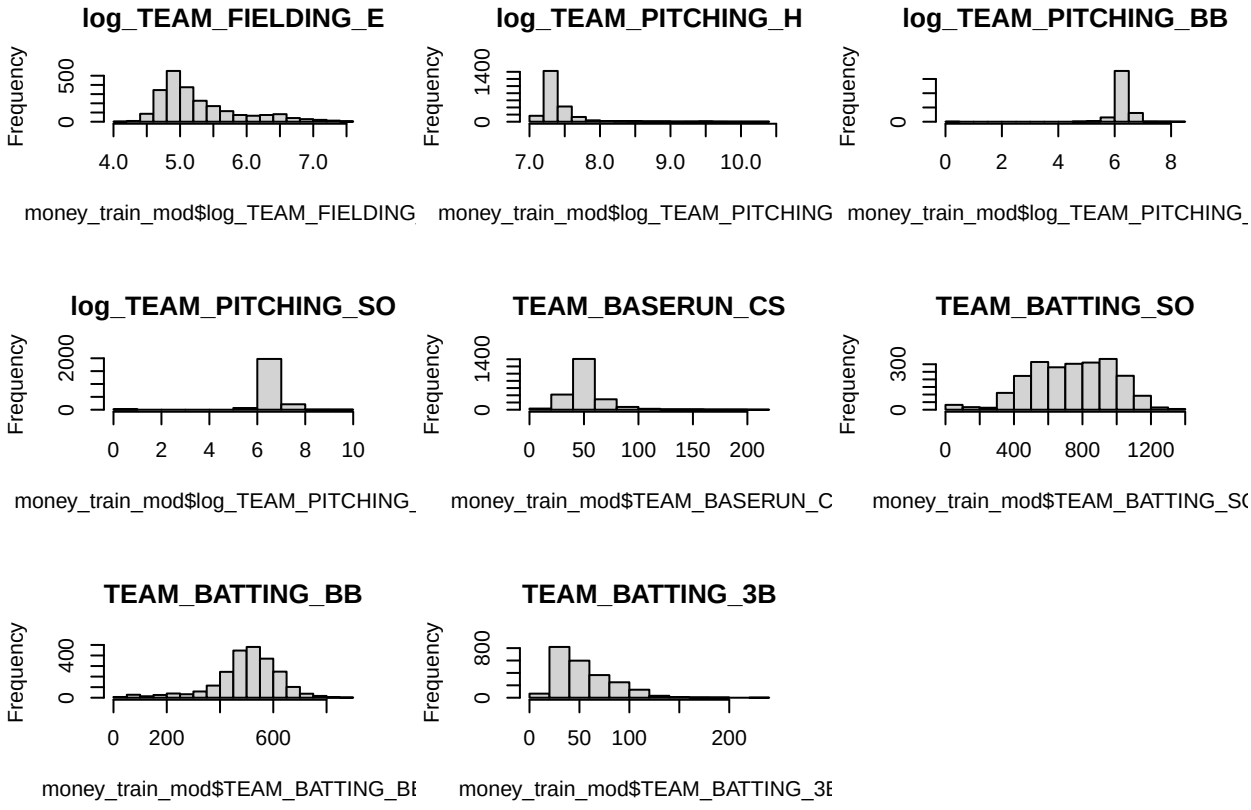
```
## TEAM_BATTING_2B              NA              NA              NA
## TEAM_BATTING_3B      -0.6355669              NA      -0.6557096
## TEAM_BATTING_HR       1.0000000       0.5137348       0.6930076
## TEAM_BATTING_BB       0.5137348       1.0000000              NA
## TEAM_BATTING_SO       0.6930076              NA       1.0000000
## TEAM_BASERUN_SB              NA              NA              NA
## TEAM_BASERUN_CS              NA              NA              NA
## TEAM_BATTING_HBP             NA              NA              NA
## TEAM_PITCHING_H              NA              NA              NA
## TEAM_PITCHING_HR      0.9693714              NA       0.6328603
## TEAM_PITCHING_BB             NA              NA              NA
## TEAM_PITCHING_SO             NA              NA              NA
## TEAM_FIELDING_E      -0.5873391      -0.6559708      -0.5825930
## TEAM_FIELDING_DP             NA              NA              NA
## INDEX                        NA              NA              NA
##                  TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_BATTING_HBP
## TARGET_WINS                   NA              NA               NA
## 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
## INDEX                         NA              NA               NA
##                  TEAM_PITCHING_H TEAM_PITCHING_HR TEAM_PITCHING_BB
## TARGET_WINS                   NA               NA               NA
## TEAM_BATTING_H                NA               NA               NA
## TEAM_BATTING_2B               NA               NA               NA
## TEAM_BATTING_3B               NA       -0.5678367               NA
## TEAM_BATTING_HR               NA        0.9693714               NA
## TEAM_BATTING_BB               NA               NA               NA
## TEAM_BATTING_SO               NA        0.6328603               NA
## TEAM_BASERUN_SB               NA               NA               NA
## TEAM_BASERUN_CS               NA               NA               NA
## TEAM_BATTING_HBP              NA               NA               NA
## TEAM_PITCHING_H         1.000000               NA               NA
## TEAM_PITCHING_HR              NA        1.0000000               NA
## TEAM_PITCHING_BB              NA               NA                1
## TEAM_PITCHING_SO              NA               NA               NA
## TEAM_FIELDING_E         0.667759               NA               NA
## TEAM_FIELDING_DP              NA               NA               NA
## INDEX                         NA               NA               NA
##                  TEAM_PITCHING_SO TEAM_FIELDING_E TEAM_FIELDING_DP INDEX
## TARGET_WINS                    NA              NA               NA    NA
## TEAM_BATTING_H                 NA              NA               NA    NA
```

```
## TEAM_BATTING_2B                NA             NA              NA    NA
## TEAM_BATTING_3B                NA      0.5097784              NA    NA
## TEAM_BATTING_HR                NA     -0.5873391              NA    NA
## TEAM_BATTING_BB                NA     -0.6559708              NA    NA
## TEAM_BATTING_SO                NA     -0.5825930              NA    NA
## TEAM_BASERUN_SB                NA             NA              NA    NA
## TEAM_BASERUN_CS                NA             NA              NA    NA
## TEAM_BATTING_HBP               NA             NA              NA    NA
## TEAM_PITCHING_H                NA      0.6677590              NA    NA
## TEAM_PITCHING_HR               NA             NA              NA    NA
## TEAM_PITCHING_BB               NA             NA              NA    NA
## TEAM_PITCHING_SO                1             NA              NA    NA
## TEAM_FIELDING_E                NA      1.0000000              NA    NA
## TEAM_FIELDING_DP               NA             NA               1    NA
## INDEX                          NA             NA              NA     1
```

Most of the variables are not highly correlated, but there are some potential predictors that are ~ 50-60% correlated. Most concerningly would be the ~90% correlation between TEAM_PITCHING_HR and TEAM_BATTING_HR. Because of the very high correlation between these variables, it may be best to avoid using them in the same model. Otherwise, it may be necessary to transform the variables via mean centering or PCA.

Log Transformations

Because some variables are heavily affected by negative skew, we can systematically remove outliers that are present



Part 3 – Building the Models

Model #1 –> Basic Multiple Linear Regression

For our first model, I decided to go with a basic multiple linear regression. I will choose variables that show the least violations of linear regression assumptions and make sure to keep minimal multicollinearity.

