DATA 607 - Project 1

Vincent Bianco

2/22/2020

Introdution: The purpose of this project is to create a methodology for reading specific fields of data about Chess players from a strucutured list of chess tournament results. The end result will be file containing each players name, state, number of points, pre-rating, and the average pre-rating of the opponents they faced.

First, we start by importing the stringr library:

```
library(stringr)
```

Next, we read in the structured tournament results file:

```
results_file <- readLines("https://raw.githubusercontent.com/biancov/DATA607-Project1/master/tournamenti
```

```
## Warning in readLines("https://raw.githubusercontent.com/biancov/DATA607-
## Project1/master/tournamentinfo.txt"): incomplete final line found on 'https://
## raw.githubusercontent.com/biancov/DATA607-Project1/master/tournamentinfo.txt'
```

Showing the format of the results_file:

```
results_file
```

```
##   [1] "-----------------------------------------------------------------------------------------"
##   [2] " Pair | Player Name                     |Total|Round|Round|Round|Round|Round|Round|Round| "
##   [3] " Num  | USCF ID / Rtg (Pre->Post)       | Pts |  1  |  2  |  3  |  4  |  5  |  6  |  7  | "
##   [4] "-----------------------------------------------------------------------------------------"
##   [5] "    1 | GARY HUA                        |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|"
##   [6] "   ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
##   [7] "-----------------------------------------------------------------------------------------"
##   [8] "    2 | DAKSHESH DARURI                 |6.0  |W  63|W  58|L   4|W  17|W  16|W  20|W   7|"
##   [9] "   MI | 14598900 / R: 1553   ->1663     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
##  [10] "-----------------------------------------------------------------------------------------"
##  [11] "    3 | ADITYA BAJAJ                    |6.0  |L   8|W  61|W  25|W  21|W  11|W  13|W  12|"
##  [12] "   MI | 14959604 / R: 1384   ->1640     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
##  [13] "-----------------------------------------------------------------------------------------"
##  [14] "    4 | PATRICK H SCHILLING             |5.5  |W  23|D  28|W   2|W  26|D   5|W  19|D   1|"
##  [15] "   MI | 12616049 / R: 1716   ->1744     |N:2  |W    |B    |W    |B    |W    |B    |B    |"
##  [16] "-----------------------------------------------------------------------------------------"
##  [17] "    5 | HANSHI ZUO                      |5.5  |W  45|W  37|D  12|D  13|D   4|W  14|W  17|"
##  [18] "   MI | 14601533 / R: 1655   ->1690     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
##  [19] "-----------------------------------------------------------------------------------------"
##  [20] "    6 | HANSEN SONG                     |5.0  |W  34|D  29|L  11|W  35|D  10|W  27|W  21|"
```

```
##  [21] "   OH | 15055204 / R: 1686   ->1687     |N:3  |W    |B    |W    |B    |B    |W    |B    |"
##  [22] "-----------------------------------------------------------------------------------------"
##  [23] "    7 | GARY DEE SWATHELL               |5.0  |W  57|W  46|W  13|W  11|L   1|W   9|L   2|"
##  [24] "   MI | 11146376 / R: 1649   ->1673     |N:3  |W    |B    |W    |B    |B    |W    |W    |"
##  [25] "-----------------------------------------------------------------------------------------"
##  [26] "    8 | EZEKIEL HOUGHTON                |5.0  |W   3|W  32|L  14|L   9|W  47|W  28|W  19|"
##  [27] "   MI | 15142253 / R: 1641P17->1657P24  |N:3  |B    |W    |B    |W    |B    |W    |W    |"
##  [28] "-----------------------------------------------------------------------------------------"
##  [29] "    9 | STEFANO LEE                     |5.0  |W  25|L  18|W  59|W   8|W  26|L   7|W  20|"
##  [30] "   ON | 14954524 / R: 1411   ->1564     |N:2  |W    |B    |W    |B    |W    |B    |B    |"
##  [31] "-----------------------------------------------------------------------------------------"
##  [32] "   10 | ANVIT RAO                       |5.0  |D  16|L  19|W  55|W  31|D   6|W  25|W  18|"
##  [33] "   MI | 14150362 / R: 1365   ->1544     |N:3  |W    |W    |B    |B    |W    |B    |W    |"
##  [34] "-----------------------------------------------------------------------------------------"
##  [35] "   11 | CAMERON WILLIAM MC LEMAN        |4.5  |D  38|W  56|W   6|L   7|L   3|W  34|W  26|"
##  [36] "   MI | 12581589 / R: 1712   ->1696     |N:3  |B    |W    |B    |W    |B    |W    |B    |"
##  [37] "-----------------------------------------------------------------------------------------"
##  [38] "   12 | KENNETH J TACK                  |4.5  |W  42|W  33|D   5|W  38|H    |D   1|L   3|"
##  [39] "   MI | 12681257 / R: 1663   ->1670     |N:3  |W    |B    |W    |B    |     |W    |B    |"
##  [40] "-----------------------------------------------------------------------------------------"
##  [41] "   13 | TORRANCE HENRY JR               |4.5  |W  36|W  27|L   7|D   5|W  33|L   3|W  32|"
##  [42] "   MI | 15082995 / R: 1666   ->1662     |N:3  |B    |W    |B    |B    |W    |W    |B    |"
##  [43] "-----------------------------------------------------------------------------------------"
##  [44] "   14 | BRADLEY SHAW                    |4.5  |W  54|W  44|W   8|L   1|D  27|L   5|W  31|"
##  [45] "   MI | 10131499 / R: 1610   ->1618     |N:3  |W    |B    |W    |W    |B    |B    |W    |"
##  [46] "-----------------------------------------------------------------------------------------"
##  [47] "   15 | ZACHARY JAMES HOUGHTON          |4.5  |D  19|L  16|W  30|L  22|W  54|W  33|W  38|"
##  [48] "   MI | 15619130 / R: 1220P13->1416P20  |N:3  |B    |B    |W    |W    |B    |B    |W    |"
##  [49] "-----------------------------------------------------------------------------------------"
##  [50] "   16 | MIKE NIKITIN                    |4.0  |D  10|W  15|H    |W  39|L   2|W  36|U    |"
##  [51] "   MI | 10295068 / R: 1604   ->1613     |N:3  |B    |W    |     |B    |W    |B    |     |"
##  [52] "-----------------------------------------------------------------------------------------"
##  [53] "   17 | RONALD GRZEGORCZYK              |4.0  |W  48|W  41|L  26|L   2|W  23|W  22|L   5|"
##  [54] "   MI | 10297702 / R: 1629   ->1610     |N:3  |W    |B    |W    |B    |W    |B    |W    |"
##  [55] "-----------------------------------------------------------------------------------------"
##  [56] "   18 | DAVID SUNDEEN                   |4.0  |W  47|W   9|L   1|W  32|L  19|W  38|L  10|"
##  [57] "   MI | 11342094 / R: 1600   ->1600     |N:3  |B    |W    |B    |W    |B    |W    |B    |"
##  [58] "-----------------------------------------------------------------------------------------"
##  [59] "   19 | DIPANKAR ROY                    |4.0  |D  15|W  10|W  52|D  28|W  18|L   4|L   8|"
##  [60] "   MI | 14862333 / R: 1564   ->1570     |N:3  |W    |B    |W    |B    |W    |W    |B    |"
##  [61] "-----------------------------------------------------------------------------------------"
##  [62] "   20 | JASON ZHENG                     |4.0  |L  40|W  49|W  23|W  41|W  28|L   2|L   9|"
##  [63] "   MI | 14529060 / R: 1595   ->1569     |N:4  |W    |B    |W    |B    |W    |B    |W    |"
##  [64] "-----------------------------------------------------------------------------------------"
##  [65] "   21 | DINH DANG BUI                   |4.0  |W  43|L   1|W  47|L   3|W  40|W  39|L   6|"
##  [66] "   ON | 15495066 / R: 1563P22->1562     |N:3  |B    |W    |B    |W    |W    |B    |W    |"
##  [67] "-----------------------------------------------------------------------------------------"
##  [68] "   22 | EUGENE L MCCLURE                |4.0  |W  64|D  52|L  28|W  15|H    |L  17|W  40|"
##  [69] "   MI | 12405534 / R: 1555   ->1529     |N:4  |W    |B    |W    |B    |     |W    |B    |"
##  [70] "-----------------------------------------------------------------------------------------"
##  [71] "   23 | ALAN BUI                        |4.0  |L   4|W  43|L  20|W  58|L  17|W  37|W  46|"
##  [72] "   ON | 15030142 / R: 1363   ->1371     |     |B    |W    |B    |W    |B    |W    |B    |"
##  [73] "-----------------------------------------------------------------------------------------"
##  [74] "   24 | MICHAEL R ALDRICH               |4.0  |L  28|L  47|W  43|L  25|W  60|W  44|W  39|"
```

```
##  [75] "   MI | 13469010 / R: 1229   ->1300     |N:4  |B    |W    |B    |B    |W    |W    |B    |"
##  [76] "-----------------------------------------------------------------------------------------"
##  [77] "   25 | LOREN SCHWIEBERT                |3.5  |L   9|W  53|L   3|W  24|D  34|L  10|W  47|"
##  [78] "   MI | 12486656 / R: 1745   ->1681     |N:4  |B    |W    |B    |W    |B    |W    |B    |"
##  [79] "-----------------------------------------------------------------------------------------"
##  [80] "   26 | MAX ZHU                         |3.5  |W  49|W  40|W  17|L   4|L   9|D  32|L  11|"
##  [81] "   ON | 15131520 / R: 1579   ->1564     |N:4  |B    |W    |B    |W    |B    |W    |W    |"
##  [82] "-----------------------------------------------------------------------------------------"
##  [83] "   27 | GAURAV GIDWANI                  |3.5  |W  51|L  13|W  46|W  37|D  14|L   6|U    |"
##  [84] "   MI | 14476567 / R: 1552   ->1539     |N:4  |W    |B    |W    |B    |W    |B    |     |"
##  [85] "-----------------------------------------------------------------------------------------"
##  [86] "   28 | SOFIA ADINA STANESCU-BELLU       |3.5  |W  24|D   4|W  22|D  19|L  20|L   8|D  36|"
##  [87] "   MI | 14882954 / R: 1507   ->1513     |N:3  |W    |W    |B    |W    |B    |B    |W    |"
##  [88] "-----------------------------------------------------------------------------------------"
##  [89] "   29 | CHIEDOZIE OKORIE                |3.5  |W  50|D   6|L  38|L  34|W  52|W  48|U    |"
##  [90] "   MI | 15323285 / R: 1602P6 ->1508P12  |N:4  |B    |W    |B    |W    |W    |B    |     |"
##  [91] "-----------------------------------------------------------------------------------------"
##  [92] "   30 | GEORGE AVERY JONES              |3.5  |L  52|D  64|L  15|W  55|L  31|W  61|W  50|"
##  [93] "   ON | 12577178 / R: 1522   ->1444     |     |W    |B    |B    |W    |W    |B    |B    |"
##  [94] "-----------------------------------------------------------------------------------------"
##  [95] "   31 | RISHI SHETTY                    |3.5  |L  58|D  55|W  64|L  10|W  30|W  50|L  14|"
##  [96] "   MI | 15131618 / R: 1494   ->1444     |     |B    |W    |B    |W    |B    |W    |B    |"
##  [97] "-----------------------------------------------------------------------------------------"
##  [98] "   32 | JOSHUA PHILIP MATHEWS           |3.5  |W  61|L   8|W  44|L  18|W  51|D  26|L  13|"
##  [99] "   ON | 14073750 / R: 1441   ->1433     |N:4  |W    |B    |W    |B    |W    |B    |W    |"
## [100] "-----------------------------------------------------------------------------------------"
## [101] "   33 | JADE GE                         |3.5  |W  60|L  12|W  50|D  36|L  13|L  15|W  51|"
## [102] "   MI | 14691842 / R: 1449   ->1421     |     |B    |W    |B    |W    |B    |W    |B    |"
## [103] "-----------------------------------------------------------------------------------------"
## [104] "   34 | MICHAEL JEFFERY THOMAS          |3.5  |L   6|W  60|L  37|W  29|D  25|L  11|W  52|"
## [105] "   MI | 15051807 / R: 1399   ->1400     |     |B    |W    |B    |B    |W    |B    |W    |"
## [106] "-----------------------------------------------------------------------------------------"
## [107] "   35 | JOSHUA DAVID LEE                |3.5  |L  46|L  38|W  56|L   6|W  57|D  52|W  48|"
## [108] "   MI | 14601397 / R: 1438   ->1392     |     |W    |W    |B    |W    |B    |B    |W    |"
## [109] "-----------------------------------------------------------------------------------------"
## [110] "   36 | SIDDHARTH JHA                   |3.5  |L  13|W  57|W  51|D  33|H    |L  16|D  28|"
## [111] "   MI | 14773163 / R: 1355   ->1367     |N:4  |W    |B    |W    |B    |     |W    |B    |"
## [112] "-----------------------------------------------------------------------------------------"
## [113] "   37 | AMIYATOSH PWNANANDAM            |3.5  |B    |L   5|W  34|L  27|H    |L  23|W  61|"
## [114] "   MI | 15489571 / R:  980P12->1077P17  |     |     |B    |W    |W    |     |B    |W    |"
## [115] "-----------------------------------------------------------------------------------------"
## [116] "   38 | BRIAN LIU                       |3.0  |D  11|W  35|W  29|L  12|H    |L  18|L  15|"
## [117] "   MI | 15108523 / R: 1423   ->1439     |N:4  |W    |B    |W    |W    |     |B    |B    |"
## [118] "-----------------------------------------------------------------------------------------"
## [119] "   39 | JOEL R HENDON                   |3.0  |L   1|W  54|W  40|L  16|W  44|L  21|L  24|"
## [120] "   MI | 12923035 / R: 1436P23->1413     |N:4  |B    |W    |B    |W    |B    |W    |W    |"
## [121] "-----------------------------------------------------------------------------------------"
## [122] "   40 | FOREST ZHANG                    |3.0  |W  20|L  26|L  39|W  59|L  21|W  56|L  22|"
## [123] "   MI | 14892710 / R: 1348   ->1346     |     |B    |B    |W    |W    |B    |W    |W    |"
## [124] "-----------------------------------------------------------------------------------------"
## [125] "   41 | KYLE WILLIAM MURPHY             |3.0  |W  59|L  17|W  58|L  20|X    |U    |U    |"
## [126] "   MI | 15761443 / R: 1403P5 ->1341P9   |     |B    |W    |B    |W    |     |     |     |"
## [127] "-----------------------------------------------------------------------------------------"
## [128] "   42 | JARED GE                        |3.0  |L  12|L  50|L  57|D  60|D  61|W  64|W  56|"
```

