Data 621

Group 4 - Homework1

Devin Teran, Atina Karim, Tom Hill, Amit Kapoor

3/1/2020

Contents

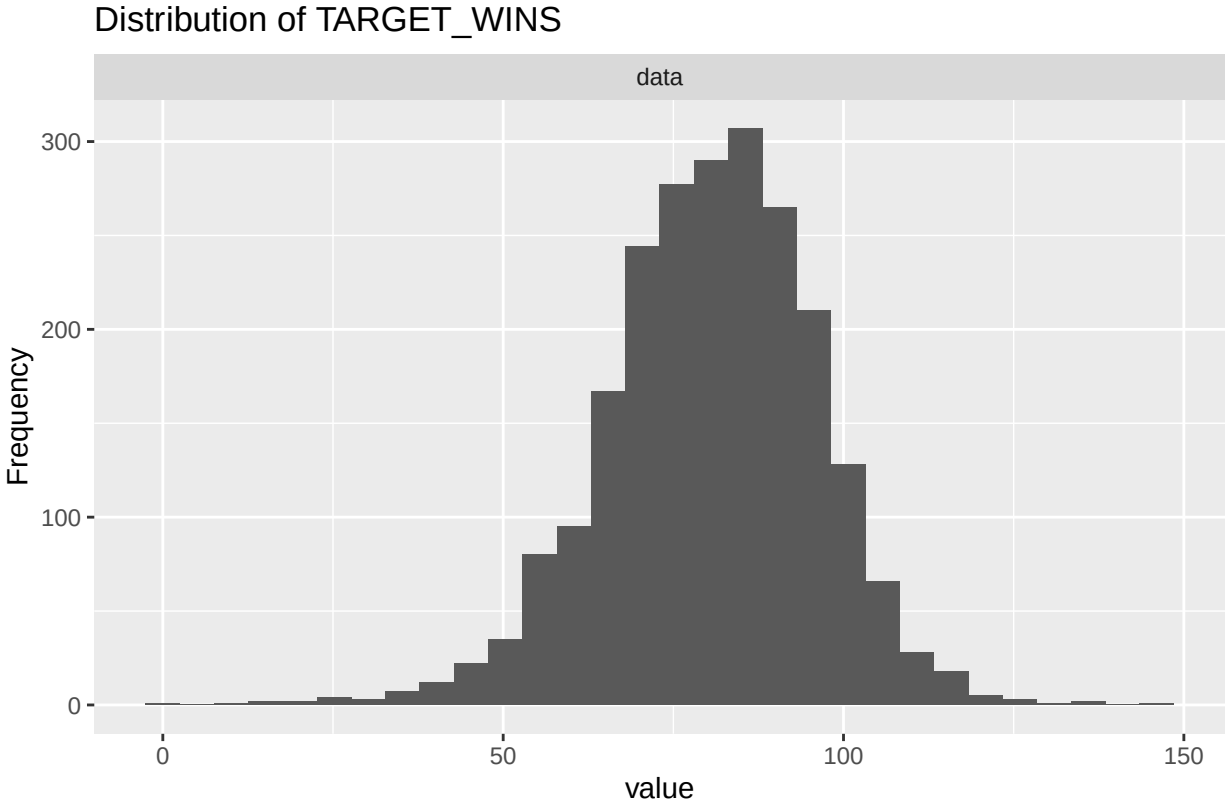
DATA EXPLORATION

The data set contains approximately 2276 records. Each record represents a professional baseball team from the years 1871 to 2006 inclusive. Each record has the performance of the team for the given year, with all of the statistics adjusted to match the performance of a 162 game season.Below is a short description of the variables

- INDEX - Identification Variable
- TARGET_WINS - Number of wins
- TEAM_BATTING_H - Base Hits by batters (1B,2B,3B,HR)
- TEAM_BATTING_2B - Doubles by batters (2B)
- TEAM_BATTING_3B - Triples by batters (3B)
- TEAM_BATTING_HR - Homeruns by batters (4B)
- TEAM_BATTING_BB - Walks by batters

- TEAM_BATTING_HBP - Batters hit by pitch (get a free base)
- TEAM_BATTING_SO - Strikeouts by batters
- TEAM_BASERUN_SB - Stolen bases
- TEAM_BASERUN_CS - Caught stealing
- TEAM_FIELDING_E - Errors
- TEAM_FIELDING_DP - Double Plays
- TEAM_PITCHING_BB - Walks allowed
- TEAM_PITCHING_H - Hits allowed
- TEAM_PITCHING_HR - Homeruns allowed
- TEAM_PITCHING_SO - Strikeouts by pitchers

The wins distribution seems nearly normal distributed. It indicates that seasons do not have many too many high or low number of wins.



Objective To build a multiple linear regression model on the training data to predict *TARGET_WINS*, which is the number of wins for the team.

Summary

Table 1: Data summary

Name	baseball_df
Number of rows	2276
Number of columns	17
Column type frequency:	
numeric	17

Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
INDEX	0	1.00	1268.46	736.35	1	630.75	1270.5	1915.50	2535
TARGET_WINS	0	1.00	80.79	15.75	0	71.00	82.0	92.00	146
TEAM_BATTING_H	0	1.00	1469.27	144.59	891	1383.00	1454.0	1537.25	2554
TEAM_BATTING_2B	0	1.00	241.25	46.80	69	208.00	238.0	273.00	458
TEAM_BATTING_3B	0	1.00	55.25	27.94	0	34.00	47.0	72.00	223
TEAM_BATTING_HR	0	1.00	99.61	60.55	0	42.00	102.0	147.00	264
TEAM_BATTING_BB	0	1.00	501.56	122.67	0	451.00	512.0	580.00	878
TEAM_BATTING_SO	102	0.96	735.61	248.53	0	548.00	750.0	930.00	1399
TEAM_BASERUN_SB	131	0.94	124.76	87.79	0	66.00	101.0	156.00	697
TEAM_BASERUN_CS	772	0.66	52.80	22.96	0	38.00	49.0	62.00	201
TEAM_BATTING_HBP	2085	0.08	59.36	12.97	29	50.50	58.0	67.00	95
TEAM_PITCHING_H	0	1.00	1779.21	1406.84	1137	1419.00	1518.0	1682.50	30132
TEAM_PITCHING_HR	0	1.00	105.70	61.30	0	50.00	107.0	150.00	343
TEAM_PITCHING_BB	0	1.00	553.01	166.36	0	476.00	536.5	611.00	3645
TEAM_PITCHING_SO	102	0.96	817.73	553.09	0	615.00	813.5	968.00	19278
TEAM_FIELDING_E	0	1.00	246.48	227.77	65	127.00	159.0	249.25	1898
TEAM_FIELDING_DP	286	0.87	146.39	26.23	52	131.00	149.0	164.00	228

```
##      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
```

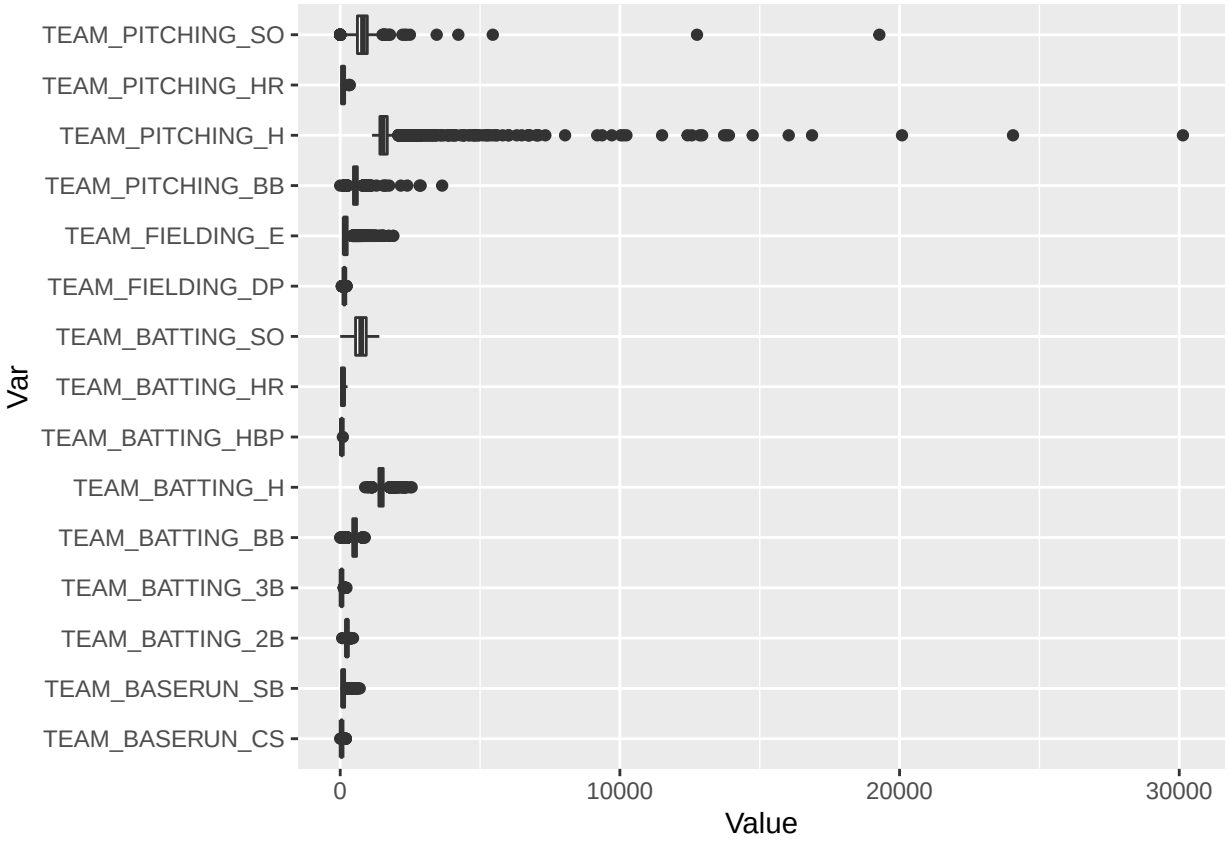
```
## [1] "Observations per year, 1871 - 2006:"
```

```
## [1] 16.86
```

The summary views above gives a quick overview of the dataset in terms of missing observation (and subsequently the completion % out of 2276 records for each variable) averages, standard deviations, quartiles and percentiles, minimum and maximum values and distributions. All the datatypes seem to be numeric. Observations span 128 years, with an average of 17 teams playing per year.

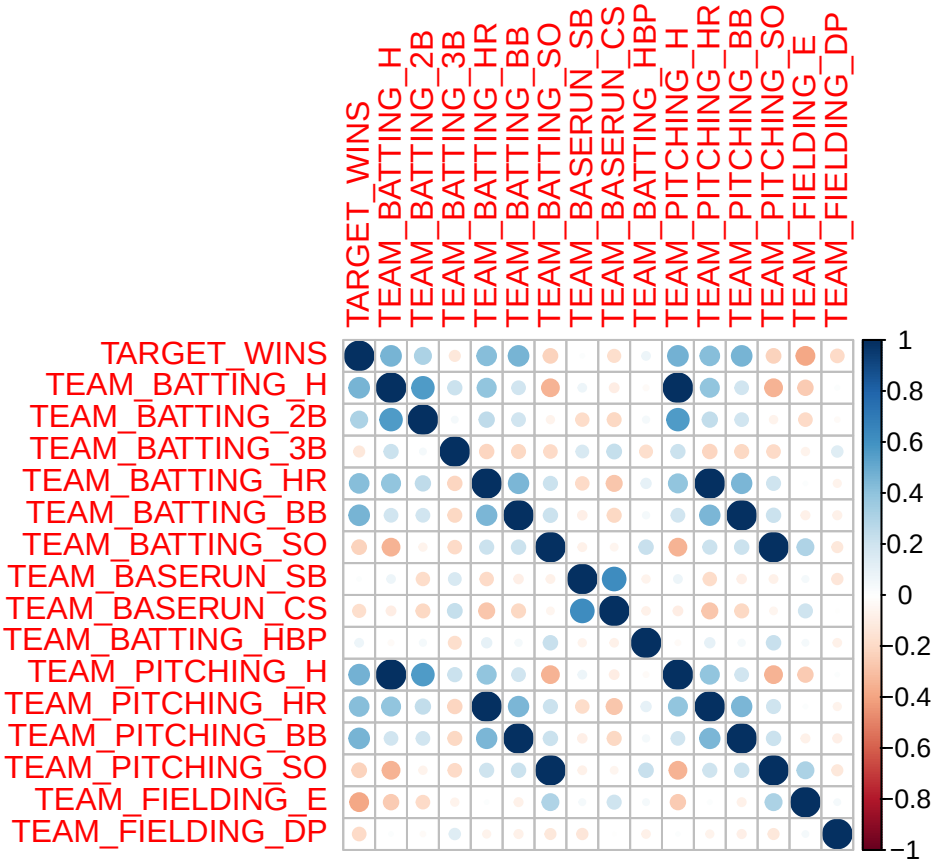
There are several variables with skewed distributions that could benefit from transformation.Additionally, there are a few variables with bi-modal distributions. Moreover, certain variables such as TEAM_BATTING_HBP have a lot of missing data (2085 out of 2276 obs.) which lowers its completion rate to about just 8%.

Outliers



From the boxplot above, we can see that several data columns like TEAM_PITCHING_H AND TEAM_PITCHING_SO have extreme outliers. The assignment mentions that some of the season records were adjusted to match the performance during a 162-game season.

Correlation and Collinearity



Looking at the correlation plot, there appear to be several statistically significant correlations between explanatory variables and the target.
From an initial inspection, it appears the team should focus on getting players on base through hits or walks.Contrary to what was expected, teams can still win if the pitchers allow homeruns, hits and walks to the other team.

Variables with Highest Positive Correlation with TARGET_WINS:
* TEAM_BATTING_H = 0.47 * TEAM_BATTING_HR = 0.42 * TEAM_BATTING_BB = 0.47 * TEAM_PITCHING_H = 0.47 * TEAM_PITCHING_HR = 0.42 * TEAM_PITCHING_BB = 0.47

To win more games it makes sense the team will need to make fewer errors.

Within this group, we detected collinearity between some of the variables:

Positive Correlations between predictors:
* TEAM_PITCHING_H and TEAM_BATTING_H = 0.99
* TEAM_PITCHING_HR and TEAM_BATTING_HR = 0.99
* TEAM_PITCHING_BB and TEAM_BATTING_BB = 0.99

Negative Correlations between predictors:
* TEAM_PITCHING_SO and TEAM_BATTING_H = -0.34

* TEAM_PITCHING_SO and TEAM_PITCHING_H = -0.34

DATA PREPARATION

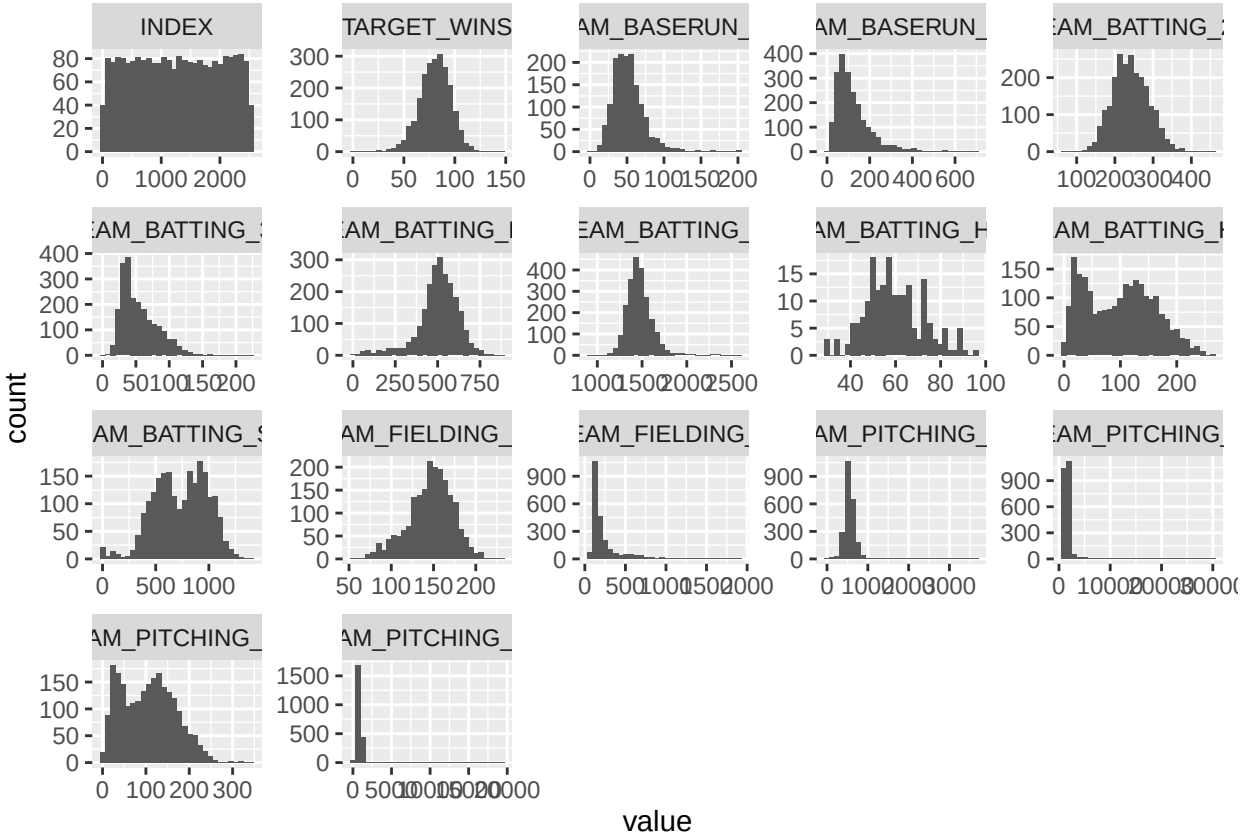
Missing values

In terms of missing values, there are two variables missing many observations. TEAM_BATTING_HBP is missing over 90% of its values, while TEAM_BASERUN_CS is missing just around 30%. Since TEAM_BATTING_HBP is barely complete and, deleting this variable would make sense.