The following is why I chose each variable:

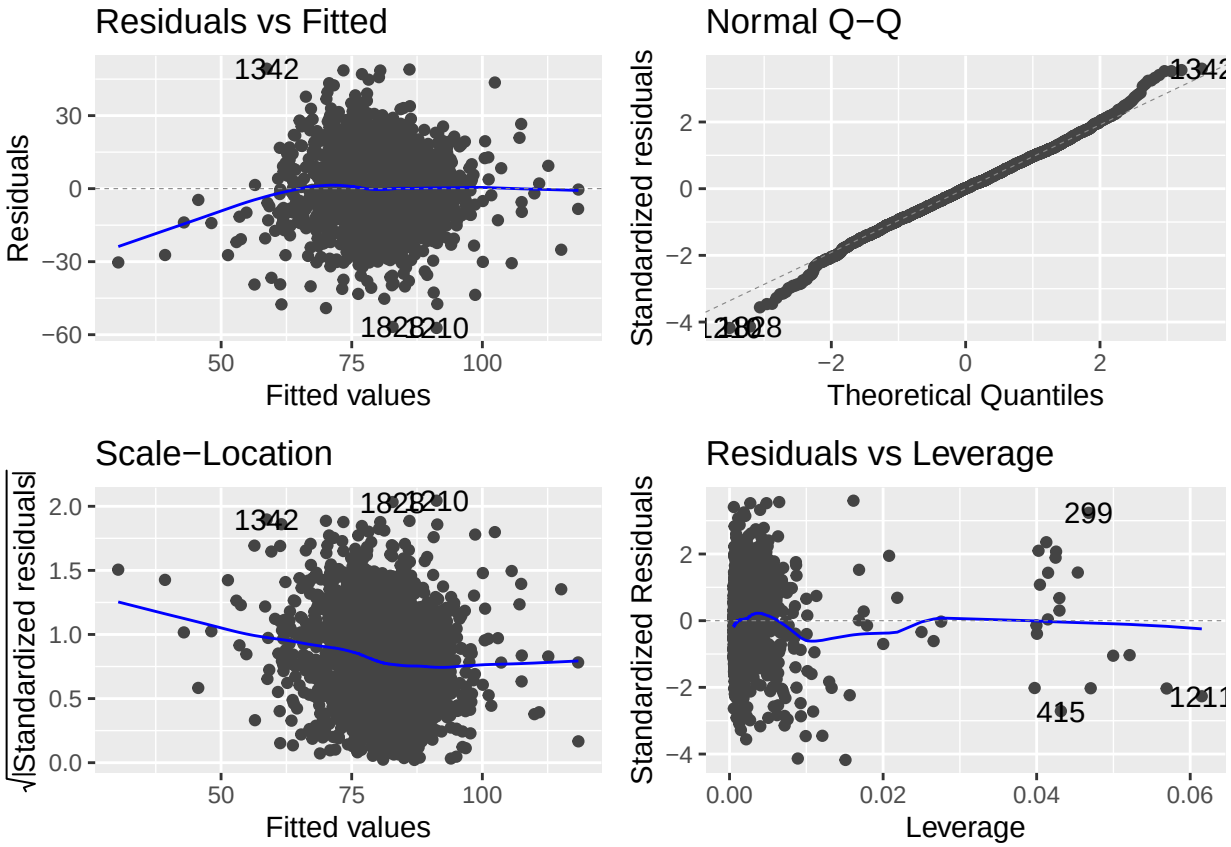
TEAM_BATTING_H –> Number of hits a team gets increases the chances of them winning... Mostly normally distributed

TEAM_BATTING_BB –> Greater walks for batting team increases the chances of winning... That means the pitching team does not pitch properly (keep throwing ball instead of strike)

log_TEAM_PITCHING_SO –> A team that pitches strikeouts might influence winning on the pitching team. Since the log was more normally distributed, I decided to include this one.

log_TEAM_FIELDING_E –> A pitching team that commits errors more likely to lose,using log because it is more normally distributed.

```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_BB +
##     log_TEAM_PITCHING_SO + log_TEAM_FIELDING_E, data = money_train_mod)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.223  -8.879   0.134   8.931  49.277
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          15.690061   6.342337   2.474   0.0134 *
## TEAM_BATTING_H        0.048173   0.002128  22.638  < 2e-16 ***
## TEAM_BATTING_BB       0.021721   0.003015   7.205 7.87e-13 ***
## log_TEAM_PITCHING_SO  0.486568   0.460910   1.056   0.2912
## log_TEAM_FIELDING_E  -3.750732   0.628940  -5.964 2.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.81 on 2271 degrees of freedom
## Multiple R-squared:  0.2333, Adjusted R-squared:  0.2319
## F-statistic: 172.7 on 4 and 2271 DF,  p-value: < 2.2e-16
```



Interpreting Coefficients TEAM_BATTING_H and TEAM_BATTING_BB both had positive coefficients (and statistically significant p-values), which makes sense. As TEAM_BATTING_H and TEAM_BATTING_BB increases, the chances of the batting team winning increases.

This makes sense in baseball world. As a batting team gets more hits, they have more chances of moving around the bases and scoring. Even if the batter hits a foul, as long as they do not get strikes out they can still get a hit and score (you can have 2 strikes and keep hitting fouls and still have a chance of getting a hit.)

The walking makes sense as well. If the pitching team keeps throwing bad pitches (that keep being balls instead of strikes), and the batter has enough skill to not blindly swing at these bad pitches, the batter will walk and go to first base, automatically increasing the chances of them scoring.

log_TEAM_PITCHING_SO has a positive coefficient but has a high p-value. This means that as the log(# of strike outs the pitching team got) increases, so does the chances of the pitching team winning. This makes sense in terms of baseball. If a team has a good pitcher that throws good strikes, the chances of the other team winning decreases. Sometimes, there are even no-hitter games (where the batting team didn't even get a hit/ all strike outs). However, since the p-value was not significant, this relationship is possibly not as strong in influencing the win of a game.

log_TEAM_FIELDING_E had a negative coefficient and a statistically significant p-value. As log(# of fielding error) increases, the chances of that team winning decreases. This makes sense in the baseball context. If players on the fields are missing and not catching the ball properly, they give the batting team an advantage and they can score more (AKA pitching team more likely to lose).