```
## [129] "   MI | 14462326 / R: 1332   ->1256     |     |B    |W    |B    |B    |W    |W    |B    |"
## [130] "-----------------------------------------------------------------------------------------"
## [131] "   43 | ROBERT GLEN VASEY               |3.0  |L  21|L  23|L  24|W  63|W  59|L  46|W  55|"
## [132] "   MI | 14101068 / R: 1283   ->1244     |     |W    |B    |W    |W    |B    |B    |W    |"
## [133] "-----------------------------------------------------------------------------------------"
## [134] "   44 | JUSTIN D SCHILLING              |3.0  |B    |L  14|L  32|W  53|L  39|L  24|W  59|"
## [135] "   MI | 15323504 / R: 1199   ->1199     |     |     |W    |B    |B    |W    |B    |W    |"
## [136] "-----------------------------------------------------------------------------------------"
## [137] "   45 | DEREK YAN                       |3.0  |L   5|L  51|D  60|L  56|W  63|D  55|W  58|"
## [138] "   MI | 15372807 / R: 1242   ->1191     |     |W    |B    |W    |B    |W    |B    |W    |"
## [139] "-----------------------------------------------------------------------------------------"
## [140] "   46 | JACOB ALEXANDER LAVALLEY        |3.0  |W  35|L   7|L  27|L  50|W  64|W  43|L  23|"
## [141] "   MI | 15490981 / R:  377P3 ->1076P10  |     |B    |W    |B    |W    |B    |W    |W    |"
## [142] "-----------------------------------------------------------------------------------------"
## [143] "   47 | ERIC WRIGHT                     |2.5  |L  18|W  24|L  21|W  61|L   8|D  51|L  25|"
## [144] "   MI | 12533115 / R: 1362   ->1341     |     |W    |B    |W    |B    |W    |B    |W    |"
## [145] "-----------------------------------------------------------------------------------------"
## [146] "   48 | DANIEL KHAIN                    |2.5  |L  17|W  63|H    |D  52|H    |L  29|L  35|"
## [147] "   MI | 14369165 / R: 1382   ->1335     |     |B    |W    |     |B    |     |W    |B    |"
## [148] "-----------------------------------------------------------------------------------------"
## [149] "   49 | MICHAEL J MARTIN                |2.5  |L  26|L  20|D  63|D  64|W  58|H    |U    |"
## [150] "   MI | 12531685 / R: 1291P12->1259P17  |     |W    |W    |B    |W    |B    |     |     |"
## [151] "-----------------------------------------------------------------------------------------"
## [152] "   50 | SHIVAM JHA                      |2.5  |L  29|W  42|L  33|W  46|H    |L  31|L  30|"
## [153] "   MI | 14773178 / R: 1056   ->1111     |     |W    |B    |W    |B    |     |B    |W    |"
## [154] "-----------------------------------------------------------------------------------------"
## [155] "   51 | TEJAS AYYAGARI                  |2.5  |L  27|W  45|L  36|W  57|L  32|D  47|L  33|"
## [156] "   MI | 15205474 / R: 1011   ->1097     |     |B    |W    |B    |W    |B    |W    |W    |"
## [157] "-----------------------------------------------------------------------------------------"
## [158] "   52 | ETHAN GUO                       |2.5  |W  30|D  22|L  19|D  48|L  29|D  35|L  34|"
## [159] "   MI | 14918803 / R:  935   ->1092     |N:4  |B    |W    |B    |W    |B    |W    |B    |"
## [160] "-----------------------------------------------------------------------------------------"
## [161] "   53 | JOSE C YBARRA                   |2.0  |H    |L  25|H    |L  44|U    |W  57|U    |"
## [162] "   MI | 12578849 / R: 1393   ->1359     |     |     |B    |     |W    |     |W    |     |"
## [163] "-----------------------------------------------------------------------------------------"
## [164] "   54 | LARRY HODGE                     |2.0  |L  14|L  39|L  61|B    |L  15|L  59|W  64|"
## [165] "   MI | 12836773 / R: 1270   ->1200     |     |B    |B    |W    |     |W    |B    |W    |"
## [166] "-----------------------------------------------------------------------------------------"
## [167] "   55 | ALEX KONG                       |2.0  |L  62|D  31|L  10|L  30|B    |D  45|L  43|"
## [168] "   MI | 15412571 / R: 1186   ->1163     |     |W    |B    |W    |B    |     |W    |B    |"
## [169] "-----------------------------------------------------------------------------------------"
## [170] "   56 | MARISA RICCI                    |2.0  |H    |L  11|L  35|W  45|H    |L  40|L  42|"
## [171] "   MI | 14679887 / R: 1153   ->1140     |     |     |B    |W    |W    |     |B    |W    |"
## [172] "-----------------------------------------------------------------------------------------"
## [173] "   57 | MICHAEL LU                      |2.0  |L   7|L  36|W  42|L  51|L  35|L  53|B    |"
## [174] "   MI | 15113330 / R: 1092   ->1079     |     |B    |W    |W    |B    |W    |B    |     |"
## [175] "-----------------------------------------------------------------------------------------"
## [176] "   58 | VIRAJ MOHILE                    |2.0  |W  31|L   2|L  41|L  23|L  49|B    |L  45|"
## [177] "   MI | 14700365 / R:  917   -> 941     |     |W    |B    |W    |B    |W    |     |B    |"
## [178] "-----------------------------------------------------------------------------------------"
## [179] "   59 | SEAN M MC CORMICK               |2.0  |L  41|B    |L   9|L  40|L  43|W  54|L  44|"
## [180] "   MI | 12841036 / R:  853   -> 878     |     |W    |     |B    |B    |W    |W    |B    |"
## [181] "-----------------------------------------------------------------------------------------"
## [182] "   60 | JULIA SHEN                      |1.5  |L  33|L  34|D  45|D  42|L  24|H    |U    |"
```

```
## [183] "   MI | 14579262 / R:  967   -> 984     |     |W    |B    |B    |W    |B    |     |     |"
## [184] "-----------------------------------------------------------------------------------------"
## [185] "   61 | JEZZEL FARKAS                   |1.5  |L  32|L   3|W  54|L  47|D  42|L  30|L  37|"
## [186] "   ON | 15771592 / R:  955P11-> 979P18  |     |B    |W    |B    |W    |B    |W    |B    |"
## [187] "-----------------------------------------------------------------------------------------"
## [188] "   62 | ASHWIN BALAJI                   |1.0  |W  55|U    |U    |U    |U    |U    |U    |"
## [189] "   MI | 15219542 / R: 1530   ->1535     |     |B    |     |     |     |     |     |     |"
## [190] "-----------------------------------------------------------------------------------------"
## [191] "   63 | THOMAS JOSEPH HOSMER            |1.0  |L   2|L  48|D  49|L  43|L  45|H    |U    |"
## [192] "   MI | 15057092 / R: 1175   ->1125     |     |W    |B    |W    |B    |B    |     |     |"
## [193] "-----------------------------------------------------------------------------------------"
## [194] "   64 | BEN LI                          |1.0  |L  22|D  30|L  31|D  49|L  46|L  42|L  54|"
## [195] "   MI | 15006561 / R: 1163   ->1112     |     |B    |W    |W    |B    |W    |B    |B    |"
## [196] "-----------------------------------------------------------------------------------------"
```

