607 - Chess Tournament Project

Karim Hammoud

2020-09-19

Overview

In this project, you're given a text file with chess tournament results where the information has some structure. Your job is to create an R Markdown file that generates a .CSV file (that could for example be imported into a SQL database) with the following information for all of the players:

Player's Name, Player's State, Total Number of Points, Player's Pre-Rating, and Average Pre Chess Rating of Opponents

For the first player, the information would be:

Gary Hua, ON, 6.0, 1794, 1605

1605 was calculated by using the pre-tournament opponents' ratings of 1436, 1563, 1600, 1610, 1649, 1663, 1716, and dividing by the total number of games played.

The chess rating system (invented by a Minnesota statistician named Arpad Elo) has been used in many other contexts, including assessing relative strength of employment candidates by human resource departments.

Github link here

Rpubs link here

Import the required libraries

```
library(tidyverse)
library(openintro)
library(stringr)
```

Read the text file

```
url <- "https://raw.githubusercontent.com/akarimhammoud/CUNY-SPS/master/607-Data-Acquisition-and-Manage

tournamentinfo <- read.csv(paste0(url), header=F)
head (tournamentinfo)
```

```
##                                                                                          V1
## 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    |
```

```
tail(tournamentinfo)
```

```
##                                                                                          V1
## 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 -----------------------------------------------------------------------------------------
```

Data wrangling

Taking out the first four rows

```
tournamentinfo <- tournamentinfo[-c(1:4),]
head(tournamentinfo)
```

```
## [1] "    1 | GARY HUA                        |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|"
## [2] "   ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
## [3] "-----------------------------------------------------------------------------------------"
## [4] "    2 | DAKSHESH DARURI                 |6.0  |W  63|W  58|L   4|W  17|W  16|W  20|W   7|"
## [5] "   MI | 14598900 / R: 1553   ->1663     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
## [6] "-----------------------------------------------------------------------------------------"
```

Checking the remaining rows

```
length(tournamentinfo)
```

```
## [1] 192
```

Pulling the first and second rows of each three rows.

```
first_row <- tournamentinfo[seq(1, length(tournamentinfo), 3)]
head(first_row,2)
```

```
## [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|"
```

```
second_row <- tournamentinfo[seq(2, length(tournamentinfo), 3)]
head(second_row,2)
```

```
## [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    |"
```

Using regular expression to extract the Data.

```
#pair number
number <- as.integer(str_extract(first_row,'\\d+'))
number
```

```
##  [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] 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] 51 52 53 54 55 56 57 58 59 60 61 62 63 64
```

```
#player's name
player_name <- str_trim(str_extract(first_row, '(\\w+\\s){2,3}'))
#player's state
player_state <- str_extract(second_row, "\\w+")
#points
player_points <- as.numeric(str_extract(first_row, '\\d+\\.\\d+'))
head(player_points)
```

```
## [1] 6.0 6.0 6.0 5.5 5.5 5.0
```

```
# the rating
player_rating <- as.integer(str_extract(str_extract(second_row, '[^\\d]\\d{3,4}[^\\d]'), '\\d+'))
head(player_rating)
```

```
## [1] 1794 1553 1384 1716 1655 1686
```

```
# the opponents
opponents <- str_extract_all(str_extract_all(first_row, "\\d+\\|"), "\\d+")
```

```
## Warning in stri_extract_all_regex(string, pattern, simplify = simplify, :
## argument is not an atomic vector; coercing
```

```
head(opponents)
```

```
## [[1]]
## [1] "39" "21" "18" "14" "7"  "12" "4"
##
## [[2]]
## [1] "63" "58" "4"  "17" "16" "20" "7"
##
## [[3]]
## [1] "8"  "61" "25" "21" "11" "13" "12"
##
## [[4]]
## [1] "23" "28" "2"  "26" "5"  "19" "1"
##
## [[5]]
## [1] "45" "37" "12" "13" "4"  "14" "17"
##
## [[6]]
## [1] "34" "29" "11" "35" "10" "27" "21"
```

```
#count the result
won <- str_count(first_row,'\\Q|W \\E')
lost <- str_count(first_row, '\\Q|L, \\E')
draw <- str_count(first_row, '\\Q|D \\E')
```

Calculate the mean rating

```
mean_rating <- length(first_row)

for (i in 1:length(first_row)) {
  mean_rating[i] <- round(mean(player_rating[as.numeric(unlist(opponents[number[i]]))]), digits = 0)
}
```

The final data frame

```
final_data <- data.frame(player_name, player_state, player_points, player_rating, mean_rating)
head(final_data)
```

```
##           player_name player_state player_points player_rating mean_rating
## 1            GARY HUA           ON           6.0          1794        1605
## 2     DAKSHESH DARURI           MI           6.0          1553        1469
## 3        ADITYA BAJAJ           MI           6.0          1384        1564
## 4 PATRICK H SCHILLING           MI           5.5          1716        1574
## 5          HANSHI ZUO           MI           5.5          1655        1501
## 6         HANSEN SONG           OH           5.0          1686        1519
```

Change the heading names

```
colnames(final_data) <- c("Name", "State", "Points", "Rating", "Average Rating")
head(final_data)
```

```
##                 Name State Points Rating Average Rating
## 1            GARY HUA    ON    6.0   1794           1605
## 2     DAKSHESH DARURI    MI    6.0   1553           1469
## 3        ADITYA BAJAJ    MI    6.0   1384           1564
## 4 PATRICK H SCHILLING    MI    5.5   1716           1574
## 5          HANSHI ZUO    MI    5.5   1655           1501
## 6         HANSEN SONG    OH    5.0   1686           1519
```

```
tail(final_data)
```

```
##                   Name State Points Rating Average Rating
## 59            SEAN M MC    MI    2.0    853           1319
## 60           JULIA SHEN    MI    1.5    967           1330
## 61        JEZZEL FARKAS    ON    1.5    955           1327
## 62        ASHWIN BALAJI    MI    1.0   1530           1186
## 63 THOMAS JOSEPH HOSMER    MI    1.0   1175           1350
## 64               BEN LI    MI    1.0   1163           1263
```

Create the CSV file in the general folder in Mac

```
write.csv(final_data, file = "../Project1_607.csv")
```