The rest of the variables with missing values are: TEAM_BATTING_SO TEAM_BASERUN_SB TEAM_BASERUN_CS TEAM_FIELDING_DP TEAM_PITCHING_SO

Multiple Imputation

We will also attempt multiple imputation on the original dataset.Multiple imputation assumes normailty of data so let's check for skewness once again among the dataset:



It seems like TEAM_BASERUN_SB, TEAM_PITCHING_SO and TEAM_FIELDING_E are skewed. Let's confirm this using the skewness function. Anything that has a skewness above 1 is thought to be highly skewed.

```
## [1] 22.17455
```

```
## [1] 1.972414
```

```
## [1] 2.990466
```

Let's log transform these variables prior to multiple imputation.

Now that we have log transformed most of the variable and all our data are numeric, let's impute the data.

```
## -- Imputation 1 --
##
##   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##  41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##  61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##  81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##  101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##  121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##  141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##  161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##  181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##  201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##  221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##  241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##  261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
##  281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
##  301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
##  321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
##  341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##  361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
##  381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
##  401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
##  421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
##  441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
##  461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
##  481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
##  501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
##  521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
##  541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
##  561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
##  581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
##  601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
##  621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
##  641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
##  661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
##  681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
##  701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
##  721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
##  741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
##  761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
##  781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
##  801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
##  821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
##  841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
##  861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
##  881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
##  901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
##  921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
##  941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
##  961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
##  981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
```

```
##  1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
##  1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
##  1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
##  1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
##  1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
##  1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
##  1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
##  1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
##  1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
##  1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
##   1201 1202 1203 1204
##
## -- Imputation 2 --
##
##    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##   41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##   61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##   81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##   101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##   121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##   141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##   161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##   181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##   201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##   221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##   241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##   261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
##   281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
##   301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
##   321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
##   341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##   361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
##   381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
##   401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
##   421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
##   441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
##   461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
##   481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
##   501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
##   521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
##   541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
##   561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
##   581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
##   601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
##   621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
##   641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
##   661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
##   681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
##   701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
##   721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
##   741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
##   761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
##   781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
```

```
##  801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
##  821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
##  841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
##  861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
##  881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
##  901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
##  921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
##  941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
##  961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
##  981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
## 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
## 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
## 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
## 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
## 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
## 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
## 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
## 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
## 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
## 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
## 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
## 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240
## 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
## 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
## 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
## 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
## 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
## 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
## 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380
## 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400
## 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420
## 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
## 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460
## 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480
## 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500
## 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
## 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540
## 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560
## 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580
## 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
## 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620
## 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640
## 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660
## 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
## 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700
## 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720
## 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740
## 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
## 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780
## 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800
## 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820
## 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840
## 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860
## 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880
```

```
##  1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
##  1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920
##  1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940
##  1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960
##  1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980
##  1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000
##  2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
##  2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040
##  2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060
##  2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080
##  2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100
##   2101 2102 2103 2104 2105 2106 2107 2108 2109 2110
##
## -- Imputation 3 --
##
##    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##   41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##   61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##   81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##   101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##   121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##   141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##   161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##   181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##   201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##   221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##   241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##   261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
##   281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
##   301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
##   321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
##   341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##   361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
##   381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
##   401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
##   421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
##   441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
##   461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
##   481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
##   501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
##   521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
##   541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
##   561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
##   581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
##   601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
##   621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
##   641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
##   661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
##   681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
##   701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
##   721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
##   741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
##   761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
```

```
##  781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
##  801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
##  821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
##  841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
##  861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
##  881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
##  901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
##  921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
##  941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
##  961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
##  981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
## 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
## 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
## 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
## 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
## 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
## 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
## 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
## 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
## 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
## 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
## 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
## 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240
## 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
## 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
## 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
## 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
## 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
## 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
## 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380
## 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400
## 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420
## 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
## 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460
## 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480
## 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500
## 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
## 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540
## 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560
## 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580
## 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
## 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620
## 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640
## 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660
## 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
## 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700
## 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720
## 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740
## 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
## 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780
## 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800
## 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820
## 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840
## 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860
```

```
##  1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880
##  1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
##  1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920
##  1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940
##  1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960
##  1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980
##  1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000
##  2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
##  2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040
##  2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060
##  2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080
##  2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100
##  2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120
##  2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140
##  2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160
##  2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180
##  2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200
##  2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220
##  2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240
##  2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260
##  2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280
##  2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300
##  2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320
##  2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340
##  2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360
##  2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380
##  2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400
##  2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420
##  2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440
##  2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460
##  2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480
##  2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500
##  2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520
##  2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540
##  2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560
##  2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580
##  2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600
##  2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620
##  2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640
##  2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660
##  2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680
##  2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700
##  2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720
##  2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740
##  2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760
##  2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780
##  2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800
##  2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820
##  2821 2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840
##  2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 2860
##  2861 2862 2863 2864 2865 2866 2867 2868 2869 2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 2880
##  2881 2882 2883 2884 2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 2900
##  2901 2902 2903 2904 2905 2906 2907 2908 2909 2910 2911 2912 2913 2914 2915 2916 2917 2918 2919 2920
##  2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940
```

```
##  2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953 2954 2955 2956 2957 2958 2959 2960
##  2961 2962 2963 2964 2965 2966 2967 2968 2969 2970 2971 2972 2973 2974 2975 2976 2977 2978 2979 2980
##  2981 2982 2983 2984 2985 2986 2987 2988 2989 2990 2991 2992 2993 2994 2995 2996 2997 2998 2999 3000
##  3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 3011 3012 3013 3014 3015 3016 3017 3018 3019 3020
##  3021 3022 3023 3024 3025 3026 3027 3028 3029 3030 3031 3032 3033 3034 3035 3036 3037 3038 3039 3040
##  3041 3042 3043 3044 3045 3046 3047 3048 3049 3050 3051 3052 3053 3054 3055 3056 3057 3058 3059 3060
##  3061 3062 3063 3064 3065 3066 3067 3068 3069 3070 3071 3072 3073 3074 3075 3076 3077 3078 3079 3080
##  3081 3082 3083 3084 3085 3086 3087 3088 3089 3090 3091 3092 3093 3094 3095 3096 3097 3098 3099 3100
##  3101 3102 3103 3104 3105 3106 3107 3108 3109 3110 3111 3112 3113 3114 3115 3116 3117 3118 3119 3120
##  3121 3122 3123 3124 3125 3126 3127 3128 3129 3130 3131 3132 3133 3134 3135 3136 3137 3138 3139 3140
##  3141 3142 3143 3144 3145 3146 3147 3148 3149 3150 3151 3152 3153 3154 3155 3156 3157 3158 3159
##
## -- Imputation 4 --
##
##    1   2   3   4   5   6   7   8   9 10 11 12 13 14 15 16 17 18 19 20
##   21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##   41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##   61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##   81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##   101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##   121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##   141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##   161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##   181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##   201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##   221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##   241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##   261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
##   281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
##   301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
##   321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
##   341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##   361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
##   381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
##   401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
##   421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
##   441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
##   461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
##   481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
##   501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
##   521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
##   541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
##   561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
##   581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
##   601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
##   621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
##   641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
##   661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
##   681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
##   701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
##   721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
##   741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
##   761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
##   781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
```

```
##  801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
##  821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
##  841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
##  861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
##  881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
##  901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
##  921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
##  941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
##  961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
##  981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
## 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
## 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
## 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
## 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
## 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
## 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
## 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
## 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
## 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
## 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
## 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
## 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240
## 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
## 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
## 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
## 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
## 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
## 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
## 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380
## 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400
## 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420
## 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
## 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460
## 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480
## 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500
## 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
## 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540
## 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560
## 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580
## 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
## 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620
## 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640
## 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660
## 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
## 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700
## 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720
## 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740
## 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
##  1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776
##
## -- Imputation 5 --
##
##   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
```

```
##  41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##  61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##  81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##  101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##  121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##  141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##  161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##  181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##  201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##  221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##  241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##  261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
##  281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
##  301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
##  321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
##  341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##  361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
##  381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
##  401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
##  421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
##  441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
##  461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
##  481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
##  501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
##  521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
##  541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
##  561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
##  581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
##  601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
##  621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
##  641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
##  661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
##  681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
##  701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
##  721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
##  741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
##  761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
##  781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
##  801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
##  821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
##  841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
##  861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
##  881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
##  901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
##  921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
##  941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
##  961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
##  981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
## 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
## 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
## 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
## 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
## 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
## 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
```

```
##  1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
##  1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
##  1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
##  1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
##  1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
##  1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240
##  1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
##  1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
##  1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
##  1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
##  1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
##  1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
##  1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380
##  1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400
##  1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420
##  1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
##  1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460
##  1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480
##  1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500
##  1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
##  1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540
##  1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560
##  1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580
##  1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
##  1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620
##  1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640
##  1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660
##  1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
##  1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700
##  1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720
##  1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740
##  1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
##  1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780
##  1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800
##  1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820
##  1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840
##  1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860
##  1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880
##  1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
##  1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920
##  1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940
##  1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960
##  1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980
##  1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000
##  2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
##  2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040
##  2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060
##  2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080
##   2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098
```

Notice that we have imputed the entire dataset.This is because, although certain variables do not have any missing data, they maybe helpful in predicting missing values in the variables that do have them (Faraway, 2014)

BUILDMODELS

Model 1

We will build the initial models on the original dataset first (prior to multiple imputation) for comparison purposes.

We want to try creating a simple model with fewer predictors to see how it performs compared to our other models. To start, we chose a few variables that were highly positively and negatively correlated with TARGET_WINS.

From there we removed multiple predictors at once. To do this we need to construct a null hypothesis test which states that removing the variables doesn't make a better model. We construct a F-test and compare both versions of the model. If the p-value is under 0.05 we reject the null hypothesis, which indicates our new model isn't different than the first model. If the p-value is greater than 0.05, the model isn't better with those variables, so we will remove them. The simpler the model the better.

To determine which variables we removed,we chose the variable that was not proving to be significant in the linear regression (where the p-value was greater than 0.05). While this doesn't mean the variable itself isn't signficant, it means the variable alongside the other combination of variables in the model is not significant.

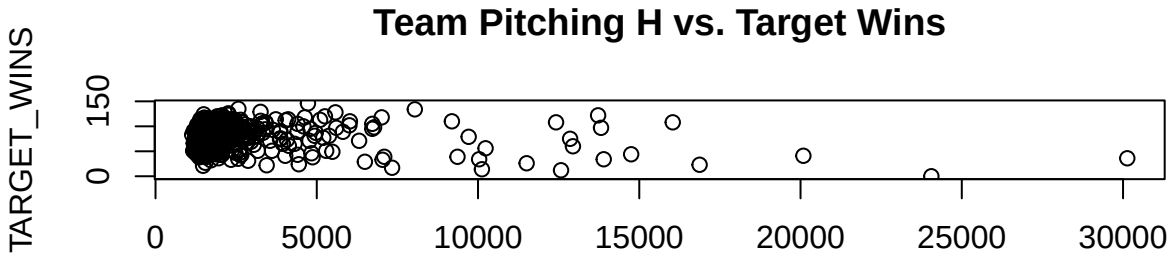
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +
##     TEAM_BATTING_BB + TEAM_BATTING_SO + TEAM_PITCHING_H + TEAM_PITCHING_HR +
##     TEAM_PITCHING_BB + TEAM_PITCHING_SO + TEAM_FIELDING_E + TEAM_FIELDING_DP,
##     data = baseball_df_fix)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.378  -7.898   0.200   7.831  39.906
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      54.808616   5.980591   9.164  < 2e-16 ***
## TEAM_BATTING_H    0.016475   0.004154   3.966 7.58e-05 ***
## TEAM_BATTING_HR   0.266831   0.057338   4.654 3.49e-06 ***
## TEAM_BATTING_BB   0.042737   0.019148   2.232 0.025733 *
## TEAM_BATTING_SO  -0.028802   0.009339  -3.084 0.002071 **
## TEAM_PITCHING_H   0.016827   0.002304   7.303 4.13e-13 ***
## TEAM_PITCHING_HR -0.207321   0.054299  -3.818 0.000139 ***
## TEAM_PITCHING_BB -0.010049   0.017800  -0.565 0.572447
## TEAM_PITCHING_SO  0.011148   0.008467   1.317 0.188131
## TEAM_FIELDING_E  -0.057509   0.005488 -10.478  < 2e-16 ***
## TEAM_FIELDING_DP -0.158032   0.012962 -12.192  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.42 on 1877 degrees of freedom
##   (388 observations deleted due to missingness)
## Multiple R-squared:  0.2991, Adjusted R-squared:  0.2954
## F-statistic:  80.1 on 10 and 1877 DF,  p-value: < 2.2e-16
```

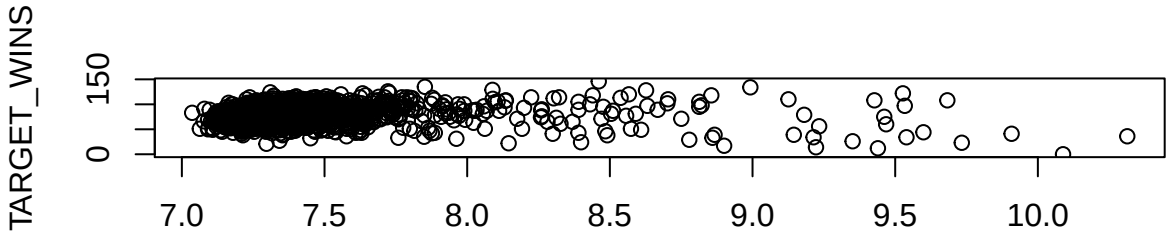
```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +
```

```
##     TEAM_BATTING_BB + TEAM_BATTING_SO + TEAM_PITCHING_H + TEAM_PITCHING_HR +
##     TEAM_FIELDING_E + TEAM_FIELDING_DP, data = baseball_df_fix)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.327  -7.799   0.209   7.875  40.210
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      54.574784   5.954596   9.165  < 2e-16 ***
## TEAM_BATTING_H    0.015205   0.003802   4.000 6.59e-05 ***
## TEAM_BATTING_HR   0.243859   0.032695   7.459 1.33e-13 ***
## TEAM_BATTING_BB   0.032029   0.003446   9.294  < 2e-16 ***
## TEAM_BATTING_SO  -0.016874   0.002207  -7.646 3.29e-14 ***
## TEAM_PITCHING_H   0.017731   0.001833   9.671  < 2e-16 ***
## TEAM_PITCHING_HR -0.184892   0.031112  -5.943 3.33e-09 ***
## TEAM_FIELDING_E  -0.055757   0.005293 -10.535  < 2e-16 ***
## TEAM_FIELDING_DP -0.156392   0.012867 -12.155  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.42 on 1879 degrees of freedom
##   (388 observations deleted due to missingness)
## Multiple R-squared:  0.2984, Adjusted R-squared:  0.2955
## F-statistic: 99.92 on 8 and 1879 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Model 1: TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR + TEAM_BATTING_BB +
##     TEAM_BATTING_SO + TEAM_PITCHING_H + TEAM_PITCHING_HR + TEAM_PITCHING_BB +
##     TEAM_PITCHING_SO + TEAM_FIELDING_E + TEAM_FIELDING_DP
## Model 2: TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR + TEAM_BATTING_BB +
##     TEAM_BATTING_SO + TEAM_PITCHING_H + TEAM_PITCHING_HR + TEAM_FIELDING_E +
##     TEAM_FIELDING_DP
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1877 244715
## 2   1879 244944 -2   -229.32 0.8794 0.4152
```

- Took the log of TEAM_PITCHING_H it's relationship to TARGET_WINS more linear