Here we see that results_file shows each separate line as a string. All the metrics and information that we need to collect for each player lies on two separate and consecutive lines. The player name, total points, and player id are on the first line, and the player state and pre-rating are on the second line. The first line for player 1 (Gary Hua) starts on row [5], and the next first line for player 2 (Dakshesh Daruri) is on row [8]. So here, we see that the first and second lines for each individual player are spaced out every 3 rows in results_file. We can then separate out all the line1s and line2s for each player into two separate vectors, this will now match the player id to their row index:

```
all_line1s <- results_file[seq(5, length(results_file),
    3)]
all_line2s <- results_file[seq(6, length(results_file),
    3)]
head(all_line1s)
```

```
## [1] "    1 | GARY HUA                       |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|"
## [2] "    2 | DAKSHESH DARURI                |6.0  |W  63|W  58|L   4|W  17|W  16|W  20|W   7|"
## [3] "    3 | ADITYA BAJAJ                   |6.0  |L   8|W  61|W  25|W  21|W  11|W  13|W  12|"
## [4] "    4 | PATRICK H SCHILLING            |5.5  |W  23|D  28|W   2|W  26|D   5|W  19|D   1|"
## [5] "    5 | HANSHI ZUO                     |5.5  |W  45|W  37|D  12|D  13|D   4|W  14|W  17|"
## [6] "    6 | HANSEN SONG                    |5.0  |W  34|D  29|L  11|W  35|D  10|W  27|W  21|"
```

```
head(all_line2s)
```

```
## [1] "   ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
## [2] "   MI | 14598900 / R: 1553   ->1663     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
## [3] "   MI | 14959604 / R: 1384   ->1640     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
## [4] "   MI | 12616049 / R: 1716   ->1744     |N:2  |W    |B    |W    |B    |W    |B    |B    |"
## [5] "   MI | 14601533 / R: 1655   ->1690     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
## [6] "   OH | 15055204 / R: 1686   ->1687     |N:3  |W    |B    |W    |B    |B    |W    |B    |"
```