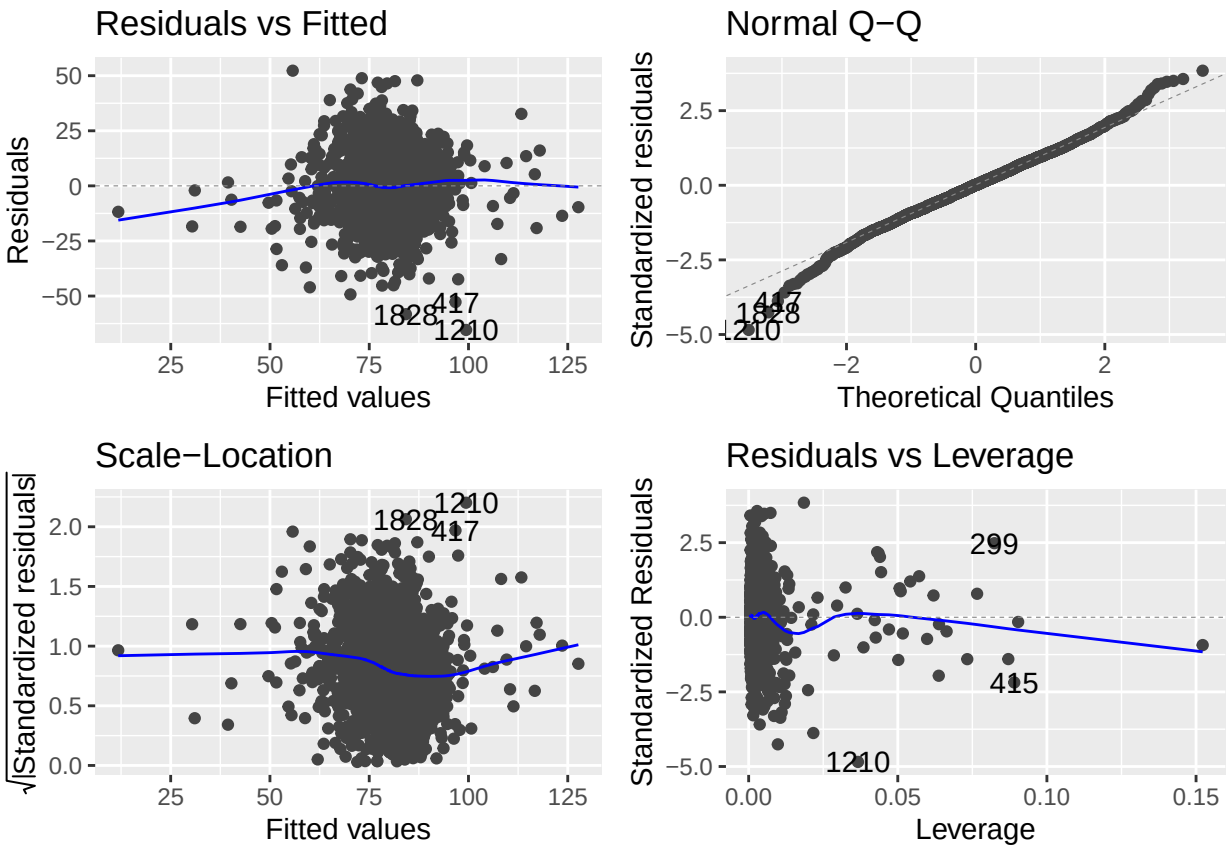
Model #2 –> Multiple Linear Regression with Interaction Terms

To win in baseball, you need to have BOTH good defense and good offense.

A team can have the best batters/ designated hitters... but what good is that if you have bad pitches/ players that can't throw and catch. You need a good mix of both to win.

In this model, I perform a multiple regression using interaction terms. I am using the same predictors that I used above. Interaction terms in regression models allow us to explore how the effect of one variable on the outcome changes depending on the level of another variable. In the context of baseball, adding interaction terms between offensive and defensive statistics can help quantify the idea that the impact of having strong hitters (offense) on winning games might vary depending on the strength of the team's pitching (defense), and vice versa.

```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H * TEAM_BATTING_BB +
##     log_TEAM_PITCHING_SO * log_TEAM_FIELDING_E, data = money_train_mod)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.387  -8.731   0.171   9.096  52.275
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           4.889e+01  3.001e+01   1.629  0.10346
## TEAM_BATTING_H                        6.635e-02  4.766e-03  13.921  < 2e-16
## TEAM_BATTING_BB                       9.502e-02  1.697e-02   5.600 2.41e-08
## log_TEAM_PITCHING_SO                 -8.602e+00  4.185e+00  -2.056  0.03992
## log_TEAM_FIELDING_E                  -1.275e+01  4.238e+00  -3.010  0.00264
## TEAM_BATTING_H:TEAM_BATTING_BB       -4.822e-05  1.086e-05  -4.441 9.38e-06
## log_TEAM_PITCHING_SO:log_TEAM_FIELDING_E  1.335e+00  6.062e-01   2.203  0.02771
##
## (Intercept)
## TEAM_BATTING_H                       ***
## TEAM_BATTING_BB                      ***
## log_TEAM_PITCHING_SO                 *
## log_TEAM_FIELDING_E                  **
## TEAM_BATTING_H:TEAM_BATTING_BB       ***
## log_TEAM_PITCHING_SO:log_TEAM_FIELDING_E *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.74 on 2269 degrees of freedom
## Multiple R-squared:  0.2407, Adjusted R-squared:  0.2387
## F-statistic: 119.9 on 6 and 2269 DF,  p-value: < 2.2e-16
```



The coefficient for TEAM_BATTING_H is positive (0.06802), indicating that an increase in team hits is associated with an increase in wins. This makes sense as more hits generally lead to more scoring opportunities.

Similarly, the positive coefficient for TEAM_BATTING_BB suggests that teams that walk more win more. This makes sense.

The negative coefficient for log_TEAM_PITCHING_SO might initially seem counterintuitive, as strikeouts are generally beneficial for pitchers. However, this effect is contextualized by its interaction with fielding errors. The benefits of high strike out number decreases if your team is making alot of errors (and letting the other team score). Even if you have a good pitcher, if your field players aren't good, you will lose.

The negative coefficient for log_TEAM_FIELDING_E aligns with expectations, as more errors typically hurt a team's chances of winning.

Interaction between TEAM_BATTING_H and TEAM_BATTING_BB: The negative term indicates that the positive effect of having more hits or walks on winning games diminishes slightly when both are high. Which means that if a team is already good at hitting, getting better at walks doesn't increase their wins as much as you'd expect.

Interaction between log_TEAM_PITCHING_SO and log_TEAM_FIELDING_E: The positive coefficient means that even though teams that strike out a lot of batters but make many errors might seem at a disadvantage, these errors don't hurt them as much in terms of winning games.

Model #3

In this model, I add TEAM_BASERUN_SB and TEAM_FIELDING_DP as two additional predictors.

TEAM_BASERUN_SB is number of stolen bases and might influence wins (since players that successfully steal bases are literally closer to scoring).

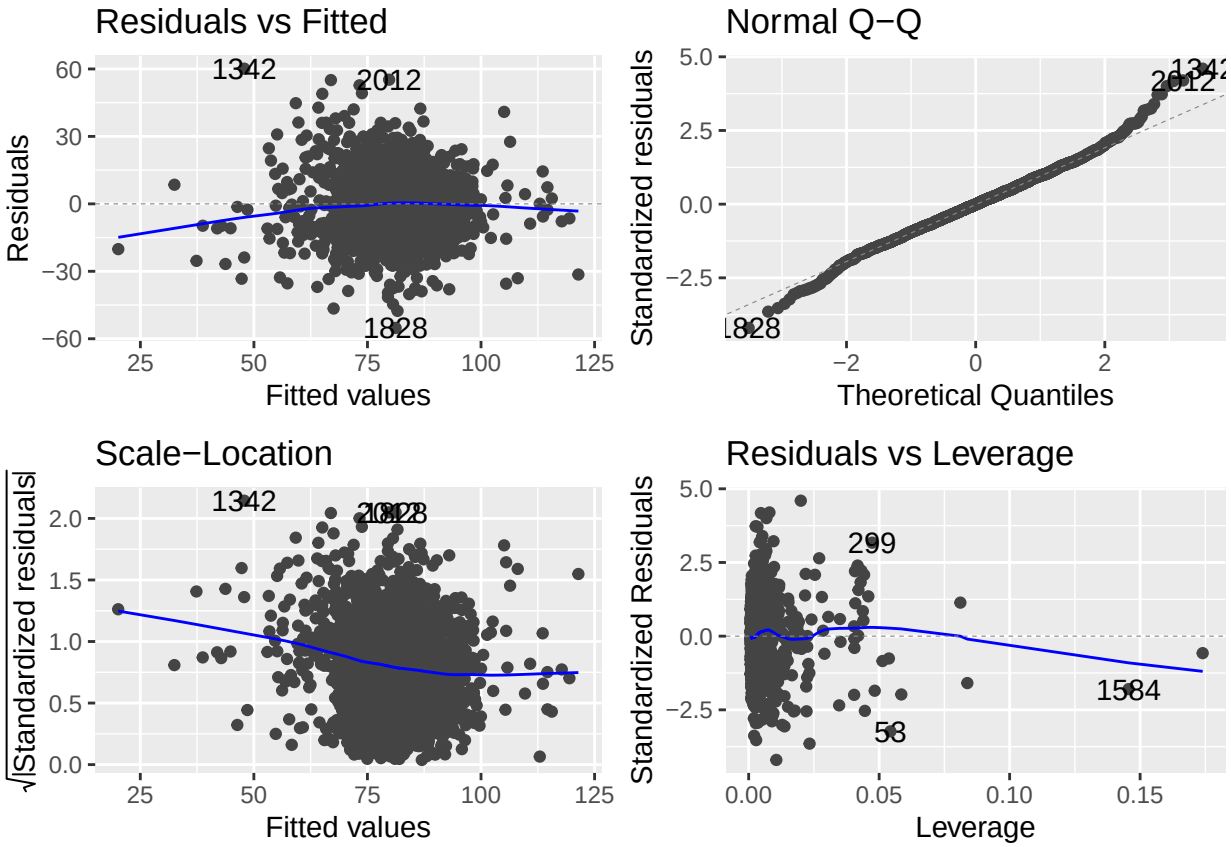
TEAM_FIELDING_DP tells us the number of double plays (or two outs in one play). Teams that have high numbers of double play are probably good in defense.'