```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +
##     TEAM_BATTING_BB + TEAM_BATTING_SO + log(TEAM_PITCHING_H) +
##     TEAM_PITCHING_HR + TEAM_FIELDING_E + TEAM_FIELDING_DP, data = baseball_df_fix)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.996  -7.809   0.122   7.874  37.186
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -3.017e+02  3.377e+01  -8.933  < 2e-16 ***
## TEAM_BATTING_H       -1.227e-03  4.861e-03  -0.252    0.801
## TEAM_BATTING_HR       3.815e-01  4.108e-02   9.286  < 2e-16 ***
## TEAM_BATTING_BB       3.238e-02  3.437e-03   9.421  < 2e-16 ***
## TEAM_BATTING_SO      -1.557e-02  2.183e-03  -7.132 1.40e-12 ***
## log(TEAM_PITCHING_H)  5.567e+01  5.435e+00  10.243  < 2e-16 ***
## TEAM_PITCHING_HR     -3.208e-01  3.984e-02  -8.051 1.44e-15 ***
## TEAM_FIELDING_E      -6.390e-02  5.740e-03 -11.133  < 2e-16 ***
## TEAM_FIELDING_DP     -1.564e-01  1.283e-02 -12.194  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.38 on 1879 degrees of freedom
##   (388 observations deleted due to missingness)
## Multiple R-squared:  0.3025, Adjusted R-squared:  0.2995
## F-statistic: 101.8 on 8 and 1879 DF,  p-value: < 2.2e-16
```

- Remove TEAM_BATTING_H

```
##
## Call:
## lm(formula = TARGET_WINS ~ TEAM_BATTING_HR + TEAM_BATTING_BB +
##     TEAM_BATTING_SO + log(TEAM_PITCHING_H) + TEAM_PITCHING_HR +
##     TEAM_FIELDING_E + TEAM_FIELDING_DP, data = baseball_df_fix)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.997  -7.787   0.124   7.878  37.288
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -2.957e+02  2.409e+01 -12.274  < 2e-16 ***
## TEAM_BATTING_HR       3.747e-01  3.103e-02  12.073  < 2e-16 ***
## TEAM_BATTING_BB       3.231e-02  3.423e-03   9.438  < 2e-16 ***
## TEAM_BATTING_SO      -1.535e-02  1.994e-03  -7.696 2.25e-14 ***
## log(TEAM_PITCHING_H)  5.458e+01  3.284e+00  16.618  < 2e-16 ***
## TEAM_PITCHING_HR     -3.146e-01  3.134e-02 -10.037  < 2e-16 ***
## TEAM_FIELDING_E      -6.300e-02  4.494e-03 -14.018  < 2e-16 ***
## TEAM_FIELDING_DP     -1.564e-01  1.282e-02 -12.194  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.38 on 1880 degrees of freedom
##   (388 observations deleted due to missingness)
## Multiple R-squared:  0.3024, Adjusted R-squared:  0.2998
## F-statistic: 116.4 on 7 and 1880 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Model 1: TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR + TEAM_BATTING_BB +
##     TEAM_BATTING_SO + log(TEAM_PITCHING_H) + TEAM_PITCHING_HR +
##     TEAM_FIELDING_E + TEAM_FIELDING_DP
## Model 2: TARGET_WINS ~ TEAM_BATTING_HR + TEAM_BATTING_BB + TEAM_BATTING_SO +
##     log(TEAM_PITCHING_H) + TEAM_PITCHING_HR + TEAM_FIELDING_E +
##     TEAM_FIELDING_DP
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1879 243539
## 2   1880 243547 -1   -8.2595 0.0637 0.8007
```

Model 2

This model eliminates several features altogether from Model 1 including those with missing values, transforms three, and considers four different interaction effects.

```
##
## Call:
## lm(formula = TARGET_WINS ~ . - INDEX + log(TEAM_FIELDING_E) +
##     log(TEAM_PITCHING_BB) - TEAM_PITCHING_H - TEAM_BATTING_BB -
##     TEAM_PITCHING_HR - TEAM_PITCHING_BB - TEAM_FIELDING_E + log(TEAM_FIELDING_E) +
##     TEAM_BATTING_3B:TEAM_BATTING_HR + TEAM_BATTING_2B:TEAM_BATTING_HR +
##     TEAM_BATTING_H:TEAM_BATTING_HR + TEAM_BATTING_H:TEAM_BATTING_3B -
##     TEAM_BATTING_3B - TEAM_BATTING_SO - TEAM_BATTING_2B - TEAM_BATTING_BB -
##     TEAM_BATTING_HR - TEAM_BATTING_H - TEAM_BATTING_HR - TEAM_PITCHING_HR -
```

```
##     TEAM_BATTING_HBP - TEAM_FIELDING_DP - TEAM_PITCHING_SO -
##     TEAM_BASERUN_SB - TEAM_BASERUN_CS, data = baseball_df_fix)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.8825  -5.7136  -0.1331   6.3792  22.9085
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     2.143e+00  4.139e+01   0.052  0.95876
## log(TEAM_FIELDING_E)           -2.392e+01  4.325e+00  -5.530 1.09e-07 ***
## log(TEAM_PITCHING_BB)           2.508e+01  5.440e+00   4.611 7.49e-06 ***
## TEAM_BATTING_3B:TEAM_BATTING_HR -4.306e-03  1.427e-03  -3.018  0.00290 **
## TEAM_BATTING_2B:TEAM_BATTING_HR -3.449e-05  1.703e-04  -0.203  0.83971
## TEAM_BATTING_H:TEAM_BATTING_HR   1.434e-04  4.261e-05   3.366  0.00093 ***
## TEAM_BATTING_H:TEAM_BATTING_3B   4.332e-04  1.556e-04   2.783  0.00594 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.06 on 184 degrees of freedom
##   (2085 observations deleted due to missingness)
## Multiple R-squared:  0.4584, Adjusted R-squared:  0.4408
## F-statistic: 25.96 on 6 and 184 DF,  p-value: < 2.2e-16
```

The R-squared statistic indicates that this model predicts about half of the variation in wins with the included features. The model is statistically significant at p<.05, however the F-Statistic seems to have fallen quite a bit from the initial model.

Model 3

This model takes into account the imputed dataset and models TARGET_WINS against all variables present in the dataset except for INDEX (since this is the id variables and had a very weak negative association with TARGET_WINS)

```
##
## Call:
## lm(formula = TARGET_WINS ~ . - INDEX, data = missmod$imputations[[i]])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.105  -7.879  -0.086   8.055  43.164
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -23.836016  28.445175  -0.838 0.402140
## TEAM_BATTING_H     0.153522   0.063906   2.402 0.016371 *
## TEAM_BATTING_2B   -0.033816   0.008848  -3.822 0.000136 ***
## TEAM_BATTING_3B    0.079610   0.016635   4.786 1.82e-06 ***
## TEAM_BATTING_HR    0.368897   0.165439   2.230 0.025858 *
## TEAM_BATTING_BB   -0.344387   0.232176  -1.483 0.138134
## TEAM_BATTING_SO   -0.050566   0.006119  -8.265 2.36e-16 ***
## TEAM_BASERUN_SB    3.987815   0.771488   5.169 2.56e-07 ***
## TEAM_BASERUN_CS    0.074637   0.017358   4.300 1.78e-05 ***
## TEAM_BATTING_HBP   0.033273   0.021003   1.584 0.113289
```

```
## TEAM_PITCHING_H   -0.103930   0.064977  -1.599 0.109852
## TEAM_PITCHING_HR  -0.291142   0.165017  -1.764 0.077813 .
## TEAM_PITCHING_BB   0.363834   0.232686   1.564 0.118045
## TEAM_PITCHING_SO  23.548478   4.552359   5.173 2.51e-07 ***
## TEAM_FIELDING_E  -19.679926   1.056598 -18.626  < 2e-16 ***
## TEAM_FIELDING_DP  -0.159912   0.012724 -12.567  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.14 on 2260 degrees of freedom
## Multiple R-squared:  0.4096, Adjusted R-squared:  0.4057
## F-statistic: 104.5 on 15 and 2260 DF,  p-value: < 2.2e-16
```

The model above is for all of the predictors in the dataset(except for Index). It is statistically significant at p<.05 and the adjusted r squared for the model is 0.405, which means that about 40.5% of the variance in the dataset is explained by the model.

Model 4

We will modify the model a bit and eliminate variables that we had previously flagged for multicollinearity such as TEAM_PITCHING_HR,TEAM_PITCHING_BB and TEAM_PITCHING_H. This is important since multicollinearity can significantly reduce model performance.Out of these predictors, TEAM_PITCHING_H also had extreme outliers, along with TEAM_PITCHING_SO. Since the r-square is computed using the mean, variables with outliers will throw off this value. Therefore, although we have transformed TEAM_PITCHING_SO, it maybe best to still remove this variable from the model.

```
##
## Call:
## lm(formula = TARGET_WINS ~ . - INDEX - TEAM_PITCHING_HR - TEAM_PITCHING_BB -
##     TEAM_PITCHING_H - TEAM_PITCHING_SO, data = missmod$imputations[[i]])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.237  -7.926  -0.228   7.878  46.116
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.110e+02  7.640e+00  14.532  < 2e-16 ***
## TEAM_BATTING_H    4.342e-02  3.479e-03  12.480  < 2e-16 ***
## TEAM_BATTING_2B  -2.566e-02  8.537e-03  -3.006 0.002678 **
## TEAM_BATTING_3B   5.995e-02  1.616e-02   3.709 0.000213 ***
## TEAM_BATTING_HR   8.397e-02  9.464e-03   8.872  < 2e-16 ***
## TEAM_BATTING_BB   1.501e-02  2.939e-03   5.106 3.56e-07 ***
## TEAM_BATTING_SO  -2.028e-02  2.310e-03  -8.781  < 2e-16 ***
## TEAM_BASERUN_SB   3.505e+00  7.579e-01   4.624 3.97e-06 ***
## TEAM_BASERUN_CS   9.757e-02  1.677e-02   5.819 6.74e-09 ***
## TEAM_BATTING_HBP -7.407e-05  1.069e-04  -0.693 0.488362
## TEAM_FIELDING_E  -1.770e+01  9.918e-01 -17.849  < 2e-16 ***
## TEAM_FIELDING_DP -1.576e-01  1.277e-02 -12.340  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.26 on 2264 degrees of freedom
## Multiple R-squared:  0.3968, Adjusted R-squared:  0.3938
```

```
## F-statistic: 135.4 on 11 and 2264 DF,  p-value: < 2.2e-16
```

After removing variables that we had previously flagged for multicollinearity and outliers, we can see that the adjusted r-squared for the model drops a bit. However the F-Statistic seems to have improved.

Model 5

In addition to the above, this model considers the four different interaction effects from Model 2 above.