The first piece of information we want to extract is the player's name. Here, we have to consider that all the players have a least a first and last name, but some players have up to 4 separate names and may also have hyphenated names. So to ensure that all the names have been extracted, we use the regular expression below:

```
names <- str_extract_all(all_line1s, "\\|\\s[\\D|[:punct:]]+\\s[\\D|[:punct:]]{2,}\\s+[\\D|[:punct:]]*\\
head(names)
```

```
## [[1]]
## [1] "| GARY HUA                         |"
##
## [[2]]
## [1] "| DAKSHESH DARURI                  |"
##
## [[3]]
## [1] "| ADITYA BAJAJ                     |"
##
## [[4]]
## [1] "| PATRICK H SCHILLING              |"
##
## [[5]]
## [1] "| HANSHI ZUO                       |"
##
## [[6]]
## [1] "| HANSEN SONG                      |"
```

This gave us all the player names and some extra space around them, so we can replace the | characters with blanks and trim the white spaces at the front:

```
clean_names <- str_trim(str_replace_all(names,
    "\\|", ""))
clean_names
```

```
##  [1] "GARY HUA"                   "DAKSHESH DARURI"
##  [3] "ADITYA BAJAJ"               "PATRICK H SCHILLING"
##  [5] "HANSHI ZUO"                 "HANSEN SONG"
##  [7] "GARY DEE SWATHELL"          "EZEKIEL HOUGHTON"
##  [9] "STEFANO LEE"                "ANVIT RAO"
## [11] "CAMERON WILLIAM MC LEMAN"   "KENNETH J TACK"
## [13] "TORRANCE HENRY JR"          "BRADLEY SHAW"
## [15] "ZACHARY JAMES HOUGHTON"     "MIKE NIKITIN"
## [17] "RONALD GRZEGORCZYK"         "DAVID SUNDEEN"
## [19] "DIPANKAR ROY"               "JASON ZHENG"
## [21] "DINH DANG BUI"              "EUGENE L MCCLURE"
## [23] "ALAN BUI"                   "MICHAEL R ALDRICH"
## [25] "LOREN SCHWIEBERT"           "MAX ZHU"
## [27] "GAURAV GIDWANI"             "SOFIA ADINA STANESCU-BELLU"
## [29] "CHIEDOZIE OKORIE"           "GEORGE AVERY JONES"
## [31] "RISHI SHETTY"               "JOSHUA PHILIP MATHEWS"
## [33] "JADE GE"                    "MICHAEL JEFFERY THOMAS"
## [35] "JOSHUA DAVID LEE"           "SIDDHARTH JHA"
## [37] "AMIYATOSH PWNANANDAM"       "BRIAN LIU"
## [39] "JOEL R HENDON"              "FOREST ZHANG"
## [41] "KYLE WILLIAM MURPHY"        "JARED GE"
## [43] "ROBERT GLEN VASEY"          "JUSTIN D SCHILLING"
## [45] "DEREK YAN"                  "JACOB ALEXANDER LAVALLEY"
## [47] "ERIC WRIGHT"                "DANIEL KHAIN"
```

```
## [49] "MICHAEL J MARTIN"           "SHIVAM JHA"
## [51] "TEJAS AYYAGARI"             "ETHAN GUO"
## [53] "JOSE C YBARRA"              "LARRY HODGE"
## [55] "ALEX KONG"                  "MARISA RICCI"
## [57] "MICHAEL LU"                 "VIRAJ MOHILE"
## [59] "SEAN M MC CORMICK"          "JULIA SHEN"
## [61] "JEZZEL FARKAS"              "ASHWIN BALAJI"
## [63] "THOMAS JOSEPH HOSMER"       "BEN LI"
```