I am also adding interaction terms:

TEAM_BATTING_H and TEAM_BASERUN_SB: explores how the effect of hits on winning is affected by the team's ability to steal bases (aggressive offensive strategy).

TEAM_PITCHING_SO and TEAM_FIELDING_DP: examines how the effect of strikeouts on wins is influenced by the team's ability to execute double plays

```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_BB +
##     TEAM_BASERUN_SB + log_TEAM_FIELDING_DP + log_TEAM_PITCHING_SO +
##     log_TEAM_FIELDING_E + TEAM_BATTING_H:log_TEAM_BASERUN_SB +
##     log_TEAM_PITCHING_SO:log_TEAM_FIELDING_DP, data = money_train_mod)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -55.197  -8.785  -0.108   8.380  60.117
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              6.005e+00  1.748e+02   0.034
## TEAM_BATTING_H                           5.069e-02  4.200e-03  12.069
## TEAM_BATTING_BB                          2.002e-02  3.175e-03   6.304
## TEAM_BASERUN_SB                          2.806e-02  8.698e-03   3.226
## log_TEAM_FIELDING_DP                     6.298e+00  3.498e+01   0.180
## log_TEAM_PITCHING_SO                     1.760e+01  2.640e+01   0.667
## log_TEAM_FIELDING_E                     -8.290e+00  7.366e-01 -11.255
## TEAM_BATTING_H:log_TEAM_BASERUN_SB       3.252e-04  7.482e-04   0.435
## log_TEAM_FIELDING_DP:log_TEAM_PITCHING_SO -3.631e+00  5.281e+00  -0.688
##                                          Pr(>|t|)
## (Intercept)                               0.97261
## TEAM_BATTING_H                            < 2e-16 ***
## TEAM_BATTING_BB                          3.47e-10 ***
## TEAM_BASERUN_SB                           0.00127 **
## log_TEAM_FIELDING_DP                      0.85713
## log_TEAM_PITCHING_SO                      0.50500
## log_TEAM_FIELDING_E                       < 2e-16 ***
## TEAM_BATTING_H:log_TEAM_BASERUN_SB        0.66392
## log_TEAM_FIELDING_DP:log_TEAM_PITCHING_SO  0.49174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.21 on 2267 degrees of freedom
## Multiple R-squared:  0.299,  Adjusted R-squared:  0.2966
## F-statistic: 120.9 on 8 and 2267 DF,  p-value: < 2.2e-16
```



TEAM_BATTING_H and TEAM_BATTING_BB both had positive coefficients (and statistically significant p-values), which makes sense. As TEAM_BATTING_H and TEAM_BATTING_BB increases, the chances of the batting team winning increases.

This makes sense in baseball world. As a batting team gets more hits, they have more chances of moving around the bases and scoring. Even if the batter hits a foul, as long as they do not get strikes out they can still get a hit and score (you can have 2 strikes and keep hitting fouls and still have a chance of getting a hit.)

The walking makes sense as well. If the pitching team keeps throwing bad pitches (that keep being balls instead of strikes), and the batter has enough skill to not blindly swing at these bad pitches, the batter will walk and go to first base, automatically increasing the chances of them scoring.

The TEAM_BASERUN_SB (stolen bases) predictor had a negative coefficient and was not statistically significant. This means that stolen bases alone do not impact a teams win. This makes sense. You can have increase numbers of stolen bases but that doesn't mean anything. For example, lets say a player steals 2 bases during an inning (and now is on 3rd base). If there is are 3 outs and that inning ends, those stolen bases would have been for nothing (did not lead the batting team closer to winning).

The TEAM_FIELDING_DP (double play) predictor had a negative coefficient and also was not statistically significant. This also means that double play count doesn't really predict a teams ability to win. This kind of makes sense. A team with large double play count for instance might have less strike out count (since the opposing team got more hits that led to those double plays).

log_TEAM_PITCHING_SO has a positive coefficient but has a high p-value. This means that as the log(# of strike outs the pitching team got) increases, so does the chances of the pitching team winning. This makes sense in terms of baseball. If a team has a good pitcher that throws good strikes, the chances of the other team winning decreases. Sometimes, there are even no-hitter games (where the batting team didn't even

get a hit/ all strike outs). However, since the p-value was not significant, this relationship is possibly not as strong in influencing the win of a game.

log_TEAM_FIELDING_E had a negative coefficient and a statistically significant p-value. As log(# of fielding error) increases, the chances of that team winning decreases. This makes sense in the baseball context. If players on the fields are missing and not catching the ball properly, they give the batting team an advantage and they can score more (AKA pitching team more likely to lose).

The interaction coefficient between TEAM_BASERUN_SB and TEAM_BATTING_H was positive and not significant, meaning hits and stolen bases combined do not impact the wins of this model. This makes sense. In baseball, hits seem more important than stolen bases. Without hits, you can't even steal bases (unless you walk but there has to be an opportunity to steal created).

The interaction coefficient between TEAM_FIELDING_DP and log_TEAM_PITCHING_SO was negative and significant. This means that pitching strikeouts and having alot of double plays contributes to winning. The negative coefficient makes sense. As more strikeouts are pitched, there is less of a chance of the opposing team getting a hit (which decreases your chances of having a double play). After all, double play happens when the batter hits (aka not a strike). But for a team that has less strikeouts, those double plays play an important role in preventing the other team from scoring (which increases your team from winning).

Part 4: SELECT MODELS

The objective of this project is to build a multiple linear regression model on the training data to predict the number of wins for the team. It is ideal to build a model with a lower level of complexity that can accomplish a higher level of accuracy, and if we can achieve a higher level of accuracy with a more complex model, that added complexity may be necessary.

In "Part 3" our team created three different models: a Basic Multiple Linear Regression, a Multiple Linear Regression with Interaction Terms, and a model that took into account a few additional variables (TEAM_BASERUN_SB and TEAM_FIELDING_DP) with the same basic concepts (Multiple Regression with Interactive Relationships) applied in "Model 2."