```
##
## Call:
## lm(formula = TARGET_WINS ~ . - INDEX - TEAM_PITCHING_HR - TEAM_PITCHING_BB -
##     TEAM_PITCHING_H - TEAM_PITCHING_SO + (TEAM_BATTING_H * TEAM_BATTING_2B +
##     TEAM_BATTING_H * TEAM_BATTING_3B + TEAM_BATTING_H * TEAM_BATTING_HR),
##     data = missmod$imputations[[i]])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.645  -7.961  -0.071   7.958  53.271
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.181e+02  1.352e+01   8.734  < 2e-16 ***
## TEAM_BATTING_H                  3.572e-02  8.336e-03   4.285 1.91e-05 ***
## TEAM_BATTING_2B                -1.802e-01  5.069e-02  -3.555 0.000385 ***
## TEAM_BATTING_3B                 2.834e-01  9.183e-02   3.086 0.002052 **
## TEAM_BATTING_HR                 4.098e-01  6.257e-02   6.550 7.10e-11 ***
## TEAM_BATTING_BB                 1.596e-02  2.930e-03   5.448 5.64e-08 ***
## TEAM_BATTING_SO                -2.250e-02  2.349e-03  -9.580  < 2e-16 ***
## TEAM_BASERUN_SB                 3.820e+00  7.630e-01   5.006 5.97e-07 ***
## TEAM_BASERUN_CS                 1.013e-01  1.695e-02   5.978 2.62e-09 ***
## TEAM_BATTING_HBP               -9.007e-05  1.084e-04  -0.831 0.405920
## TEAM_FIELDING_E                -1.778e+01  9.863e-01 -18.029  < 2e-16 ***
## TEAM_FIELDING_DP               -1.561e-01  1.274e-02 -12.252  < 2e-16 ***
## TEAM_BATTING_H:TEAM_BATTING_2B  1.105e-04  3.377e-05   3.272 0.001083 **
## TEAM_BATTING_H:TEAM_BATTING_3B -1.431e-04  5.800e-05  -2.468 0.013673 *
## TEAM_BATTING_H:TEAM_BATTING_HR -2.143e-04  4.088e-05  -5.242 1.74e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.19 on 2261 degrees of freedom
## Multiple R-squared:  0.4047, Adjusted R-squared:  0.4011
## F-statistic: 109.8 on 14 and 2261 DF,  p-value: < 2.2e-16
```

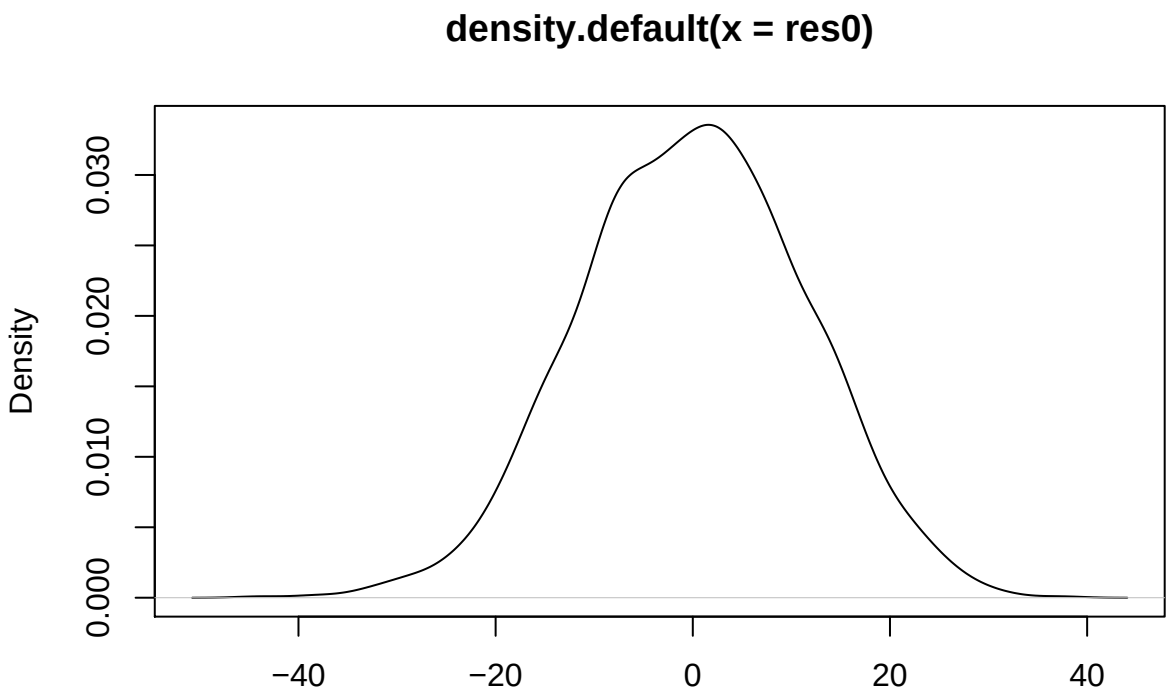
After adding the interaction effects, it seems that out adjusted r-sqaured has gone up to 0.40 (model explains 40% of the variance in the data). The model is statistically significant at $p<0.05$.

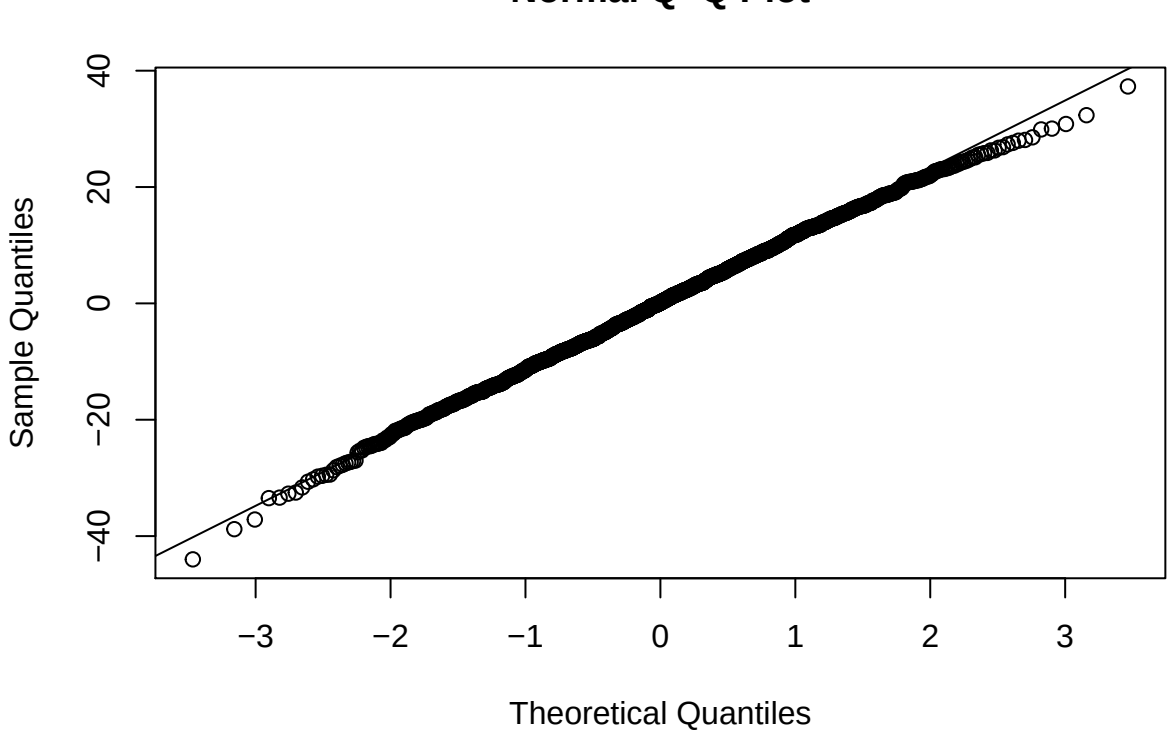
SELECT MODELS

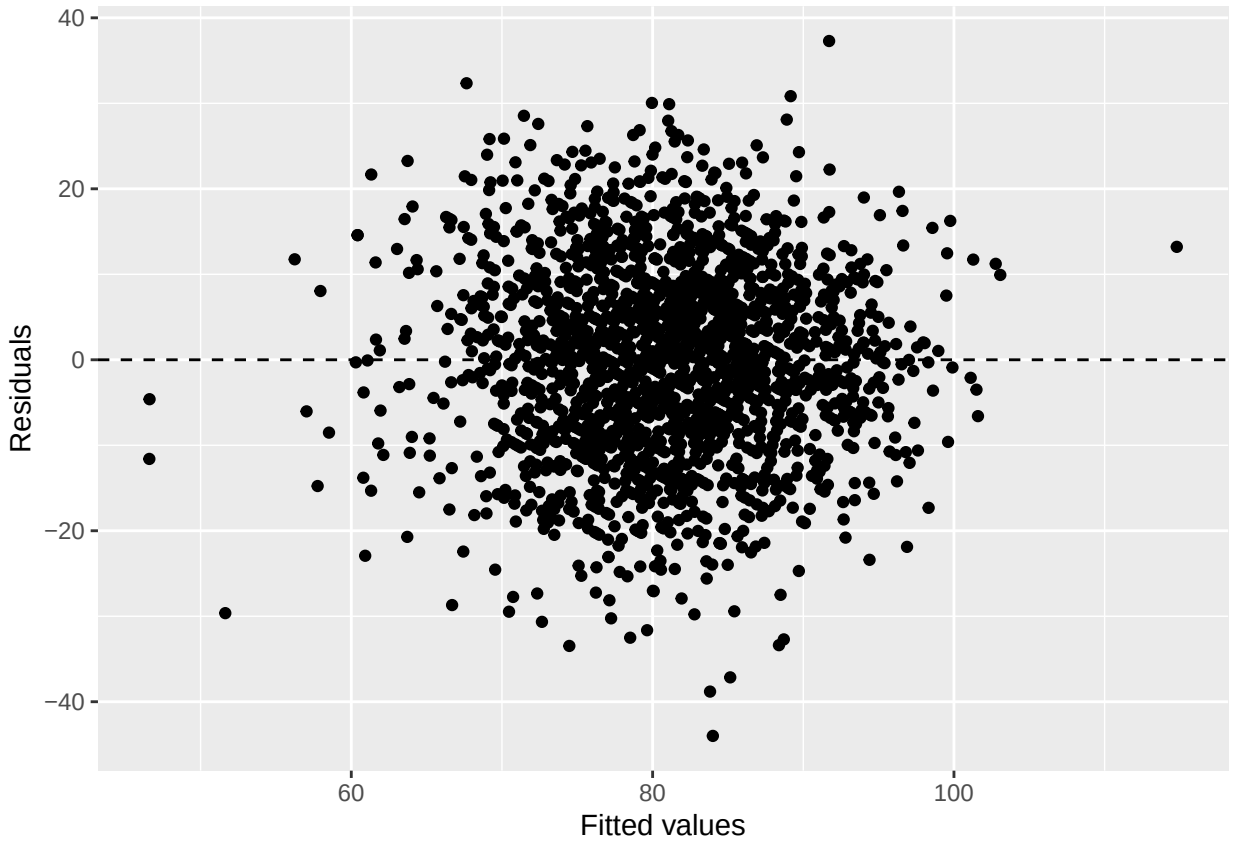
Diagnostics

We will look at the residual plots and model performance statistics (MSE and RMSE) for each of the models.

Model 1

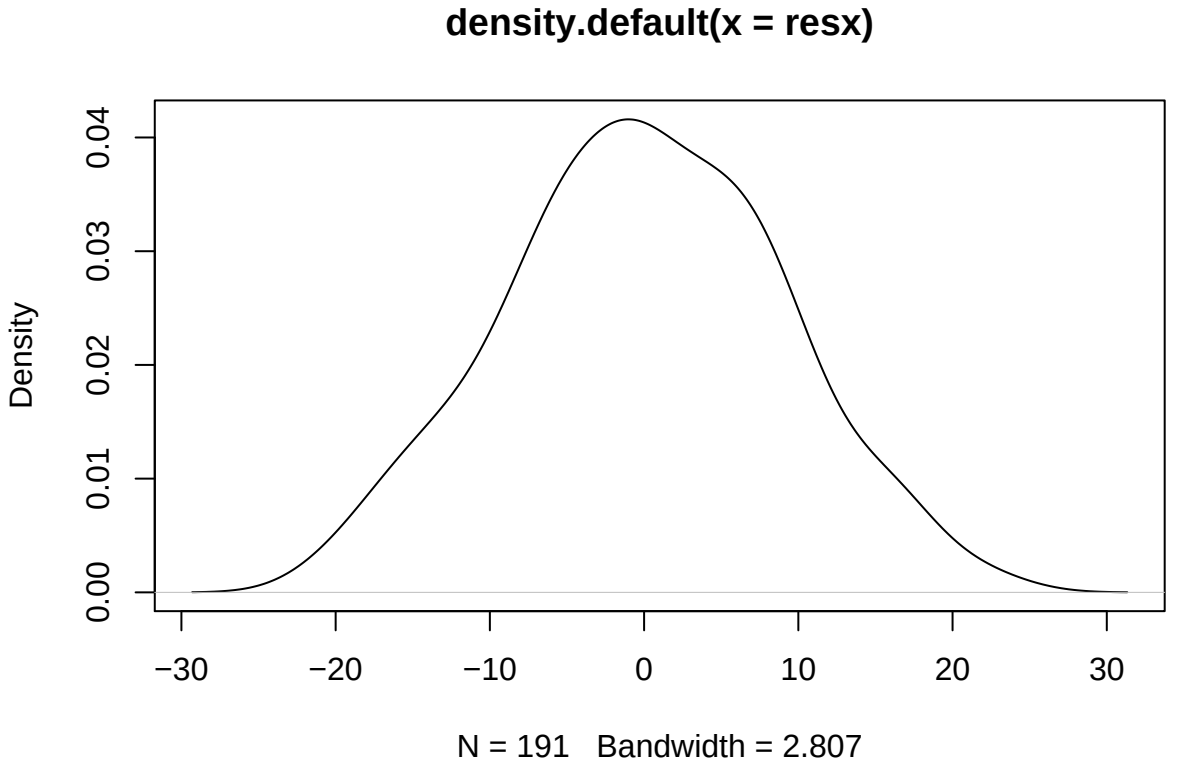

density.default(x = res0)
Density
0.000
0.010
0.020
0.030
−40
−20
0
20
40
N = 1888 Bandwidth = 2.262

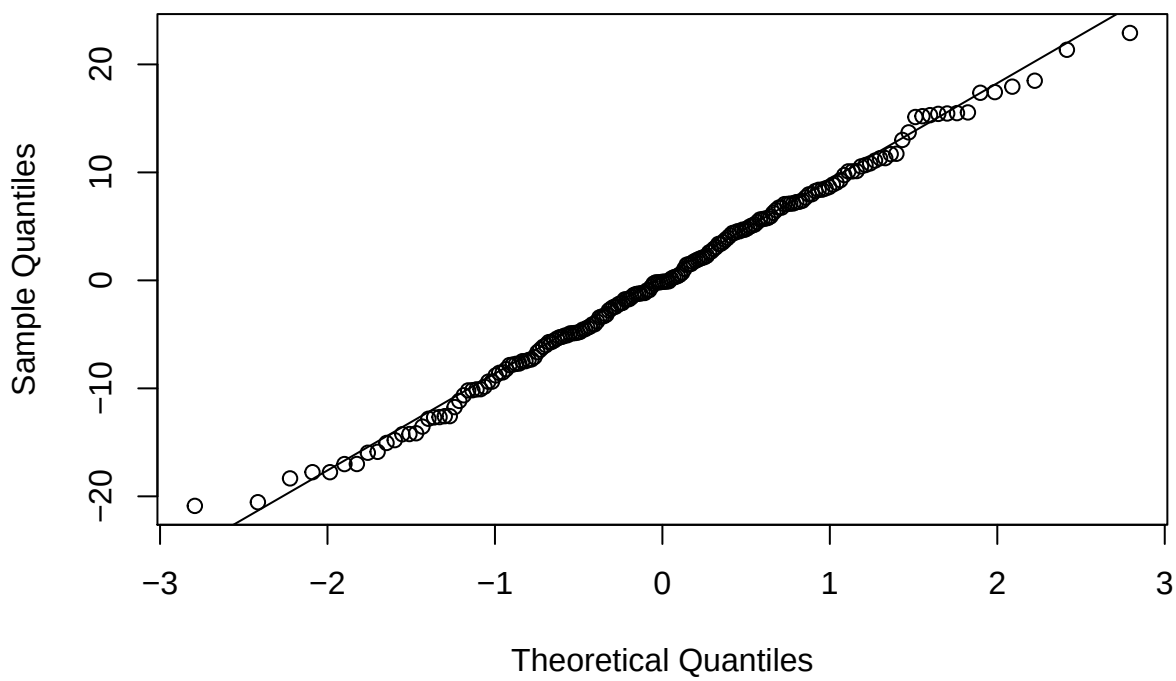

Normal Q−Q Plot
Sample Quantiles
−40
−20
0
20
40
−3
−2
−1
0
1
2
3
Theoretical Quantiles

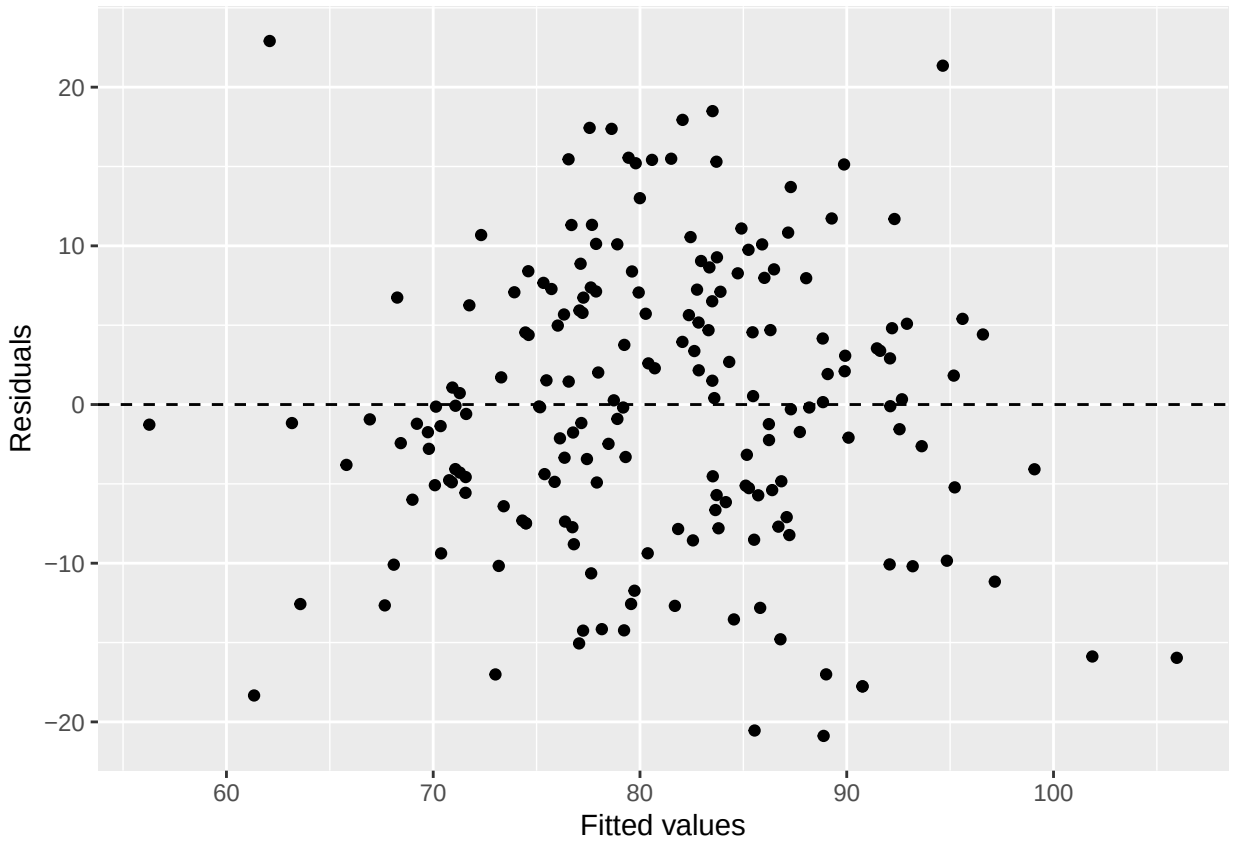