Next, we extract the states from the all_line2s vector and trim the extra white space. Here, we can take advantage of the fact that the state acronym is the only occurence of two consecutive letters in each string:

```
state <- str_trim(str_extract(all_line2s,
    "\\s+\\w+"))
state
```

```
##  [1] "ON" "MI" "MI" "MI" "MI" "OH" "MI" "MI" "ON" "MI" "MI" "MI" "MI" "MI" "MI"
## [16] "MI" "MI" "MI" "MI" "MI" "ON" "MI" "ON" "MI" "MI" "ON" "MI" "MI" "MI" "ON"
## [31] "MI" "ON" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI"
## [46] "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI"
## [61] "ON" "MI" "MI" "MI"
```

For extracting the player's total points, we extract all instances of a "digit . digit" and then convert it to a numeric type:

```
total_points <- as.numeric(str_extract(all_line1s,
    "\\d\\.\\d"))
total_points
```

```
##  [1] 6.0 6.0 6.0 5.5 5.5 5.0 5.0 5.0 5.0 5.0 4.5 4.5 4.5 4.5 4.5 4.0 4.0 4.0 4.0
## [20] 4.0 4.0 4.0 4.0 4.0 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.0
## [39] 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 2.5 2.5 2.5 2.5 2.5 2.5 2.0 2.0 2.0 2.0 2.0
## [58] 2.0 2.0 1.5 1.5 1.0 1.0 1.0
```

To get the pre-rating, we use a regular expression which takes advantage of the fact that it is the series of 3 or 4 digits after the "R:". The "R:" part of the string is then removed and the remaining digits are converted to numerics:

```
pre_rating <- as.numeric(str_replace(str_extract(all_line2s,
    "R:[:blank:]+[:digit:]+"), "R: ", ""))
pre_rating
```

```
##  [1] 1794 1553 1384 1716 1655 1686 1649 1641 1411 1365 1712 1663 1666 1610 1220
## [16] 1604 1629 1600 1564 1595 1563 1555 1363 1229 1745 1579 1552 1507 1602 1522
## [31] 1494 1441 1449 1399 1438 1355  980 1423 1436 1348 1403 1332 1283 1199 1242
## [46]  377 1362 1382 1291 1056 1011  935 1393 1270 1186 1153 1092  917  853  967
## [61]  955 1530 1175 1163
```