When looking at the three models predictive metrics, I want to start off with a simple AIC (Akaike Information Criterion) comparison. For context, when comparing models' AIC values, the model with the lowest AIC value is generally the best model in terms of "goodness of fit."

```
##        df      AIC
## model1  6 18415.20
## model2  8 18397.07
## model3 10 18219.07
```

After running instance of AIC() on all three models we get a result of very similar scores with different degrees of freedom. Looking at the instance of AIC() above, my initial choice would be to choose Model 3. Model 3 is definitely the most complex of all the three models ran; however, it also yields the minimum AIC score of the three models ran in this project.

I also want to go ahead and compare the three models using a BIC comparison (Bayesian Information Criterion). Similarly to AIC, we are looking for the lowest score, though we do want to consider the tradeoffs presented by model complexity.

```
##        df      BIC
## model1  6 18449.58
## model2  8 18442.91
## model3 10 18276.38
```

Again, similarly to the results yielded using the AIC comparison, Model 3 has the lowest score by a significant margin.

I do acknowledge that there is an about 15.6% size difference in the number of observations, and I also acknowledge that running both AIC and BIC comparisons yield a clear favorite in terms of goodness of fit. I will not base my final decision of model choice based on these metrics and I will use there comparisons to inform the ultimate decision.

Now, I want to look at the R^2 values for the respective models, so that I can assess how the models deal with variance present in the data.

```
## [1] 0.2332753
```

```
## [1] 0.2406953
```

```
## [1] 0.2990461
```

When comparing R^2 values for the respective models, it is clear that Model 3 yields the best R^2 value. However, it is notable that all R^2 are rather low, indicating that our models may not be explaining much of the variance in our outcome variable (TARGET_WINS).

Now, I want to check the RMSE for each model:

```
## [1] 13.79
```

```
## [1] 13.72311
```

```
## [1] 13.18528
```

Based on these RMSE values, again, it seems that Model 3 is the best predictor (the RMSE value for Model 3 is the lowest).

Considering the results from the comparison of AIC, BIC, R-squared values and RMSE values, it seems that Model 3 is the best fit.

Now, before I check the model's respective performance on the evaluation data, I have to prepare my model evaluation data set.

```
## Rows: 259 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (16): INDEX, TEAM_BATTING_H, TEAM_BATTING_2B, TEAM_BATTING_3B, TEAM_BATT...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Now, with my test data set prepared, I can run predictions:

Model 1 Predictions:

```
##        1        2        3        4        5        6        7        8
##  68.47796 70.61698 78.24544 85.59280 66.68353 68.68774 76.76128 71.60137
##        9       10       11       12       13       14       15       16
##  69.32554 76.35625 78.46261 82.16513 79.96128 82.00717 79.53065 78.69903
##       17       18       19       20       21       22       23       24
##  73.29053 83.17820 67.31757 92.24841 86.37149 90.45745 83.57861 77.78718
##       25       26       27       28       29       30       31       32
##  83.90215 86.85764 63.03390 77.16114 87.28807 80.70296 91.84798 87.72541
##       33       34       35       36       37       38       39       40
##  88.91603 90.74701 82.00144 82.76391 77.31562 87.37086 88.97987 88.16849
##       41       42       43       44       45       46       47       48
##  85.72882 90.97221 42.80070 90.28044 82.58886 87.75615 90.96612 70.26505
##       49       50       51       52       53       54       55       56
##  69.47034 76.33569 80.31980 89.31970 77.81323 73.57267 77.83776 78.61910
##       57       58       59       60       61       62       63       64
##  83.16827 68.29232 60.39180 74.11891 84.39369 84.18934 85.88091 83.93885
##       65       66       67       68       69       70       71       72
##  81.27166 86.34320 73.67254 81.15016 73.14794 80.90349 87.27874 79.73900
##       73       74       75       76       77       78       79       80
##  82.31907 84.31382 84.36246 85.17998 81.75111 79.87221 68.10919 74.10621
##       81       82       83       84       85       86       87       88
##  89.79843 90.08507 97.66952 85.10145 85.05696 82.33294 78.02291 83.69728
##       89       90       91       92       93       94       95       96
##  85.12017 88.80507 75.75173 89.69255 72.96295 77.08946 76.71142 74.80313
##       97       98       99      100      101      102      103      104
##  82.82936 99.27811 90.72468 90.11700 84.96323 76.94113 86.85963 82.12934
##      105      106      107      108      109      110      111      112
##  80.66905 72.84876 58.40121 82.28715 84.75566 65.24190 81.69256 81.60244
##      113      114      115      116      117      118      119      120
##  89.99723 87.54025 81.02702 82.80402 89.32796 80.27260 79.73554 69.38841
##      121      122      123      124      125      126      127      128
##  83.63929 64.51911 65.53191 58.44327 67.68797 85.60241 90.90657 73.68803
##      129      130      131      132      133      134      135      136
##  88.65674 94.15788 90.10322 83.99398 76.08153 82.52804 84.44425 70.58897
##      137      138      139      140      141      142      143      144
##  75.97798 79.92441 84.30356 81.45310 64.21630 67.02295 93.80400 80.62205
##      145      146      147      148      149      150      151      152
##  78.61276 77.10894 78.99058 84.17934 86.27933 81.84177 81.65542 83.72064
##      153      154      155      156      157      158      159      160
##  66.27771 73.58200 78.93707 72.99286 84.27667 68.22126 84.39587 70.21754
##      161      162      163      164      165      166      167      168
##  98.56829 101.61369 90.47931 101.10611 94.36859 89.80552 85.05110 82.26392
##      169      170      171      172      173      174      175      176
##  73.85485 83.34218 83.07165 82.79040 82.09016 92.55385 83.98388 80.24824
##      177      178      179      180      181      182      183      184
##  81.81402 77.67531 78.33362 81.39140 77.13893 82.32803 85.12806 84.75074
##      185      186      187      188      189      190      191      192
##  95.24747 83.41857 87.81040 66.38639 64.87952 107.22178 70.03538 76.98276
##      193      194      195      196      197      198      199      200
##  75.27298 81.67597 85.60031 73.09718 77.91672 81.84811 79.95058 85.84112
##      201      202      203      204      205      206      207      208
##  77.79563 82.10426 75.62524 86.06848 80.19496 79.16708 83.88850 79.80923
```

```
##       209       210       211       212       213       214       215       216
##  78.02861  72.59147  94.93724  87.71330  78.37485  71.18638  74.22580  87.31240
##       217       218       219       220       221       222       223       224
##  83.46706  85.26006  78.27363  77.77628  81.25923  76.25612  85.52366  79.93721
##       225       226       227       228       229       230       231       232
##  98.80188  75.78805  80.56183  81.68729  82.65557  75.45300  72.50972  89.94353
##       233       234       235       236       237       238       239       240
##  83.04689  84.96198  78.90582  74.59730  83.45875  76.46555  84.97615  71.27687
##       241       242       243       244       245       246       247       248
##  87.75740  89.59095  87.80515  84.91299  66.51165  88.85298  79.98730  82.16007
##       249       250       251       252       253       254       255       256
##  76.13087  83.83672  83.08111  65.23220  88.46968  38.25910  71.41294  79.68946
##       257       258       259
##  77.51379  79.63935  78.44970
```

Model 2 Predictions:

```
##         1         2         3         4         5         6         7         8
##  68.47498  71.93704  78.67711  84.84028  65.07076  67.62838  77.46424  71.24291
##         9        10        11        12        13        14        15        16
##  70.25938  77.00324  79.00192  82.24989  80.33076  82.36528  80.30900  78.78839
##        17        18        19        20        21        22        23        24
##  73.62634  82.79880  67.00731  91.65372  87.15673  89.87767  84.43569  78.80386
##        25        26        27        28        29        30        31        32
##  83.28074  86.13299  59.54566  77.67000  87.51138  80.85690  90.79622  87.81797
##        33        34        35        36        37        38        39        40
##  88.88810  89.96574  81.80819  83.02333  78.01767  86.41252  87.18698  87.26859
##        41        42        43        44        45        46        47        48
##  85.95980  90.39564  30.80870  88.23902  82.36397  86.35411  89.25310  70.44948
##        49        50        51        52        53        54        55        56
##  70.05753  77.29982  80.94095  89.34681  77.65008  73.46760  77.71903  78.46289
##        57        58        59        60        61        62        63        64
##  83.02545  68.47329  60.96988  74.16341  84.89284  86.26639  85.35578  83.14462
##        65        66        67        68        69        70        71        72
##  80.51057  86.42191  73.28574  81.18822  73.83179  81.47109  88.04472  80.90568
##        73        74        75        76        77        78        79        80
##  83.09089  84.23324  83.65656  84.08749  82.50976  79.50934  67.80324  73.80680
##        81        82        83        84        85        86        87        88
##  89.63085  89.46823  95.97903  85.10213  85.30748  82.75305  78.49100  84.01439
##        89        90        91        92        93        94        95        96
##  86.29283  87.60116  76.29436  89.68466  73.60820  76.66485  76.49066  75.11307
##        97        98        99       100       101       102       103       104
##  84.23573  96.52371  90.12962  89.54711  85.09470  78.49107  86.73523  82.94576
##       105       106       107       108       109       110       111       112
##  80.07050  72.27014  56.58168  81.23052  84.46044  64.47653  82.69157  82.10238
##       113       114       115       116       117       118       119       120
##  89.27039  87.11744  80.65818  82.49927  88.01777  80.34983  78.99670  69.38955
##       121       122       123       124       125       126       127       128
##  84.09426  64.99684  65.03342  57.32041  68.47565  86.93052  90.92724  74.06883
##       129       130       131       132       133       134       135       136
##  88.42089  93.08531  89.19973  85.19084  76.06382  81.43886  83.68067  69.89121
##       137       138       139       140       141       142       143       144
##  76.43999  80.20138  84.38677  81.56326  62.79413  66.82874  92.70159  80.96902
```

```
##       145       146       147       148       149       150       151       152
##  80.66299  78.30395  79.30141  84.77106  86.16393  82.01845  81.26024  83.08485
##       153       154       155       156       157       158       159       160
##  66.10823  72.74443  79.41485  72.66072  83.56926  67.38457  83.86226  70.39406
##       161       162       163       164       165       166       167       168
##  95.05767  96.92784  90.04520  96.70641  92.37055  90.08920  85.15732  83.09681
##       169       170       171       172       173       174       175       176
##  73.37678  83.15768  82.57971  83.16565  83.10378  92.13400  84.06010  82.29104
##       177       178       179       180       181       182       183       184
##  84.52903  78.11314  79.12344  82.03350  78.78140  82.25110  84.87169  84.57804
##       185       186       187       188       189       190       191       192
##  97.38994  83.04215  85.98665  65.27194  63.88790 102.36477  69.83662  76.78666
##       193       194       195       196       197       198       199       200
##  75.23245  82.10359  85.46927  74.18733  78.26737  84.01152  81.39991  85.30385
##       201       202       203       204       205       206       207       208
##  77.48646  81.74283  75.99231  85.24193  80.43516  80.68736  83.19524  79.68317
##       209       210       211       212       213       214       215       216
##  78.03977  73.16198  93.05617  87.69546  79.28612  72.50987  75.21600  86.59144
##       217       218       219       220       221       222       223       224
##  82.93894  84.58073  78.15838  78.52069  80.38516  75.34551  84.69347  78.95181
##       225       226       227       228       229       230       231       232
## 106.30597  77.05273  80.31235  81.58488  82.92132  74.95558  73.98352  89.97032
##       233       234       235       236       237       238       239       240
##  84.41221  84.65113  79.64675  75.45332  82.83548  76.41616  84.21278  71.38189
##       241       242       243       244       245       246       247       248
##  87.45530  89.82569  88.01010  84.60195  68.36469  88.69997  79.75642  81.56771
##       249       250       251       252       253       254       255       256
##  76.89641  82.76549  82.22311  63.89263  86.82126  31.88170  72.32240  82.11861
##       257       258       259
##  77.26489  80.02839  77.56642
```

Model 3 Predictions:

```
##         1         2         3         4         5         6         7         8
##  65.24781  66.45852  75.72673  87.40382  63.00840  69.09934  79.38624  76.92803
##         9        10        11        12        13        14        15        16
##  70.34468  74.17351  72.26134  77.40119  75.51190  78.85333  83.81683  73.90494
##        17        18        19        20        21        22        23        24
##  74.73702  80.24631  75.18191  88.99171  84.55310  86.53004  79.88355  71.28077
##        25        26        27        28        29        30        31        32
##  83.41293  88.59031  61.27405  74.78788  87.77528  79.53233  90.65068  84.76615
##        33        34        35        36        37        38        39        40
##  84.20598  84.08852  79.42112  82.75780  76.09789  85.12098  85.71845  89.23049
##        41        42        43        44        45        46        47        48
##  84.44720  95.45483  38.46920  98.27721  86.81218  90.05061  93.82007  76.27930
##        49        50        51        52        53        54        55        56
##  69.92968  78.99830  76.31354  87.60153  75.98822  71.91392  76.47378  79.73967
##        57        58        59        60        61        62        63        64
##  90.02578  76.25701  64.38239  80.11644  87.90866  78.56573  88.87764  82.51365
##        65        66        67        68        69        70        71        72
##  82.18337  94.78046  74.96453  81.05450  77.82726  88.33274  84.03260  73.95540
##        73        74        75        76        77        78        79        80
##  74.93669  82.54522  84.07620  82.83067  83.31270  83.32221  74.61179  78.55687
```

```
##       81       82       83       84       85       86       87       88
## 85.97531 88.14960 95.70722 77.72160 82.55871 79.12461 81.82602 84.27494
##       89       90       91       92       93       94       95       96
## 89.91667 91.55387 76.52107 91.09538 72.48822 87.87561 88.25262 86.78847
##       97       98       99      100      101      102      103      104
## 89.28615 103.19097 88.82927 87.47557 80.76605 73.40914 84.19872 81.33671
##      105      106      107      108      109      110      111      112
## 78.75580 65.83517 49.61748 80.12330 87.31284 58.89486 82.78643 84.85564
##      113      114      115      116      117      118      119      120
## 93.49715 91.02238 81.01327 82.18693 86.62908 81.15379 76.90032 71.79216
##      121      122      123      124      125      126      127      128
## 87.18376 71.70988 69.42142 69.15783 68.61261 85.31957 89.68622 75.14789
##      129      130      131      132      133      134      135      136
## 91.22414 90.24776 88.04312 82.79457 80.68485 85.26290 87.65784 75.74607
##      137      138      139      140      141      142      143      144
## 73.89148 77.85216 91.73989 82.83171 68.29466 73.81336 91.82266 74.84288
##      145      146      147      148      149      150      151      152
## 75.45360 71.51535 76.28520 81.72671 80.49289 82.82857 79.41215 86.47486
##      153      154      155      156      157      158      159      160
## 52.02487 69.82594 80.90038 70.73160 89.31595 60.99254 92.69605 75.87170
##      161      162      163      164      165      166      167      168
## 98.92922 104.37183 92.71317 99.80674 93.56078 86.10543 79.22363 78.39698
##      169      170      171      172      173      174      175      176
## 72.56338 81.68455 89.21052 89.83700 80.82248 94.76769 82.34895 75.49734
##      177      178      179      180      181      182      183      184
## 78.96178 69.47808 74.16172 79.43826 86.05608 86.90564 87.00812 84.60440
##      185      186      187      188      189      190      191      192
## 94.31720 92.44506 88.54394 57.46583 57.62907 112.59953 74.37601 82.87241
##      193      194      195      196      197      198      199      200
## 74.42755 77.97193 82.93994 68.22736 77.18908 83.75764 77.72087 85.06372
##      201      202      203      204      205      206      207      208
## 74.26834 81.68170 76.71092 89.93173 83.04825 83.94275 83.21270 82.39782
##      209      210      211      212      213      214      215      216
## 78.50719 70.62562 99.02821 88.08390 79.43841 64.26511 66.44081 83.19837
##      217      218      219      220      221      222      223      224
## 77.87329 92.04724 79.41216 80.79858 80.03289 76.40467 82.88060 75.25126
##      225      226      227      228      229      230      231      232
## 93.33115 74.63066 82.36009 81.24458 85.03099 68.26028 84.41561 89.93681
##      233      234      235      236      237      238      239      240
## 76.57952 84.30469 78.84165 76.78603 85.46341 76.73478 84.53398 70.01210
##      241      242      243      244      245      246      247      248
## 86.77123 86.96004 86.22456 83.87820 64.36277 87.23927 80.92035 83.41689
##      249      250      251      252      253      254      255      256
## 72.25294 83.20889 81.77218 65.28446 90.55543 25.07401 70.05718 76.79216
##      257      258      259
## 85.89388 83.43783 77.58740
```

Appendix

R Code