```
## [1] "Mean Squared Error: 128.997300425331"
## [1] "Root MSE: 11.3576978488306"
```

Model 2

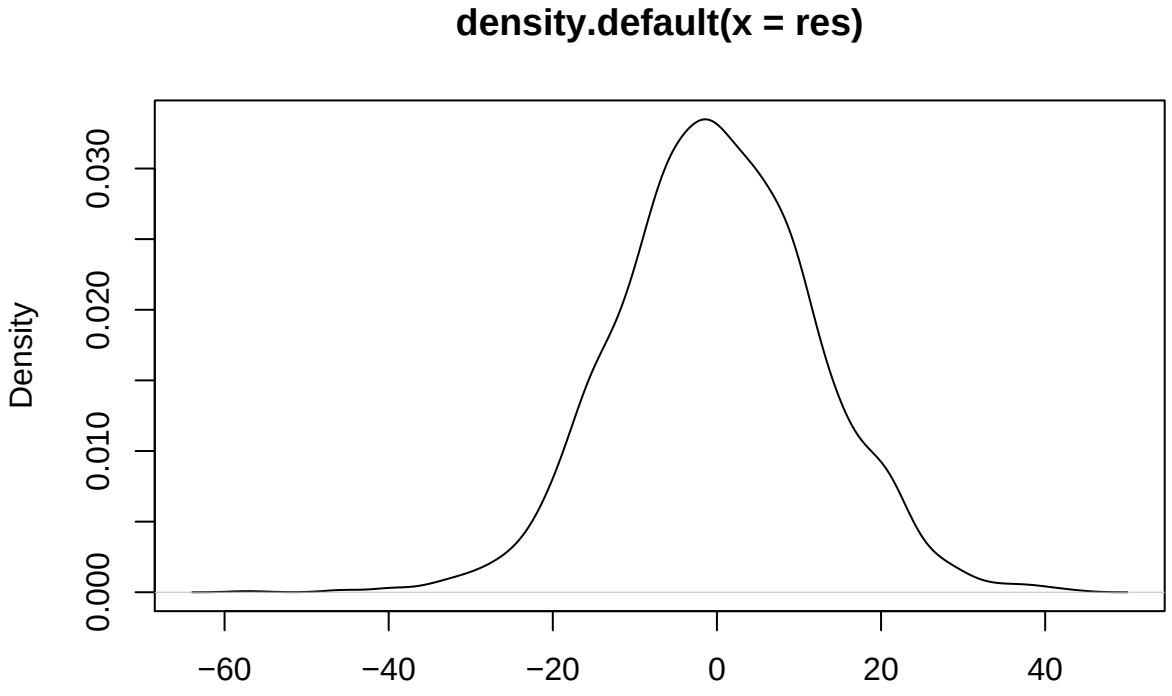

density.default(x = resx)
Density
0.00
0.01
0.02
0.03
0.04
−30
−20
−10
0
10
20
30
N = 191 Bandwidth = 2.807

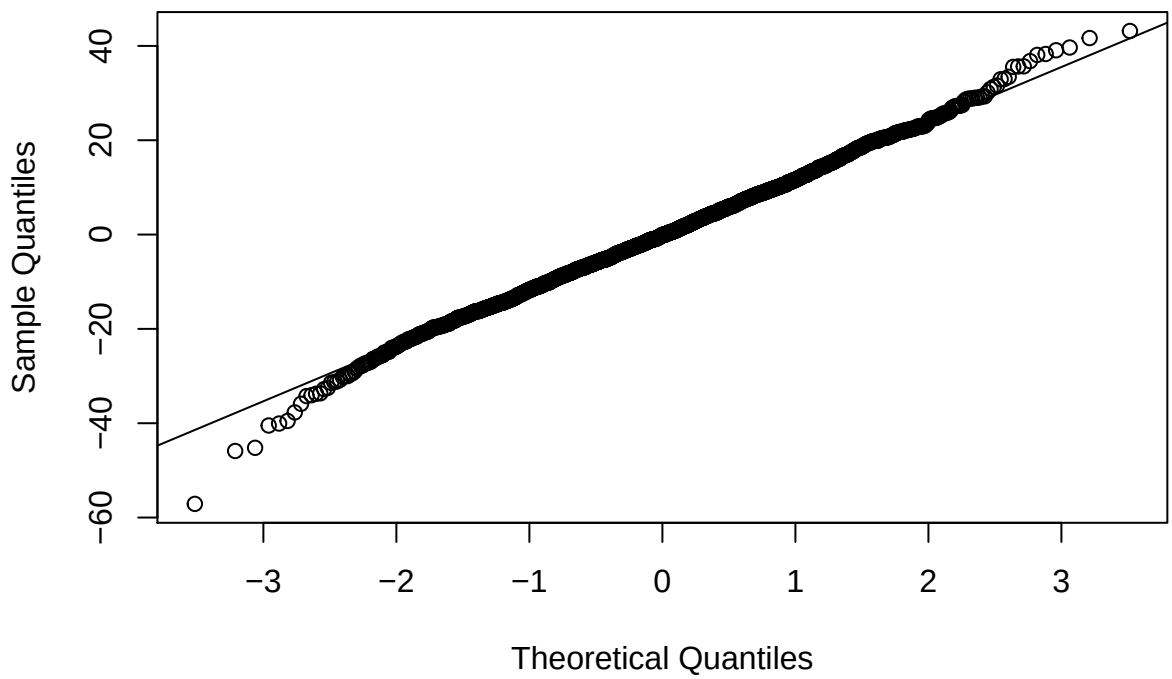

Normal Q−Q Plot
Sample Quantiles
−20
−10
0
10
20
−3
−2
−1
0
1
2
3
Theoretical Quantiles

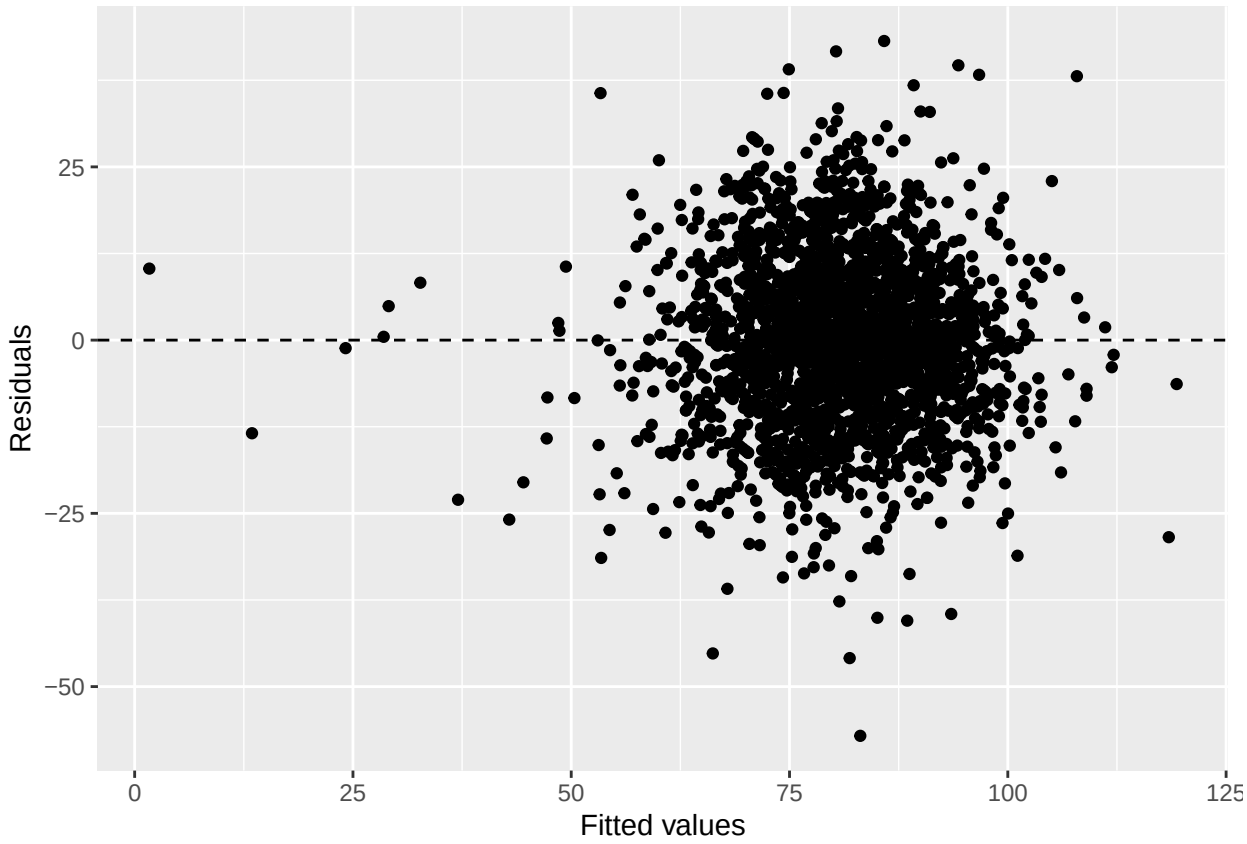



```
## [1] "Mean Squared Error: 79.0711797402695"
## [1] "Root MSE: 8.8921976890007"
```

Model 3

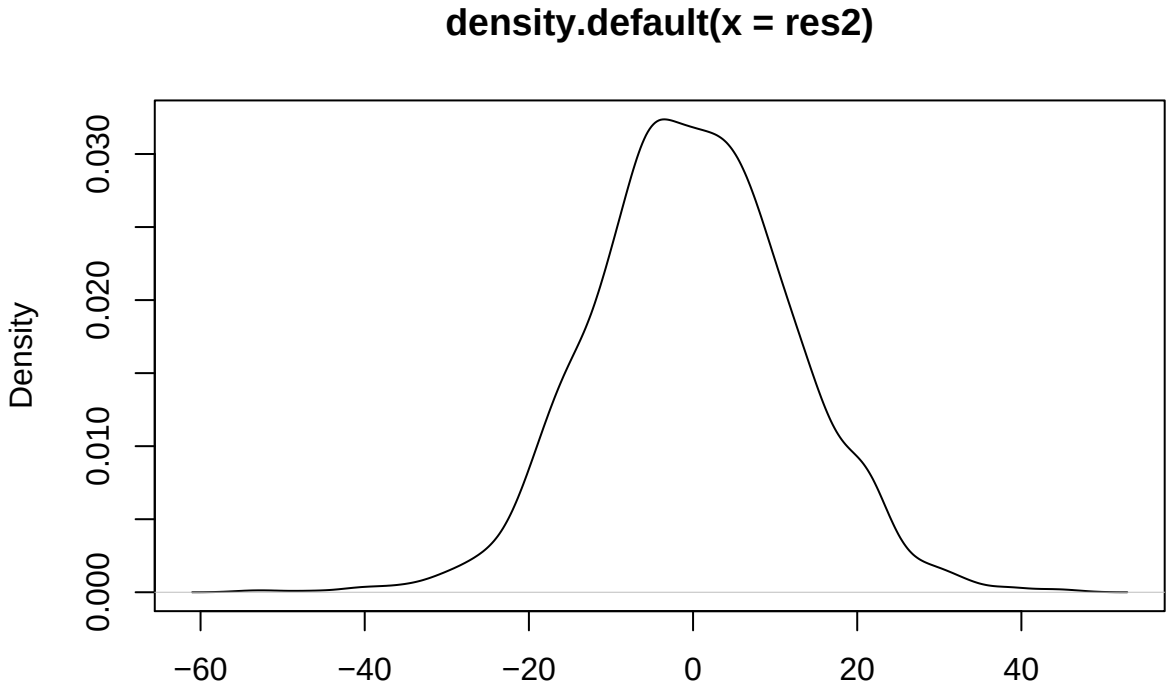

density.default(x = res)
Density
0.000
0.010
0.020
0.030
−60
−40
−20
0
20
40
N = 2276 Bandwidth = 2.28

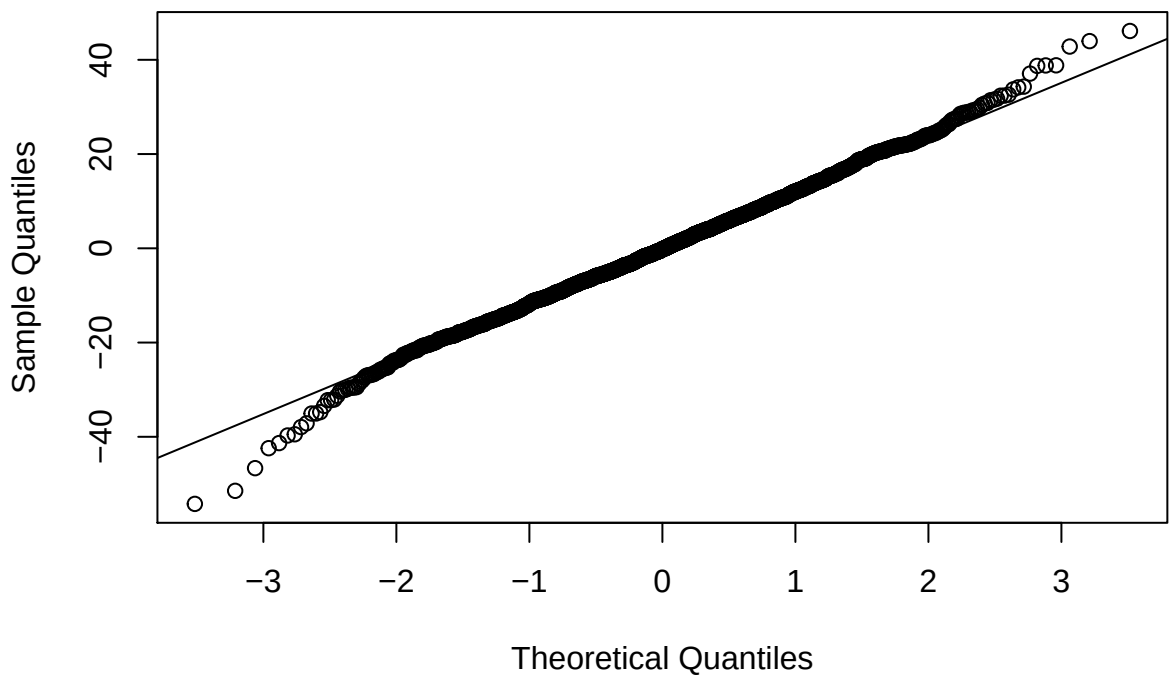

Normal Q−Q Plot
Sample Quantiles
−60
−40
−20
0
20
40
−3
−2
−1
0
1
2
3
Theoretical Quantiles

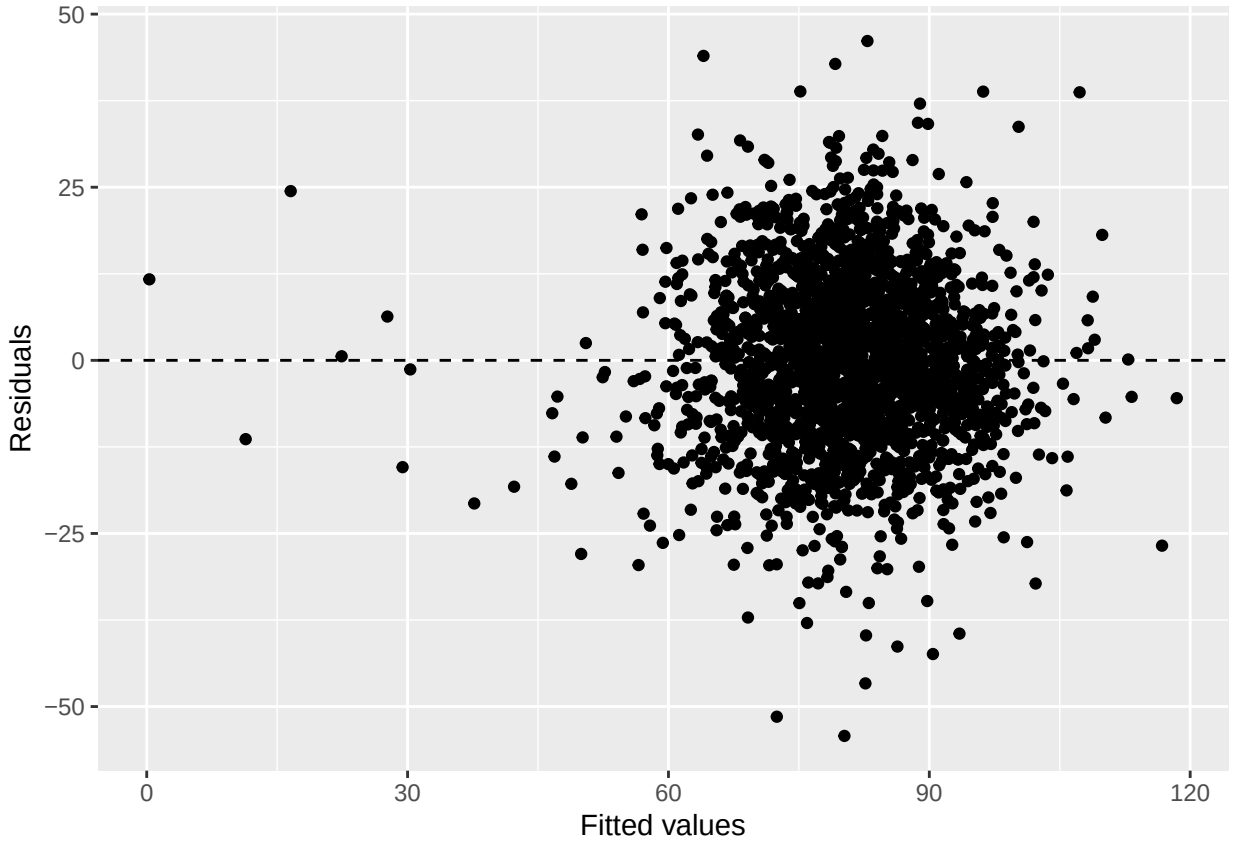