Now that we have the player name, state, total points and pre_rating, we can organize them all into the following data frame:

```
player_stats_frame <- data.frame(clean_names,
    state, total_points, pre_rating)
player_stats_frame
```

```
##                   clean_names state total_points pre_rating
## 1                    GARY HUA    ON          6.0       1794
## 2              DAKSHESH DARURI    MI          6.0       1553
## 3                 ADITYA BAJAJ    MI          6.0       1384
## 4          PATRICK H SCHILLING    MI          5.5       1716
## 5                   HANSHI ZUO    MI          5.5       1655
## 6                  HANSEN SONG    OH          5.0       1686
## 7            GARY DEE SWATHELL    MI          5.0       1649
## 8             EZEKIEL HOUGHTON    MI          5.0       1641
## 9                  STEFANO LEE    ON          5.0       1411
## 10                   ANVIT RAO    MI          5.0       1365
## 11   CAMERON WILLIAM MC LEMAN    MI          4.5       1712
## 12              KENNETH J TACK    MI          4.5       1663
## 13          TORRANCE HENRY JR    MI          4.5       1666
## 14                BRADLEY SHAW    MI          4.5       1610
## 15     ZACHARY JAMES HOUGHTON    MI          4.5       1220
## 16                MIKE NIKITIN    MI          4.0       1604
## 17         RONALD GRZEGORCZYK    MI          4.0       1629
## 18              DAVID SUNDEEN    MI          4.0       1600
## 19                DIPANKAR ROY    MI          4.0       1564
## 20                JASON ZHENG    MI          4.0       1595
## 21              DINH DANG BUI    ON          4.0       1563
## 22            EUGENE L MCCLURE    MI          4.0       1555
## 23                    ALAN BUI    ON          4.0       1363
## 24          MICHAEL R ALDRICH    MI          4.0       1229
## 25           LOREN SCHWIEBERT    MI          3.5       1745
## 26                     MAX ZHU    ON          3.5       1579
## 27              GAURAV GIDWANI    MI          3.5       1552
## 28 SOFIA ADINA STANESCU-BELLU    MI          3.5       1507
## 29            CHIEDOZIE OKORIE    MI          3.5       1602
## 30         GEORGE AVERY JONES    ON          3.5       1522
## 31                RISHI SHETTY    MI          3.5       1494
## 32      JOSHUA PHILIP MATHEWS    ON          3.5       1441
## 33                    JADE GE    MI          3.5       1449
## 34     MICHAEL JEFFERY THOMAS    MI          3.5       1399
## 35           JOSHUA DAVID LEE    MI          3.5       1438
## 36              SIDDHARTH JHA    MI          3.5       1355
## 37       AMIYATOSH PWNANANDAM    MI          3.5        980
## 38                   BRIAN LIU    MI          3.0       1423
## 39               JOEL R HENDON    MI          3.0       1436
## 40                FOREST ZHANG    MI          3.0       1348
## 41        KYLE WILLIAM MURPHY    MI          3.0       1403
## 42                    JARED GE    MI          3.0       1332
## 43          ROBERT GLEN VASEY    MI          3.0       1283
## 44         JUSTIN D SCHILLING    MI          3.0       1199
## 45                   DEREK YAN    MI          3.0       1242
## 46   JACOB ALEXANDER LAVALLEY    MI          3.0        377
## 47                ERIC WRIGHT    MI          2.5       1362
## 48               DANIEL KHAIN    MI          2.5       1382
```

```
## 49           MICHAEL J MARTIN    MI          2.5       1291
## 50                 SHIVAM JHA    MI          2.5       1056
## 51             TEJAS AYYAGARI    MI          2.5       1011
## 52                  ETHAN GUO    MI          2.5        935
## 53              JOSE C YBARRA    MI          2.0       1393
## 54                LARRY HODGE    MI          2.0       1270
## 55                   ALEX KONG    MI          2.0       1186
## 56               MARISA RICCI    MI          2.0       1153
## 57                 MICHAEL LU    MI          2.0       1092
## 58               VIRAJ MOHILE    MI          2.0        917
## 59          SEAN M MC CORMICK    MI          2.0        853
## 60                 JULIA SHEN    MI          1.5        967
## 61              JEZZEL FARKAS    ON          1.5        955
## 62              ASHWIN BALAJI    MI          1.0       1530
## 63       THOMAS JOSEPH HOSMER    MI          1.0       1175
## 64                     BEN LI    MI          1.0       1163
```

The last metric left is the average pre-rating of the opponents that each player faced in the tournament. To calculate this, we must first, for every player, extract their opponents' player ids which are listed in each line1. To do this, we create a subset of each string, only looking at the characters in the "rounds" columns.

Extracting the digits from a single string subset will result in a list of strings, represnting each opponent's player id. When done for all 64 players, the result is a list of 64 lists of strings. This makes things tricky because not only do we have to convert these player ids to numeric types, we also have to use them to reference the pre-ratings of these players in the player stats data frame. Both of these steps were done at the same time by using nested for loops, one to iterate through each player, and the other to iterate through each of their opponent's player ID strings. Once each opponent's player ID was individually converted to numeric it was then used to reference the player's pre-rating, which was then stored in another list of 64 lists, this time showing the collection of opponent pre-ratings for each player. Finally, each of these lists could be averaged over, calculating the average pre-rating of the opponents for each player, and stored in a new list.

```
opponents <- str_extract_all(str_sub(all_line1s,
    start = 47), "\\d+")

opponent_pre_ratings <- vector(mode = "list",
    length = length(opponents))
AvgOppPreRatings <- vector(mode = "list",
    length = length(opponents))

for (i in seq(1, length(all_line1s))) {
    for (j in seq(1, length(opponents[[i]]))) {

        opponents_numeric <- as.numeric(opponents[[i]][j])

        opponent_pre_ratings[[i]][j] <- player_stats_frame$pre_rating[opponents_numeric]

    }

    AvgOppPreRatings[i] <- round(mean(opponent_pre_ratings[[i]]))

}
# opponents_numeric
# player_stats_frame$pre_rating[39]
```

```
# head(opponent_pre_ratings)
# head(AvgOppPreRatings)
```