```
library(knitr)
library(readr)
library(ggplot2)
library(tidyverse)
library(missMethods)
library(dplyr)
library(naniar)
library(MASS)
library(ggfortify)
library(glmnet)
library(ggpubr)
library(corrplot)

set.seed(621)

###Import and Summarize the data
money_train <- read.csv(url("https://raw.githubusercontent.com/Mattr5541/DATA-621-Homework-1/main/moneyl



summary(money_train)

###Descriptive data analysis
dim(money_train)
colSums(is.na(money_train))
#box plot not including the index column

money_train %>%
  gather(variable, value, TARGET_WINS:TEAM_FIELDING_DP) %>%
  ggplot(., aes(x= variable, y=value)) +
  geom_boxplot() +
  facet_wrap(~variable, scales ="free", ncol = 4) +
  labs(x = element_blank(), y = element_blank())
```

```
## Warning: Removed 3478 rows containing non-finite values (`stat_boxplot()`).
```

```
money_train %>%
  keep(is.numeric) %>%
  gather(key, value, TARGET_WINS:TEAM_FIELDING_DP) %>%
  ggplot(aes(value)) +
    facet_wrap(~ key, scales = "free") +
    geom_histogram(bins = 35)
```

```
## Warning: Removed 3478 rows containing non-finite values (`stat_bin()`).
```

```
money_train <- money_train[, -1]
money_train %>%
  gather(variable, value, -TARGET_WINS) %>%
```

```
  ggplot(., aes(value, TARGET_WINS)) +
  geom_point(fill = "blue", color="blue") +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  facet_wrap(~variable, scales ="free", ncol = 4) +
  labs(x = element_blank(), y = "Wins")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 3478 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 3478 rows containing missing values (`geom_point()`).
```

```
# sorted correlation of explanatory variables correlated to TARGET_WINS, response variables.
correlation_matrix <- cor(drop_na(money_train))
correlation_coefficients <- correlation_matrix[, 1]
sorted_correlation <- sort(correlation_coefficients, decreasing = TRUE)
print(sorted_correlation)

correlation_matrix <- cor(drop_na(money_train))
corrplot(correlation_matrix, method = "color",
         addgrid.col = NA, tl.col = "black", tl.srt = 45,
         addCoef.col = "black",
         tl.cex = 0.5,
         number.cex = 0.5)

###Imputation and testing the imputation
## Vars with missing observations include: TEAM_BATTING_SO; TEAM_BASERUN_SB; TEAM_BASERUN_CS; TEAM_BATTI
###Team_BATTING_HBP contains over 2000 missing variables; everything else is within the range of 102-77

####First, I will make indices to flag missing values
TEAM_BATTING_SO_index <- is.na(money_train$TEAM_BATTING_SO)
TEAM_BASERUN_SB_index <- is.na(money_train$TEAM_BASERUN_SB)
TEAM_BASERUN_CS_index <- is.na(money_train$TEAM_BASERUN_CS)
TEAM_BATTING_HBP_index <- is.na(money_train$TEAM_BATTING_HBP)
TEAM_PITCHING_SO_index <- is.na(money_train$TEAM_PITCHING_SO)
TEAM_FIELDING_DP_index <- is.na(money_train$TEAM_FIELDING_DP)


### And now to impute

money_train_mod <- money_train %>% missMethods::impute_median()
summary(money_train_mod)

money_train <- money_train %>% mutate(INDEX = seq_len(nrow(money_train)))
money_train_mod <- money_train_mod %>% mutate(INDEX = seq_len(nrow(money_train_mod)))

### Now I will compare some variables to see if the median imputation worked
money_train_comp <- money_train$TEAM_BATTING_SO

data.frame(money_train_comp)

money_train_comp <- money_train %>% dplyr::select(INDEX, TEAM_BATTING_SO, TEAM_BASERUN_SB, TEAM_BATTING_
```

```r
money_train_mod_comp <- money_train_mod %>% dplyr::select(INDEX, TEAM_BATTING_SO, TEAM_BASERUN_SB, TEAM_

money_train_index <- cbind(TEAM_BATTING_SO_index, TEAM_BASERUN_SB_index, TEAM_BASERUN_CS_index, TEAM_BA
money_train_index <- as.data.frame(money_train_index)

money_train_comp <- merge(money_train_comp, money_train_mod_comp, by = "INDEX")
money_train_comp <- money_train_comp %>% cbind(money_train_index)


ggplot(money_train_comp, aes(TEAM_BATTING_SO.y, color = TEAM_BATTING_SO_index)) + geom_histogram()
```

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

```r
ggplot(money_train_comp, aes(TEAM_BASERUN_SB.y, color = TEAM_BASERUN_SB_index)) + geom_histogram()
```

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

```r
ggplot(money_train_comp, aes(TEAM_BASERUN_CS.y, color = TEAM_BASERUN_CS_index)) + geom_histogram()
```

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

```r
ggplot(money_train_comp, aes(TEAM_BATTING_HBP.y, color = TEAM_BATTING_HBP_index)) + geom_histogram() ##
```

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

```r
##been entirely appropriate for this variable, since such a large proportion was missing (~88%). Everyt
ggplot(money_train_comp, aes(TEAM_PITCHING_SO.y, color = TEAM_PITCHING_SO_index)) + geom_histogram()
```

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

```r
ggplot(money_train_comp, aes(TEAM_FIELDING_DP.y, color = TEAM_FIELDING_DP_index)) + geom_histogram()
```

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

```r
###QQ plots and histograms of imputed data
##TARGET_WINS
q1 <- ggplot(money_train_mod, aes(sample = TARGET_WINS)) + stat_qq() + stat_qq_line() + labs(title = "T

##TEAM_BATTING_H
q2 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_H)) + stat_qq() + stat_qq_line() + labs(title =

##TEAM_BATTING_2B
q3 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_2B)) + stat_qq() + stat_qq_line() + labs(title =

##TEAM_BATTING_3B
q4 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_3B)) + stat_qq() + stat_qq_line() + labs(title =