```
## [1] "Mean Squared Error: 146.432480093758"
## [1] "Root MSE: 12.1009288938395"
```

Model 4

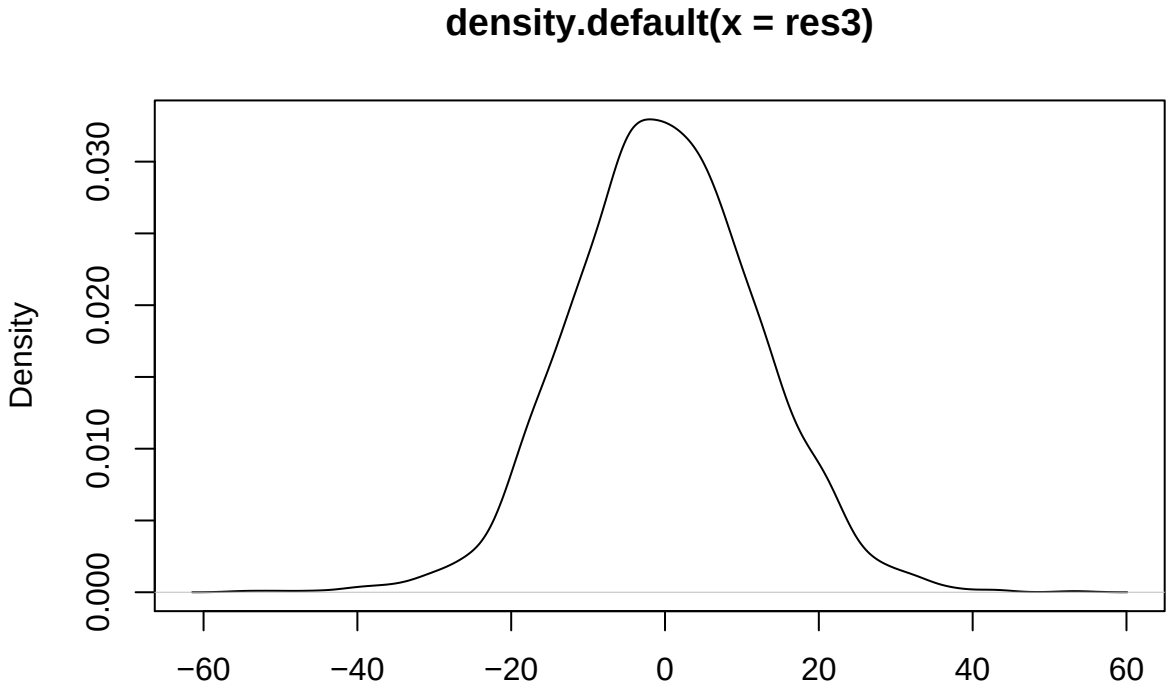

density.default(x = res2)
Density
0.000
0.010
0.020
0.030
−60
−40
−20
0
20
40
N = 2276 Bandwidth = 2.262

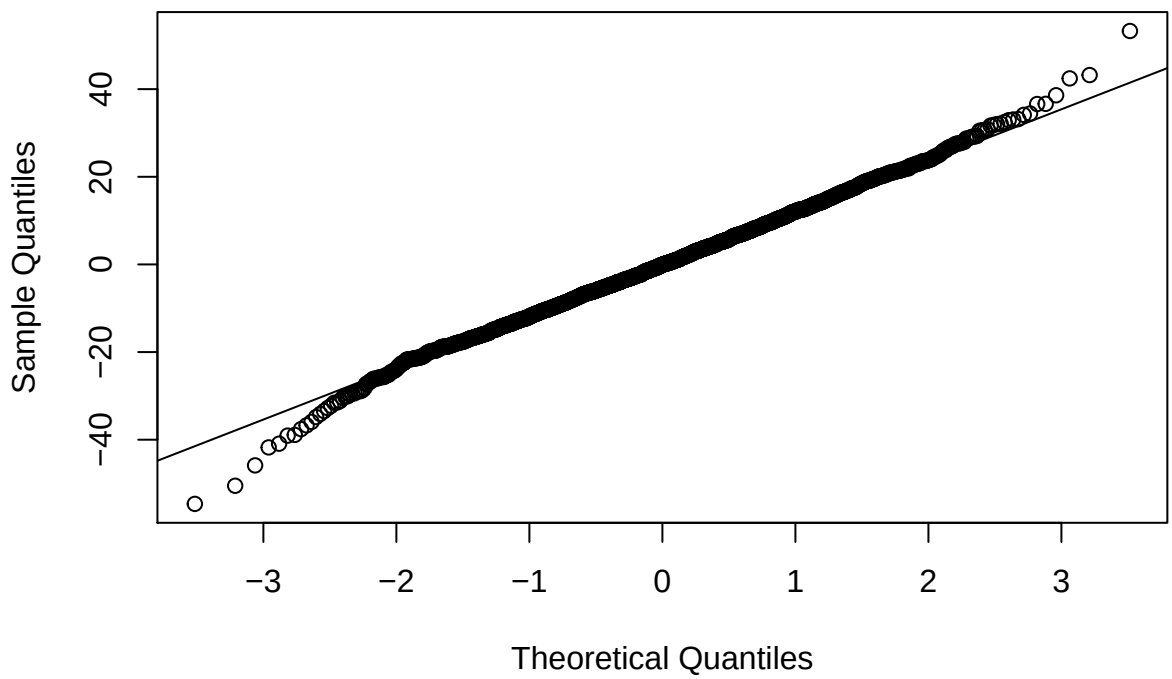

Normal Q−Q Plot
Sample Quantiles
−40
−20
0
20
40
−3
−2
−1
0
1
2
3
Theoretical Quantiles

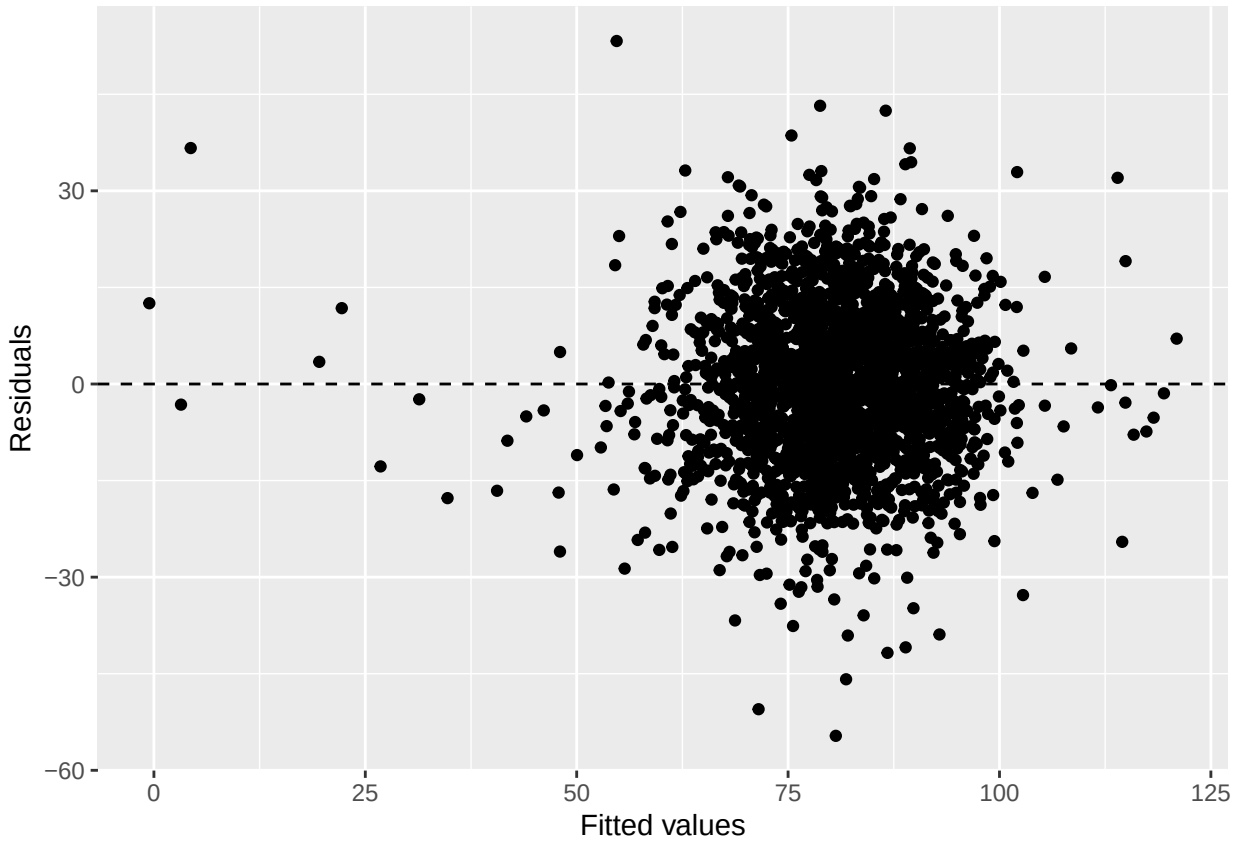



```
## [1] "Mean Squared Error: 149.617687432938"
## [1] "Root MSE: 12.2318309109036"
```

Model 5


density.default(x = res3)
Density
0.000
0.010
0.020
0.030
−60
−40
−20
0
20
40
60
N = 2276 Bandwidth = 2.278


Normal Q−Q Plot
Sample Quantiles
−40
−20
0
20
40
−3
−2
−1
0
1
2
3
Theoretical Quantiles




```
## [1] "Mean Squared Error: 147.636505909219"
```

```
## [1] "Root MSE: 12.1505763611945"
```

The diagnostic plots illustrate that our residuals for all 5 models are normally distributed. However, in terms of the residual abline plot, Model 2's residual plot seems the best compared to the other 4, which seem to have a pattern to them.In addition, Model 2 also had the lowest MSE/RMSE and highest adjusted R^2 (0.44)

Therefore, *Model 2* is the best model thus far. We will further test out this model on the evaluation dataset to see if this hold.

FURTHER EVALUATION

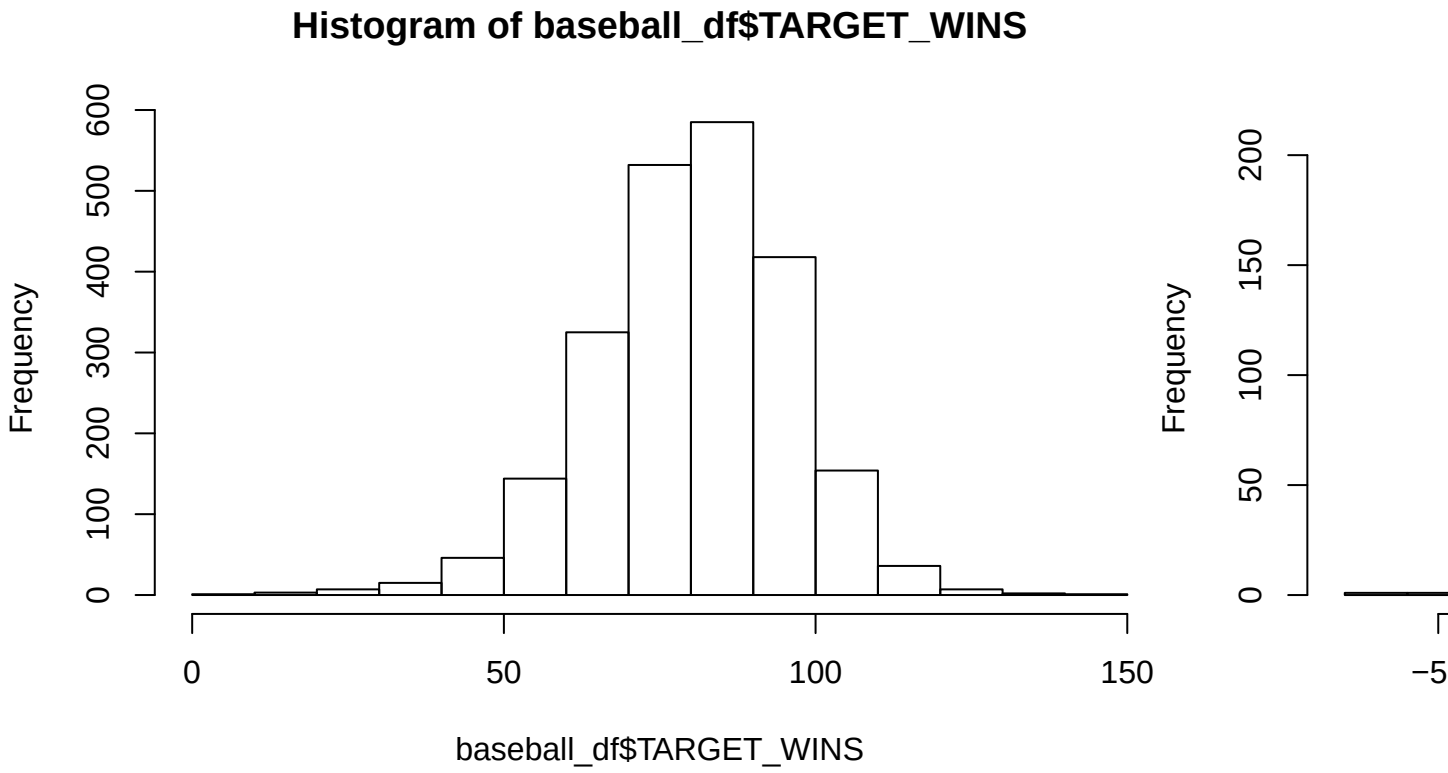
To ensure the model's efficacy when applied to the evaluation data set, we apply the same set of transformations used on the Training data set.Since the actual wins are withheld, we compared the distribution of predictions to the actual wins in the training set. The means were similar but the training data included much more variation between teams. It's also worth mentioning as well that using the predict function creates missing values as the evaluation data is missing. In fact, for TEAM_BATTING_HBP, over 90% of rows are missing entries.

```
##            INDEX   TEAM_BATTING_H  TEAM_BATTING_2B  TEAM_BATTING_3B
##             0.00             0.00             0.00             0.00
##  TEAM_BATTING_HR  TEAM_BATTING_BB  TEAM_BATTING_SO  TEAM_BASERUN_SB
##             0.00             0.00             6.95             5.02
##  TEAM_BASERUN_CS TEAM_BATTING_HBP  TEAM_PITCHING_H TEAM_PITCHING_HR
##            33.59            92.66             0.00             0.00
```

```
## TEAM_PITCHING_BB TEAM_PITCHING_SO  TEAM_FIELDING_E TEAM_FIELDING_DP
##             0.00             6.95             0.00            11.97
```

The prediction data also has missing values, which are approximately the same as the training data.

We will run our selected model on the evaluation dataset and look at the summary.