The list was then converted into a vector so that it could be easily appended to the player stats data frame.

```
AvgOppPreRatingsVector <- unlist(AvgOppPreRatings)

player_stats_frame$Avg_Opponent_Pre_Rating <- AvgOppPreRatingsVector

player_stats_frame
```

```
##                   clean_names state total_points pre_rating
## 1                    GARY HUA    ON          6.0       1794
## 2              DAKSHESH DARURI    MI          6.0       1553
## 3                 ADITYA BAJAJ    MI          6.0       1384
## 4          PATRICK H SCHILLING    MI          5.5       1716
## 5                   HANSHI ZUO    MI          5.5       1655
## 6                  HANSEN SONG    OH          5.0       1686
## 7            GARY DEE SWATHELL    MI          5.0       1649
## 8             EZEKIEL HOUGHTON    MI          5.0       1641
## 9                  STEFANO LEE    ON          5.0       1411
## 10                   ANVIT RAO    MI          5.0       1365
## 11   CAMERON WILLIAM MC LEMAN    MI          4.5       1712
## 12              KENNETH J TACK    MI          4.5       1663
## 13          TORRANCE HENRY JR    MI          4.5       1666
## 14               BRADLEY SHAW    MI          4.5       1610
## 15     ZACHARY JAMES HOUGHTON    MI          4.5       1220
## 16                MIKE NIKITIN    MI          4.0       1604
## 17         RONALD GRZEGORCZYK    MI          4.0       1629
## 18               DAVID SUNDEEN    MI          4.0       1600
## 19                DIPANKAR ROY    MI          4.0       1564
## 20                JASON ZHENG    MI          4.0       1595
## 21              DINH DANG BUI    ON          4.0       1563
## 22            EUGENE L MCCLURE    MI          4.0       1555
## 23                    ALAN BUI    ON          4.0       1363
## 24           MICHAEL R ALDRICH    MI          4.0       1229
## 25            LOREN SCHWIEBERT    MI          3.5       1745
## 26                     MAX ZHU    ON          3.5       1579
## 27              GAURAV GIDWANI    MI          3.5       1552
## 28 SOFIA ADINA STANESCU-BELLU    MI          3.5       1507
## 29            CHIEDOZIE OKORIE    MI          3.5       1602
## 30         GEORGE AVERY JONES    ON          3.5       1522
## 31                RISHI SHETTY    MI          3.5       1494
## 32      JOSHUA PHILIP MATHEWS    ON          3.5       1441
## 33                    JADE GE    MI          3.5       1449
## 34     MICHAEL JEFFERY THOMAS    MI          3.5       1399
## 35            JOSHUA DAVID LEE    MI          3.5       1438
## 36              SIDDHARTH JHA    MI          3.5       1355
## 37       AMIYATOSH PWNANANDAM    MI          3.5        980
## 38                   BRIAN LIU    MI          3.0       1423
## 39               JOEL R HENDON    MI          3.0       1436
## 40                FOREST ZHANG    MI          3.0       1348
## 41        KYLE WILLIAM MURPHY    MI          3.0       1403
```

```
## 42                  JARED GE    MI          3.0       1332
## 43          ROBERT GLEN VASEY    MI          3.0       1283
## 44         JUSTIN D SCHILLING    MI          3.0       1199
## 45                  DEREK YAN    MI          3.0       1242
## 46   JACOB ALEXANDER LAVALLEY    MI          3.0        377
## 47                ERIC WRIGHT    MI          2.5       1362
## 48               DANIEL KHAIN    MI          2.5       1382
## 49           MICHAEL J MARTIN    MI          2.5       1291
## 50                 SHIVAM JHA    MI          2.5       1056
## 51             TEJAS AYYAGARI    MI          2.5       1011
## 52                  ETHAN GUO    MI          2.5        935
## 53              JOSE C YBARRA    MI          2.0       1393
## 54                LARRY HODGE    MI          2.0       1270
## 55                  ALEX KONG    MI          2.0       1186
## 56               MARISA RICCI    MI          2.0       1153
## 57                 MICHAEL LU    MI          2.0       1092
## 58               VIRAJ MOHILE    MI          2.0        917
## 59          SEAN M MC CORMICK    MI          2.0        853
## 60                 JULIA SHEN    MI          1.5        967
## 61              JEZZEL FARKAS    ON          1.5        955
## 62              ASHWIN BALAJI    MI          1.0       1530
## 63       THOMAS JOSEPH HOSMER    MI          1.0       1175
## 64                    BEN LI    MI          1.0       1163
##    Avg_Opponent_Pre_Rating
## 1                     1605
## 2                     1469
## 3                     1564
## 4                     1574
## 5                     1501
## 6                     1519
## 7                     1372
## 8                     1468
## 9                     1523
## 10                    1554
## 11                    1468
## 12                    1506
## 13                    1498
## 14                    1515
## 15                    1484
## 16                    1386
## 17                    1499
## 18                    1480
## 19                    1426
## 20                    1411
## 21                    1470
## 22                    1300
## 23                    1214
## 24                    1357
## 25                    1363
## 26                    1507
## 27                    1222
## 28                    1522
## 29                    1314
## 30                    1144
```

```
## 31                    1260
## 32                    1379
## 33                    1277
## 34                    1375
## 35                    1150
## 36                    1388
## 37                    1385
## 38                    1539
## 39                    1430
## 40                    1391
## 41                    1248
## 42                    1150
## 43                    1107
## 44                    1327
## 45                    1152
## 46                    1358
## 47                    1392
## 48                    1356
## 49                    1286
## 50                    1296
## 51                    1356
## 52                    1495
## 53                    1345
## 54                    1206
## 55                    1406
## 56                    1414
## 57                    1363
## 58                    1391
## 59                    1319
## 60                    1330
## 61                    1327
## 62                    1186
## 63                    1350
## 64                    1263
```

Finally, we can take our player_stats data frame and write it to a .csv file.

```
write.csv(player_stats_frame, "player_stats_file.csv")
```