##TEAM_BATTING_HR
q5 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_HR)) + stat_qq() + stat_qq_line() + labs(title =
```

```r
##TEAM_BATTING_BB
q6 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_BB)) + stat_qq() + stat_qq_line() + labs(title =

##TEAM_BATTING_SO
q7 <- ggplot(money_train_mod, aes(sample = TEAM_BATTING_SO)) + stat_qq() + stat_qq_line() + labs(title =

ggarrange(q1, q2, q3, q4, q5, q6, q7)


### Log transformations and histograms of log transformations

# Apply log transformations; and safely by add 1 to avoid log(0) in variables with 0's
money_train_mod <- money_train_mod %>% mutate(log_TEAM_FIELDING_E = log(TEAM_FIELDING_E))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_PITCHING_H = log(TEAM_PITCHING_H))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_PITCHING_BB = log(TEAM_PITCHING_BB))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_PITCHING_SO = log(TEAM_PITCHING_SO))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_BASERUN_SB = log(TEAM_BASERUN_SB))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_BATTING_3B = log(TEAM_BATTING_3B))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_BATTING_HR = log(TEAM_BATTING_HR))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_BATTING_SO = log(TEAM_BATTING_SO))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_BASERUN_CS = log(TEAM_BASERUN_CS))
money_train_mod <- money_train_mod %>% mutate(log_TEAM_FIELDING_DP = log(TEAM_FIELDING_DP))
money_train_mod$log_TEAM_BASERUN_CS <- log(money_train_mod$TEAM_BASERUN_CS + 1)
money_train_mod$log_TEAM_BATTING_HR <- log(money_train_mod$TEAM_BATTING_HR + 1)
money_train_mod$log_TEAM_BATTING_SO <- log(money_train_mod$TEAM_BATTING_SO + 1)
money_train_mod$log_TEAM_BATTING_3B <- log(money_train_mod$TEAM_BATTING_3B + 1)
money_train_mod$log_TEAM_BASERUN_SB <- log(money_train_mod$TEAM_BASERUN_SB + 1)
money_train_mod$log_TEAM_PITCHING_SO <- log(money_train_mod$TEAM_PITCHING_SO + 1)
money_train_mod$log_TEAM_FIELDING_E <- log(money_train_mod$TEAM_FIELDING_E + 1)
money_train_mod$log_TEAM_PITCHING_BB <- log(money_train_mod$TEAM_PITCHING_BB + 1)


par(mfrow=c(3,3))
hist(money_train_mod$log_TEAM_FIELDING_E, main = "log_TEAM_FIELDING_E")
hist(money_train_mod$log_TEAM_PITCHING_H, main = "log_TEAM_PITCHING_H")
hist(money_train_mod$log_TEAM_PITCHING_BB, main = "log_TEAM_PITCHING_BB")
hist(money_train_mod$log_TEAM_PITCHING_SO, main = "log_TEAM_PITCHING_SO")
hist(money_train_mod$TEAM_BASERUN_CS, main = "TEAM_BASERUN_CS")
hist(money_train_mod$TEAM_BATTING_SO, main = "TEAM_BATTING_SO")
hist(money_train_mod$TEAM_BATTING_BB, main = "TEAM_BATTING_BB")
hist(money_train_mod$TEAM_BATTING_3B, main = "TEAM_BATTING_3B")


###Multiple Regression models

model1 = lm(TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_BB + log_TEAM_PITCHING_SO+ log_TEAM_FIELDING_E

summary(model1)

autoplot(model1)

model2 = lm(TARGET_WINS ~ TEAM_BATTING_H * TEAM_BATTING_BB + log_TEAM_PITCHING_SO* log_TEAM_FIELDING_E
```

```r
summary(model2)

autoplot(model2)

model3 = lm(TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_BB + TEAM_BASERUN_SB + log_TEAM_FIELDING_DP + l


summary(model3)

autoplot(model3)

###Model validation via AIC and BIC criteria
AIC_values <- AIC(model1, model2, model3)

print(AIC_values)

BIC_values <- BIC(model1, model2, model3)

print(BIC_values)

###Comparing R^2 values
summary(model1)$r.squared
summary(model2)$r.squared
summary(model3)$r.squared

###Predictive modeling and checking the RMSE
predictions_train_1 = predict(model1, newdata = money_train_mod)
predictions_train_2 = predict(model2, newdata = money_train_mod)
predictions_train_3 = predict(model3, newdata = money_train_mod)

rmse1 = sqrt(mean(na.omit(predictions_train_1 - money_train_mod$TARGET_WINS)^2))
rmse2 = sqrt(mean(na.omit(predictions_train_2 - money_train_mod$TARGET_WINS)^2))
rmse3 = sqrt(mean(na.omit(predictions_train_3 - money_train_mod$TARGET_WINS)^2))

print(rmse1)
print(rmse2)
print(rmse3)

###Importing the evaluation dataset and transforming the data
evaluation = read_csv('moneyball-evaluation-data.csv')
```

```
## Rows: 259 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (16): INDEX, TEAM_BATTING_H, TEAM_BATTING_2B, TEAM_BATTING_3B, TEAM_BATT...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
money_train_join = money_train %>%
  mutate(train_or_test = 'train')

evaluation_join = evaluation %>%
```

```
  mutate(train_or_test = 'test')

df = rbind(subset(money_train_join, select =-TARGET_WINS), evaluation_join)

df <- df %>%
  missMethods::impute_median() %>%
  filter(train_or_test == 'test')

df <- df %>% mutate(log_TEAM_FIELDING_E = log(TEAM_FIELDING_E))
df <- df %>% mutate(log_TEAM_PITCHING_H = log(TEAM_PITCHING_H))
df <- df %>% mutate(log_TEAM_PITCHING_BB = log(TEAM_PITCHING_BB))
df <- df %>% mutate(log_TEAM_PITCHING_SO = log(TEAM_PITCHING_SO))
df <- df %>% mutate(log_TEAM_BASERUN_SB = log(TEAM_BASERUN_SB))
df <- df %>% mutate(log_TEAM_BATTING_3B = log(TEAM_BATTING_3B))
df <- df %>% mutate(log_TEAM_BATTING_HR = log(TEAM_BATTING_HR))
df <- df %>% mutate(log_TEAM_BATTING_SO = log(TEAM_BATTING_SO))
df <- df %>% mutate(log_TEAM_BASERUN_CS = log(TEAM_BASERUN_CS))
df <- df %>% mutate(log_TEAM_FIELDING_DP = log(TEAM_FIELDING_DP))
df$log_TEAM_BASERUN_CS <- log(df$TEAM_BASERUN_CS + 1)
df$log_TEAM_BATTING_HR <- log(df$TEAM_BATTING_HR + 1)
df$log_TEAM_BATTING_SO <- log(df$TEAM_BATTING_SO + 1)
df$log_TEAM_BATTING_3B <- log(df$TEAM_BATTING_3B + 1)
df$log_TEAM_BASERUN_SB <- log(df$TEAM_BASERUN_SB + 1)
df$log_TEAM_PITCHING_SO <- log(df$TEAM_PITCHING_SO + 1)
df$log_TEAM_FIELDING_E <- log(df$TEAM_FIELDING_E + 1)
df$log_TEAM_PITCHING_BB <- log(df$TEAM_PITCHING_BB + 1)

###Making predictions for the evaluation data
predictions_1 = predict(model1, newdata = df)

print(predictions_1)

predictions_2 = predict(model2, newdata = df)

print(predictions_2)

predictions_3 = predict(model3, newdata = df)

print(predictions_3)
```