```
##       fit              lwr              upr
##  Min.   :-34.51   Min.   :-65.85   Min.   : -3.161
##  1st Qu.: 62.43   1st Qu.: 38.50   1st Qu.: 83.060
##  Median : 71.55   Median : 50.38   Median : 91.999
##  Mean   : 69.32   Mean   : 44.78   Mean   : 93.867
##  3rd Qu.: 78.39   3rd Qu.: 56.34   3rd Qu.: 99.615
##  Max.   :127.32   Max.   : 71.53   Max.   :206.949
```

```
## [1] 27.97456
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   71.00   82.00   80.79   92.00  146.00
```

```
## [1] 15.75215
```

Here are the predicted values for Target Wins, based on our model for the teams in the evaluation dataset.

Table 3: Predicted Target Wins

INDEX	EXAMPLE_TEAM_BATTING_H	EXAMPLE_TEAM_BATTING_2B	EXAMPLE_TEAM_BATTING_3B	EXAMPLE_TEAM_BATTING_HR	EXAMPLE_TEAM_BATTING_BB	EXAMPLE_TEAM_BATTING_SO	EXAMPLE_TEAM_BASERUN_SB	EXAMPLE_TEAM_BASERUN_CS	EXAMPLE_TEAM_FIELDING_E	EXAMPLE_TEAM_FIELDING_DP	EXAMPLE_TEAM_PITCHING_H	EXAMPLE_TEAM_PITCHING_HR	EXAMPLE_TEAM_PITCHING_BB	EXAMPLE_TEAM_PITCHING_SO	EXAMPLE_TEAM_WINS_PRED
9	1209	170	33	83	447	1080	62	50	NA	1209	83	447	1080	140	156	56
10	1221	151	29	88	516	929	54	39	NA	1221	88	516	929	135	164	61
14	1395	183	29	93	509	816	59	47	NA	1395	93	509	816	156	153	62
47	1539	309	29	159	486	914	148	57	42	1539	159	486	914	124	154	75
60	1445	203	68	5	95	416	NA	NA	NA	3902	14	257	1123	616	130	30

INDEX	EXAM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
63	1431	236	53	10	215	377	NA	NA	NA	2793	20	420	736	572	105	34
74	1430	219	55	37	568	527	365	NA	NA	1544	40	613	569	490	NA	48
83	1385	158	42	33	356	609	185	NA	NA	1626	39	418	715	328	104	41
98	1259	177	78	23	466	689	150	NA	NA	1342	25	497	734	226	132	67
120	1397	212	42	58	452	584	52	NA	NA	1489	62	482	622	184	145	58
123	1427	243	40	50	495	640	64	NA	NA	1501	53	521	673	200	183	58
135	1496	239	55	164	462	670	48	28	NA	1574	173	486	705	150	178	68
138	1420	223	57	186	511	751	31	21	NA	1494	196	538	790	137	167	68
140	1460	232	22	176	503	680	27	8	NA	1536	185	529	715	125	160	77
151	1411	195	22	141	485	665	59	48	NA	1411	141	485	665	115	114	71
153	1434	192	30	153	434	747	57	46	NA	1434	153	434	747	146	180	65
171	1297	204	22	130	491	1008	84	55	NA	1313	132	497	1021	154	126	61
184	1446	284	25	166	565	1041	77	39	55	1464	168	572	1054	115	172	78
193	1276	162	52	17	383	NA	138	NA	NA	1351	18	406	NA	301	83	44
213	1715	322	72	116	527	397	90	83	NA	1816	123	558	420	232	174	75
217	1520	295	68	49	628	459	77	49	NA	1620	52	669	489	166	158	84
226	1597	291	38	98	629	563	54	43	NA	1702	104	670	600	155	174	76
230	1453	256	67	105	653	651	40	41	NA	1559	113	701	698	179	153	75
241	1378	225	26	118	533	677	18	36	NA	1450	124	561	712	160	174	64
291	1516	277	24	152	431	902	89	36	54	1516	152	431	902	105	164	75
294	1556	288	20	164	474	878	121	32	73	1556	164	474	878	102	156	80
300	1499	183	28	3	83	0	NA	NA	NA	5167	10	286	0	1224	NA	-8
348	1464	263	58	47	385	479	63	66	NA	1540	49	405	504	232	146	57
350	1558	318	66	32	634	439	83	64	NA	1639	34	667	462	218	130	79
357	1502	308	36	39	432	602	45	46	NA	1601	42	460	642	199	135	55
367	1596	320	58	130	718	596	70	54	NA	1679	137	755	627	178	146	80
368	1546	260	59	110	630	541	72	65	NA	1648	117	671	577	167	166	78
372	1516	282	53	115	723	695	47	38	NA	1595	121	761	731	146	174	82
382	1550	275	47	146	765	723	29	20	NA	1631	154	805	761	178	177	79
388	1447	260	54	148	532	935	39	33	NA	1465	150	539	947	130	154	72
396	1450	252	28	203	594	855	50	48	NA	1450	203	594	855	156	131	75
398	1347	239	36	130	546	897	69	31	NA	1408	136	571	938	136	147	69
403	1561	260	56	214	531	911	66	47	NA	1571	215	534	917	133	163	75
407	1578	252	26	135	567	780	48	47	NA	2367	203	851	1170	137	162	86
410	1598	259	45	181	500	842	38	25	NA	1598	181	500	842	143	128	75
412	1497	322	21	145	599	711	41	34	NA	1506	146	603	715	130	147	76
414	1569	310	39	124	623	728	65	36	NA	1569	124	623	728	93	123	87
436	1119	118	33	7	37	0	NA	NA	NA	4120	26	136	0	1568	NA	-35
440	1609	196	120	62	781	599	536	NA	NA	1931	74	937	719	470	NA	92
476	1514	175	70	80	615	612	392	NA	NA	1803	95	733	729	413	NA	62
479	1657	237	119	41	593	334	325	NA	NA	2114	52	756	426	537	NA	92
481	1746	213	106	69	526	429	324	NA	NA	2176	86	655	535	500	NA	82
501	1319	224	70	56	416	677	176	131	NA	1397	59	440	717	284	100	53
503	1293	204	70	18	437	630	134	NA	NA	1360	19	460	663	281	127	58
506	1420	235	70	36	450	443	121	136	NA	1494	38	473	466	237	118	65
519	1496	269	54	76	412	500	55	NA	NA	1574	80	433	526	177	171	64
522	1625	289	38	80	517	486	72	NA	NA	1709	84	544	511	154	164	71
550	1391	239	50	145	499	1041	70	49	NA	1391	145	499	1041	162	147	63
554	1319	203	43	130	415	854	41	30	NA	1319	130	415	854	119	149	63
566	1411	251	35	107	471	912	93	64	NA	1411	107	471	912	174	149	59
578	1420	221	41	104	417	816	77	51	NA	1420	104	417	816	114	142	67
596	1552	206	106	38	566	401	334	NA	NA	1849	45	674	478	411	119	84

INDEX	EXAM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
599	1280	203	72	15	392	616	227	NA	NA	1346	16	412	648	250	100	59
605	1120	122	61	7	427	NA	194	NA	NA	1186	7	452	NA	332	106	46
607	1390	183	84	18	445	NA	216	NA	NA	1462	19	468	NA	304	107	67
614	1554	252	81	29	494	414	174	NA	NA	1798	34	572	479	200	134	85
644	1410	218	69	45	738	627	65	58	NA	1483	47	776	660	142	189	88
692	1507	262	28	159	573	907	107	52	NA	1516	160	577	913	126	132	78
699	1481	284	19	242	499	1030	78	51	63	1481	242	499	1030	100	167	89
700	1450	253	23	200	435	1002	137	67	79	1450	200	435	1002	94	166	80
716	1637	260	93	26	487	288	446	NA	NA	2088	33	621	367	321	NA	87
721	1436	202	82	44	376	681	160	NA	NA	1674	51	438	794	414	119	55
722	1600	218	89	21	344	538	152	NA	NA	1851	24	398	623	373	137	69
729	1348	168	76	23	506	NA	296	NA	NA	1427	24	536	NA	327	127	62
731	1460	191	111	22	612	629	306	NA	NA	1546	23	648	666	314	114	91
746	1621	255	126	37	478	350	54	NA	NA	1705	39	503	368	193	168	109
763	1433	241	49	45	468	501	52	NA	NA	1507	47	492	527	127	203	72
774	1440	232	48	155	586	679	49	32	NA	1515	163	616	714	144	204	73
776	1479	211	34	232	555	799	47	23	NA	1556	244	584	841	119	155	83
788	1573	281	36	106	379	938	59	55	NA	1573	106	379	938	144	144	63
789	1558	224	42	171	474	1042	79	56	NA	1558	171	474	1042	168	158	68
792	1385	225	46	130	637	961	147	66	NA	1457	137	670	1011	116	150	78
811	1419	250	27	164	488	1006	124	56	NA	1419	164	488	1006	125	131	71
835	1284	198	61	19	383	NA	186	NA	NA	1351	20	403	NA	270	100	51
837	1403	200	68	10	390	NA	201	NA	NA	1495	11	416	NA	262	119	61
861	1631	358	48	105	553	455	55	34	NA	1716	110	582	479	179	173	73
862	1666	343	82	98	487	600	67	57	NA	1764	104	516	635	184	156	81
863	1804	376	86	129	541	494	69	56	NA	1898	136	569	520	191	162	87
871	1534	284	53	74	539	624	50	44	NA	1614	78	567	656	173	202	72
879	1472	222	52	156	659	788	48	41	NA	1548	164	693	829	163	148	74
887	1489	229	21	134	467	603	61	26	NA	1566	141	491	634	133	174	70
892	1367	198	21	156	506	857	109	46	NA	1367	156	506	857	114	127	73
904	1485	222	46	101	534	692	88	88	NA	1494	102	537	696	131	146	74
909	1458	225	32	109	651	625	151	68	NA	1458	109	651	625	123	129	77
925	1530	334	30	198	630	1061	143	60	NA	1530	198	630	1061	110	146	87
940	1421	160	72	30	523	508	289	NA	NA	1731	37	637	619	445	NA	59
951	1869	301	122	58	347	127	399	NA	NA	10814	336	2008	735	1261	NA	105
976	1400	169	66	26	431	344	156	NA	NA	1680	31	517	413	398	133	53
981	1494	193	81	12	340	NA	207	NA	NA	1614	13	367	NA	285	85	66
983	1449	223	62	20	423	NA	298	NA	NA	1544	21	451	NA	286	93	58
984	1385	200	76	29	483	NA	262	NA	NA	1457	31	508	NA	296	83	64
989	1443	218	99	24	716	554	254	154	NA	1518	25	753	583	271	113	91
995	1825	284	106	61	616	398	101	94	NA	1932	65	652	421	245	113	104
1000	1627	296	95	38	630	445	93	76	NA	1712	40	663	468	207	159	97
1001	1623	299	106	54	622	445	149	77	NA	1718	57	659	471	221	183	98
1007	1556	298	82	60	500	550	72	53	NA	1637	63	526	579	187	176	81
1016	1381	228	39	80	535	501	41	42	NA	1453	84	563	527	203	149	59
1027	1556	272	46	114	532	634	32	37	NA	1637	120	560	667	138	157	76
1033	1416	206	32	168	610	775	36	18	NA	1490	177	642	815	130	138	77
1070	1413	257	21	204	546	1268	87	50	NA	1413	204	546	1268	135	157	77
1081	1504	253	102	33	262	482	NA	NA	NA	2901	64	505	930	652	154	62
1084	1193	165	68	45	299	1011	NA	NA	NA	1726	65	432	1462	743	NA	26
1098	1461	325	30	166	470	1145	89	40	67	1461	166	470	1145	103	174	76
1150	1458	294	36	187	590	999	89	30	61	1458	187	590	999	101	136	83

INDEX	EXAM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1160	1295	237	64	25	360	814	129	NA	NA	1734	33	482	1090	609	NA	37
1169	1431	263	58	118	591	675	155	75	NA	1431	118	591	675	155	151	71
1172	1469	305	59	98	498	644	216	84	NA	1469	98	498	644	150	153	70
1174	1633	266	59	115	508	709	185	43	NA	1633	115	508	709	141	150	78
1176	1603	295	58	132	442	758	133	48	NA	1603	132	442	758	127	140	75
1178	1487	269	52	117	400	832	106	64	NA	1487	117	400	832	129	157	67
1184	1474	318	44	101	501	884	108	62	NA	1483	102	504	889	123	162	72
1193	1594	296	52	152	538	938	128	39	NA	1604	153	541	944	126	190	79
1196	1415	285	42	140	524	921	140	65	52	1415	140	524	921	130	153	70
1199	1445	289	34	126	424	1008	53	33	63	1445	126	424	1008	125	163	66
1207	1362	199	81	29	408	508	386	NA	NA	1576	34	472	588	581	NA	47
1218	1572	195	106	30	522	288	297	NA	NA	1721	33	571	315	344	NA	87
1223	1209	168	56	16	435	NA	217	NA	NA	1280	17	461	NA	363	92	43
1226	1242	155	69	20	368	NA	132	NA	NA	1359	22	403	NA	287	103	52
1227	1098	116	63	29	340	NA	119	NA	NA	1155	31	358	NA	254	69	44
1229	1235	175	77	26	457	743	159	NA	NA	1299	27	481	782	246	131	62
1241	1651	247	80	59	357	335	83	63	NA	1737	62	376	352	219	146	72
1244	1712	265	85	68	463	406	39	32	NA	1813	72	490	430	221	138	83
1246	1391	206	78	41	390	523	112	NA	NA	1473	43	413	554	239	124	63
1248	1625	299	73	105	534	481	85	NA	NA	1721	111	565	509	203	120	76
1249	1740	319	77	128	506	569	56	NA	NA	1830	135	532	599	178	176	82
1253	1626	303	55	84	584	592	59	NA	NA	1733	90	622	631	192	150	75
1261	1471	277	36	65	602	509	85	NA	NA	1547	68	633	535	145	158	71
1305	1373	232	14	130	478	966	155	67	NA	1373	130	478	966	179	118	58
1314	1466	215	35	158	527	1151	143	51	NA	1649	178	593	1295	146	135	74
1323	1450	226	30	203	536	1092	102	41	62	1450	203	536	1092	73	145	90
1328	1474	223	57	18	259	391	NA	NA	NA	2985	36	524	792	780	75	36
1353	1335	228	49	120	500	909	106	75	NA	1335	120	500	909	127	168	67
1363	1455	233	36	97	435	677	52	57	NA	1464	98	438	681	137	157	64
1371	1477	272	35	82	511	779	256	115	NA	1477	82	511	779	89	146	78
1372	1426	240	25	125	555	932	138	93	NA	1426	125	555	932	131	148	71
1389	1255	183	61	11	304	814	161	NA	NA	1346	12	326	873	336	104	40
1393	1264	141	79	9	392	NA	181	NA	NA	1347	10	418	NA	294	95	59
1421	1695	310	89	66	610	421	110	44	NA	1795	70	646	446	193	173	94
1431	1460	274	66	63	538	674	54	53	NA	1536	66	566	709	222	170	68
1437	1349	237	46	53	610	639	50	39	NA	1419	56	642	672	137	160	73
1442	1340	226	40	117	554	771	14	40	NA	1410	123	583	811	135	167	69
1450	1396	257	42	150	554	969	92	33	NA	1396	150	554	969	172	158	64
1463	1472	259	47	82	604	684	99	56	NA	1472	82	604	684	146	171	74
1464	1544	256	46	112	526	693	66	45	NA	1544	112	526	693	134	203	75
1470	1453	282	41	141	502	779	68	44	NA	1453	141	502	779	120	139	73
1471	1446	257	39	196	501	977	81	61	NA	1446	196	501	977	118	168	74
1484	1468	289	30	106	506	990	119	61	NA	1486	107	512	1002	93	146	77
1495	1546	44	29	0	15	44	0	0	NA	22768	0	221	648	1473	NA	-18
1507	1372	195	31	103	353	932	36	31	NA	1372	103	353	932	166	154	51
1514	1365	203	29	98	547	958	89	43	NA	1365	98	547	958	112	135	71
1526	1314	172	26	112	436	1031	141	64	NA	1314	112	436	1031	151	171	57
1549	1469	323	41	200	547	1071	146	35	62	1469	200	547	1071	104	131	80
1552	1382	185	86	32	326	642	NA	NA	NA	2073	48	489	963	680	NA	47
1556	1642	218	135	29	449	459	252	NA	NA	2000	35	547	559	488	93	98
1564	1324	153	65	17	437	NA	201	NA	NA	1420	18	469	NA	352	101	52
1585	1770	313	116	160	677	599	96	63	NA	1862	168	712	630	219	139	86

INDEX	EXAM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1586	1765	293	83	164	792	587	146	72	NA	1869	174	839	622	177	139	92
1590	1590	277	76	113	657	510	74	50	NA	1729	123	714	554	164	124	85
1591	1775	334	88	193	741	629	82	42	NA	1879	204	785	666	173	157	88
1592	1635	297	77	183	746	639	63	38	NA	1720	193	785	672	178	141	80
1603	1557	264	79	146	655	503	25	25	NA	1638	154	689	529	126	169	85
1612	1485	210	57	153	591	746	52	40	NA	1562	161	622	785	129	193	78
1634	1461	229	41	152	515	597	66	47	NA	1479	154	521	604	124	185	74
1645	1322	208	19	147	427	1027	119	45	NA	1322	147	427	1027	126	164	64
1647	1462	281	18	163	536	903	78	37	NA	1462	163	536	903	112	165	78
1673	1537	217	115	23	517	NA	275	NA	NA	1638	25	551	NA	280	123	96
1674	1495	236	85	35	565	579	234	NA	NA	1583	37	598	613	224	114	83
1687	1468	280	70	66	565	488	60	50	NA	1585	71	610	527	187	141	76
1688	1689	296	74	59	580	343	103	66	NA	1777	62	610	361	204	130	85
1700	1533	301	59	104	536	567	64	36	NA	1634	111	571	604	224	140	66
1708	1379	229	55	64	636	592	39	35	NA	1451	67	669	623	150	169	75
1713	1373	223	37	94	718	590	55	45	NA	1444	99	755	621	147	156	74
1717	1394	215	43	118	505	765	42	32	NA	1466	124	531	805	175	197	63
1721	1371	223	36	116	540	783	17	12	NA	1442	122	568	824	134	157	69
1730	1400	210	28	148	617	953	100	39	NA	1400	148	617	953	137	162	73
1737	1327	209	33	114	596	823	343	124	NA	1335	115	600	828	145	131	67
1748	1432	263	33	199	593	1056	140	63	NA	1432	199	593	1056	142	122	75
1749	1474	251	22	156	580	926	129	54	NA	1474	156	580	926	105	151	81
1763	1450	279	28	205	609	1008	46	20	68	1450	205	609	1008	102	144	86
1768	2025	292	140	32	259	70	259	NA	NA	10935	173	1399	378	1172	NA	127
1778	1669	281	102	35	391	473	580	NA	NA	2033	43	476	576	643	NA	69
1780	1631	291	79	52	650	604	307	NA	NA	1987	63	792	736	566	NA	68
1782	1420	299	79	5	233	587	NA	NA	NA	2347	8	385	970	1056	NA	33
1784	1312	230	52	29	324	591	NA	NA	NA	1932	43	477	870	658	NA	30
1794	2058	336	90	75	573	324	341	NA	NA	2545	93	709	401	456	NA	93
1803	1351	181	58	25	402	NA	169	NA	NA	1440	27	428	NA	427	99	42
1804	1452	199	87	17	433	NA	192	NA	NA	1548	18	461	NA	293	106	72
1819	1466	242	57	68	300	562	106	88	NA	1552	72	318	595	246	143	48
1832	1534	256	44	64	406	511	59	NA	NA	1635	68	433	545	195	166	59
1833	1609	311	38	61	433	581	57	NA	NA	1749	66	471	632	214	152	58
1844	1344	207	28	59	472	527	57	NA	NA	1414	62	497	554	246	158	46
1847	1438	239	41	96	463	629	72	NA	NA	1513	101	487	662	221	133	56
1854	1368	225	53	139	686	708	46	34	NA	1439	146	722	745	116	123	79
1855	1381	218	52	127	615	708	47	24	NA	1453	134	647	745	151	147	71
1857	1498	250	59	130	603	916	54	35	NA	1576	137	634	964	136	143	79
1864	1389	206	53	145	497	1098	46	32	NA	1398	146	500	1105	158	154	64
1865	1448	224	49	117	510	969	56	42	NA	1448	117	510	969	113	147	75
1869	1307	225	58	102	522	1073	72	64	NA	1315	103	525	1080	113	135	72
1880	1517	250	38	104	563	654	156	70	NA	2297	157	852	990	130	136	85
1881	1417	245	25	112	506	831	128	76	NA	1417	112	506	831	121	138	69
1882	1352	209	45	125	640	906	143	75	NA	1352	125	640	906	152	117	70
1894	1458	296	34	106	559	995	81	28	NA	1640	119	629	1119	108	156	79
1896	1390	290	35	116	519	1032	92	56	NA	1390	116	519	1032	105	134	73
1916	1475	257	80	52	515	573	284	NA	NA	1810	64	632	703	471	NA	60
1918	1378	178	85	35	512	604	246	NA	NA	1654	42	614	725	570	NA	56
1921	1817	277	155	60	541	259	319	NA	NA	2264	75	674	323	441	NA	117
1926	1711	213	133	29	418	375	195	NA	NA	1860	32	454	408	392	NA	102
1938	1415	217	112	52	552	613	168	NA	NA	1489	55	581	645	243	138	84

INDEX	TEAM_BATTING_H	TEAM_BATTING_2B	TEAM_BATTING_3B	TEAM_BATTING_HR	TEAM_BATTING_BB	TEAM_BATTING_SO	TEAM_BASERUN_SB	TEAM_BASERUN_CS	TEAM_BATTING_HBP	TEAM_PITCHING_H	TEAM_PITCHING_HR	TEAM_PITCHING_BB	TEAM_PITCHING_SO	TEAM_FIELDING_E	TEAM_FIELDING_DP	[illegible]
1979	1263	190	32	97	511	762	45	43	NA	1329	102	538	802	190	176	55
1982	1328	221	63	96	495	686	23	23	NA	1397	101	521	722	175	184	63
1987	1571	248	59	126	511	786	36	25	NA	1653	133	538	827	135	171	78
1997	1522	235	70	130	444	871	66	34	NA	1522	130	444	871	137	195	72
2004	1550	278	57	133	474	878	260	120	NA	1550	133	474	878	145	137	72
2011	1412	237	33	98	438	841	96	62	NA	1412	98	438	841	128	142	64
2015	1344	243	46	111	560	959	120	61	NA	1361	112	567	971	126	130	71
2022	1441	276	30	141	513	1094	95	62	NA	1621	159	577	1231	136	155	72
2025	1395	271	35	107	393	1060	159	51	NA	1395	107	393	1060	140	161	59
2027	1506	320	31	168	564	1032	86	40	66	1506	168	564	1032	132	169	76
2031	1437	269	39	143	418	1073	63	40	96	1446	144	421	1080	104	190	71
2036	2170	241	70	13	111	102	92	76	NA	6893	41	353	324	1217	NA	45
2066	1324	194	53	94	537	775	101	58	NA	1332	95	540	780	141	155	68
2073	1442	239	25	136	484	917	96	68	NA	1442	136	484	917	135	135	68
2087	1413	279	37	157	602	1177	131	53	46	1413	157	602	1177	141	155	72
2092	1416	269	39	130	600	977	99	44	49	1416	130	600	977	109	136	78
2125	1523	216	97	33	360	712	NA	NA	NA	2203	48	521	1030	743	NA	58
2148	1294	169	51	24	546	NA	217	NA	NA	1370	25	578	NA	244	79	58
2162	1668	251	98	79	497	413	145	121	NA	1766	84	526	437	198	164	89
2191	1422	215	53	140	660	662	44	NA	NA	1496	147	694	696	144	190	76
2203	1524	231	31	200	513	807	72	49	NA	1496	196	504	792	139	150	76
2218	1392	227	41	134	568	842	90	59	NA	1392	134	568	842	178	136	64
2221	1318	200	44	80	512	845	101	58	NA	1326	80	515	850	157	125	62
2225	1499	229	26	112	528	980	126	76	NA	1499	112	528	980	169	134	64
2232	1345	215	48	141	471	973	95	57	NA	1345	141	471	973	108	151	70
2267	1620	210	139	66	542	355	233	NA	NA	1988	81	665	436	523	NA	88
2291	1339	185	80	34	413	579	149	NA	NA	1682	43	519	727	276	146	66
2299	1621	272	86	95	503	545	87	NA	NA	1705	100	529	573	208	148	78
2317	1585	288	62	105	572	498	39	NA	NA	1667	110	602	524	118	170	86
2318	1576	269	46	67	542	513	58	NA	NA	1658	70	570	540	143	158	75
2353	1541	300	49	101	451	781	117	54	NA	1541	101	451	781	122	174	73
2403	1149	175	18	59	529	974	133	77	NA	1209	62	556	1025	175	155	51
2411	1626	265	27	125	483	593	92	49	NA	1636	126	486	597	148	170	70
2415	1461	228	29	121	423	812	82	50	NA	1470	122	426	817	139	139	64
2424	1472	284	39	181	483	984	113	67	NA	1472	181	483	984	130	145	72
2441	1366	218	39	99	451	649	28	52	NA	1374	100	454	653	131	164	64
2464	1489	287	36	195	470	1094	156	55	74	1489	195	470	1094	97	184	80
2465	1457	305	38	187	522	1142	71	18	53	1457	187	522	1142	107	159	78
2472	1454	220	52	9	97	393	NA	NA	NA	3141	19	210	849	994	95	4
2481	1642	221	98	56	638	451	319	NA	NA	2031	69	789	558	492	NA	80
2487	819	72	72	18	198	1107	NA	NA	NA	7371	162	1782	9963	936	NA	48
2500	1251	162	23	95	492	860	71	69	NA	1299	99	511	893	139	146	60
2501	1345	190	23	125	695	777	77	68	NA	1345	125	695	777	163	156	69
2520	1381	263	37	102	463	976	196	63	NA	1381	102	463	976	124	113	66
2521	1410	270	36	122	542	860	228	56	NA	1410	122	542	860	159	144	65
2525	1423	339	34	172	420	1084	75	46	NA	1423	172	420	1084	131	150	66

CONCLUSION

Our final model predicts target wins for the team as follows:

Target Wins = 2.143e+00-2.392e+01(log(TEAM_FIELD_E))+2.508e+01(log(TEAM_PITCHING_BB))-4.306e-03(TEAM_BATTING_3B:TEAM_BATTING_HR)-3.449e-05(TEAM_BATTING_2B:TEAM_BATTING_HR)+104(TEAM_BATTING_H:TEAM_BATTING_HR)+4.332e-04(TEAM_BATTING_H:TEAM_BATTING_3B)

The model seems to suggest that in order to maximize a teams chances of winning they should focus on reducing fielding errors which makes sense. However, what is interesting from our model is the positive association between walks allowed and target wins. Moreover, the model also seems to suggest that some of the batting interaction effects may slightly lower your chances of winning. While this is definitely an interesting finding, this may just as well be because observations suggest that one may lower the chances or another and vice versa.

REFERENCES

An Introduction to Statistical Learning with Applications in R Springer Linear Models with R (2014), Julian J, Faraway

CODE APPENDIX

The code chunks below shows the R code called above throughout the analysis. They are being reproduced in the appendix for review and feedback.

```
# Libraries
library(dplyr)
library(GGally)
library(DataExplorer)
library(ggplot2)
library(readr)
library(reshape2)
library(purrr)
library(tidyr)
library(corrplot)
library(MASS)
library(caret)
library(Hmisc)
library(e1071)
library(Amelia)

set.seed(2012)
# read data
baseball_df <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW1/moneyball-training
baseball_eval <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW1/moneyball-evalua
```

```
plot_histogram(baseball_df$TARGET_WINS,
               title="Distribution of TARGET_WINS")
```

```
skimr::skim(baseball_df)
```

```
summary(baseball_df)
print('Observations per year, 1871 - 2006:')
round(nrow(baseball_df)/(2006-1871),2)
```

```
baseball_df %>%
  dplyr::select(-INDEX, -TARGET_WINS) %>%
  pivot_longer(everything(), names_to = 'Var', values_to='Value') %>%
```

```
  ggplot(aes(x = Var, y = Value)) +
  geom_boxplot() +
  coord_flip()
```

```
corrplot(cor(baseball_df[,2:17], use = 'complete.obs'))
```

```
baseball_df %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
  facet_wrap(~key,scales="free")+
  geom_histogram()
```

```
skewness(baseball_df$TEAM_PITCHING_SO,na.rm=TRUE)
skewness(baseball_df$TEAM_BASERUN_SB,na.rm=TRUE)
skewness(baseball_df$TEAM_FIELDING_E,na.rm=TRUE)
```

```
# Log Transformation
baseball_df2 <- baseball_df
baseball_df2$TEAM_PITCHING_SO <- log(baseball_df2$TEAM_PITCHING_SO)
baseball_df2$TEAM_BASERUN_SB <- log(baseball_df2$TEAM_BASERUN_SB)
baseball_df2$TEAM_FIELDING_E <- log(baseball_df2$TEAM_FIELDING_E)

#Certain values changed to -lnf afte transformation. This throws an error during imputation so we will
baseball_df2$TEAM_PITCHING_SO <- ifelse(baseball_df2$TEAM_PITCHING_SO=="-Inf",NA,baseball_df2$TEAM_PITC
baseball_df2$TEAM_BASERUN_SB <- ifelse(baseball_df2$TEAM_BASERUN_SB == "-Inf",NA,baseball_df2$TEAM_BASE
baseball_df2$TEAM_FIELDING_E <-ifelse(baseball_df2$TEAM_FIELDING_E=="-Inf",NA,baseball_df2$TEAM_FIELDIN
```

```
require(Amelia)
set.seed(123)
missmod<- amelia(baseball_df2)
```

```
baseball_df_fix <- baseball_df
```

```
m1 <- lm(TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +TEAM_BATTING_BB + TEAM_BATTING_SO  + TEAM_PITC
summary(m1)

#remove TEAM_PITCHING_BB & TEAM_PITCHING_SO
m2<- lm(TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +TEAM_BATTING_BB + TEAM_BATTING_SO  + TEAM_PITCH

summary(m2)
anova(m1, m2)
```

```
par(mfrow=c(2,1))
plot(baseball_df_fix$TEAM_PITCHING_H,baseball_df_fix$TARGET_WINS,xlab = 'TEAM_PITCHING',ylab = 'TARGET_
plot(log(baseball_df_fix$TEAM_PITCHING_H),baseball_df_fix$TARGET_WINS,xlab = 'LOG(TEAM_PITCHING)',ylab =


#log TEAM_PITCHING_H
m3 <- lm(TARGET_WINS ~ TEAM_BATTING_H + TEAM_BATTING_HR +TEAM_BATTING_BB + TEAM_BATTING_SO  + log(TEAM_
summary(m3)
```

```
#Remove TEAM_BATTING_H
m4 <- lm(TARGET_WINS ~ TEAM_BATTING_HR +TEAM_BATTING_BB + TEAM_BATTING_SO  + log(TEAM_PITCHING_H) + TEA
summary(m4)
anova(m3, m4)
```

```
baseball_lm2 <- lm (baseball_df_fix, formula = TARGET_WINS ~.-INDEX+log(TEAM_FIELDING_E) + log(TEAM_PIT
summary(baseball_lm2)
```

```
betas <- NULL
ses <- NULL
for(i in 1:missmod$m)
{
  lmod <-  lm (TARGET_WINS ~.-INDEX,  missmod$imputations[[i]])
  betas <- rbind(betas ,coef(lmod))
  ses <- rbind(ses,coef(summary(lmod))[,2])
}
summary(lmod)
```

```
betas <- NULL
ses <- NULL
for(i in 1:missmod$m)
{
  lmod2 <-  lm (TARGET_WINS ~.-INDEX-TEAM_PITCHING_HR-TEAM_PITCHING_BB-TEAM_PITCHING_H-TEAM_PITCHING_SO
  betas <- rbind(betas ,coef(lmod2))
  ses <- rbind(ses,coef(summary(lmod2))[,2])
}
summary(lmod2)
```

```
betas <- NULL
ses <- NULL
for(i in 1:missmod$m)
{
  lmod3 <-  lm (TARGET_WINS ~.-INDEX-TEAM_PITCHING_HR-TEAM_PITCHING_BB-TEAM_PITCHING_H-TEAM_PITCHING_SO
  betas <- rbind(betas ,coef(lmod3))
  ses <- rbind(ses,coef(summary(lmod3))[,2])
}
summary(lmod3)
```

```
res0 <- resid(m4)
plot(density(res0))
qqnorm(res0)
qqline(res0)
ggplot(data = m4, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```

```
RSS <- c(crossprod(m4$residuals))
MSE <- RSS/length(m4$residuals)
print(paste0("Mean Squared Error: ", MSE))
print(paste0("Root MSE: ", sqrt(MSE)))
```

```
resx <- resid(baseball_lm2)
plot(density(resx))
qqnorm(resx)
qqline(resx)
ggplot(data = baseball_lm2, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
```

```
  xlab("Fitted values") +
  ylab("Residuals")
```

```
RSS <- c(crossprod(baseball_lm2$residuals))
MSE <- RSS/length(baseball_lm2$residuals)
print(paste0("Mean Squared Error: ", MSE))
print(paste0("Root MSE: ", sqrt(MSE)))
```

```
res <- resid(lmod)
plot(density(res))
qqnorm(res)
qqline(res)
ggplot(data = lmod, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```

```
RSS <- c(crossprod(lmod$residuals))
MSE <- RSS/length(lmod$residuals)
print(paste0("Mean Squared Error: ", MSE))
print(paste0("Root MSE: ", sqrt(MSE)))
```

```
res2 <- resid(lmod2)
plot(density(res2))
qqnorm(res2)
qqline(res2)
ggplot(data = lmod2, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```

```
RSS <- c(crossprod(lmod2$residuals))
MSE <- RSS/length(lmod2$residuals)
print(paste0("Mean Squared Error: ", MSE))
print(paste0("Root MSE: ", sqrt(MSE)))
```

```
res3 <- resid(lmod3)
plot(density(res3))
qqnorm(res3)
qqline(res3)
ggplot(data = lmod3, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```

```
RSS <- c(crossprod(lmod3$residuals))
MSE <- RSS/length(lmod3$residuals)
print(paste0("Mean Squared Error: ", MSE))
print(paste0("Root MSE: ", sqrt(MSE)))
```

```
round(100*colSums(is.na(baseball_eval))/nrow(baseball_eval),2)
```

```
eval_predict <-  predict(baseball_lm2, newdata = baseball_eval, interval="prediction")
hist(baseball_df$TARGET_WINS)
hist(eval_predict)

summary(eval_predict)
sd(eval_predict)
summary(baseball_df$TARGET_WINS)
sd(baseball_df$TARGET_WINS)
```

```
pred.TW <- round(predict(baseball_lm2, baseball_eval))
baseball_eval$TARGET_WINS <- pred.TW
```

```
knitr::kable(baseball_eval,caption="Predicted Target Wins")
```