

Project 1 - Chess Tournament

John Mazon

Data Analysis for Chess Tournament

```
#calling in the library for RCurl and stringr
library(RCurl)
library(stringr)
```

Incorporating text file from Github repo source

Used to read in the original text file from an accessible location. In this case i've uploaded the tournament info text file into my gitub repo

```
txt_file2 <- readLines("https://raw.githubusercontent.com/johnm1990/DATA607/master/tournamentinfo.txt")
# preview the first 10 lines
txt_file2[1:10]
```

```
## [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    |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    |W    |B    | "
## [10] "-----"
```

```
#preview for the last lines showing the players
txt_file2[190:196]
```

```
## [1] -----
## [2] "   63 | THOMAS JOSEPH HOSMER     |1.0  |L    |2|L  |48|D  |49|L  |43|L  |45|H  |U    | "
## [3] "   MI | 15057092 / R: 1175    ->1125 |     |W    |B   |W    |B   |B    |I    |     |     | "
## [4] "-----"
## [5] "   64 | BEN LI                  |1.0  |L    |22|D  |30|L  |31|D  |49|L  |46|L  |42|L  |54| "
## [6] "   MI | 15006561 / R: 1163    ->1112 |     |B    |W    |W    |B    |W    |B    |B    |B    | "
## [7] "-----"
```

Generating a CSV file - extract the players Names

```

## purpose is to extract the players names
# below creates an empty vector to store player names ie. [player_names]. We use mode character in this
player_names <- vector(mode = "character")
# start on the 5th line since this is where the names start, this can be see from raw txt on github
l <- 5
# implementing a while loop. while l is smaller than the number of lines in the text file.
while(l < length(txt_file2)) {
  # extract 8th to 40th elements of the current line and add to vector player_names
  player_names <- c(player_names, substr(txt_file2[l], 8, 40))

  #print the name of the current in the loop just for reference
  print(substr(txt_file2[l], 8, 40))

  # increase the line number by 3 [on to the next name]
  l <- l + 3
}

```

```

## [1] " GARY HUA "
## [1] " DAKSHESH DARURI "
## [1] " ADITYA BAJAJ "
## [1] " PATRICK H SCHILLING "
## [1] " HANSHI ZUO "
## [1] " HANSEN SONG "
## [1] " GARY DEE SWATHELL "
## [1] " EZEKIEL HOUGHTON "
## [1] " STEFANO LEE "
## [1] " ANVIT RAO "
## [1] " CAMERON WILLIAM MC LEMAN "
## [1] " KENNETH J TACK "
## [1] " TORRANCE HENRY JR "
## [1] " BRADLEY SHAW "
## [1] " ZACHARY JAMES HOUGHTON "
## [1] " MIKE NIKITIN "
## [1] " RONALD GRZEGORCZYK "
## [1] " DAVID SUNDEEN "
## [1] " DIPANKAR ROY "
## [1] " JASON ZHENG "
## [1] " DINH DANG BUI "
## [1] " EUGENE L MCCLURE "
## [1] " ALAN BUI "
## [1] " MICHAEL R ALDRICH "
## [1] " LOREN SCHWIEBERT "
## [1] " MAX ZHU "
## [1] " GAURAV GIDWANI "
## [1] " SOFIA ADINA STANESCU-BELLU "
## [1] " CHIEDOZIE OKORIE "
## [1] " GEORGE AVERY JONES "
## [1] " RISHI SHETTY "
## [1] " JOSHUA PHILIP MATHEWS "
## [1] " JADE GE "
## [1] " MICHAEL JEFFERY THOMAS "
## [1] " JOSHUA DAVID LEE "

```

```

## [1] " SIDDHARTH JHA "
## [1] " AMIYATOSH PWNANANDAM "
## [1] " BRIAN LIU "
## [1] " JOEL R HENDON "
## [1] " FOREST ZHANG "
## [1] " KYLE WILLIAM MURPHY "
## [1] " JARED GE "
## [1] " ROBERT GLEN VASEY "
## [1] " JUSTIN D SCHILLING "
## [1] " DEREK YAN "
## [1] " JACOB ALEXANDER LAVALLEY "
## [1] " ERIC WRIGHT "
## [1] " DANIEL KHAIN "
## [1] " MICHAEL J MARTIN "
## [1] " SHIVAM JHA "
## [1] " TEJAS AYYAGARI "
## [1] " ETHAN GUO "
## [1] " JOSE C YBARRA "
## [1] " LARRY HODGE "
## [1] " ALEX KONG "
## [1] " MARISA RICCI "
## [1] " MICHAEL LU "
## [1] " VIRAJ MOHILE "
## [1] " SEAN M MC CORMICK "
## [1] " JULIA SHEN "
## [1] " JEZZEL FARKAS "
## [1] " ASHWIN BALAJI "
## [1] " THOMAS JOSEPH HOSMER "
## [1] " BEN LI "

```

Cleaning up the data for player names

```
# remove leading and trailing whitespace. trimws returns a character string with leading and/or trailing
player_names <- trimws(player_names)
```

```
# make only the first letters of name and surname upper case. I used str_to_title which will accomplish
player_names <- str_to_title(player_names)
# preview the first ten player names. we use this head function as reference to see our result so far
head(player_names, 10)
```

```

## [1] "Gary Hua"           "Dakshesh Daruri"      "Aditya Bajaj"
## [4] "Patrick H Schilling" "Hanshi Zuo"          "Hansen Song"
## [7] "Gary Dee Swathell"   "Ezekiel Houghton"    "Stefano Lee"
## [10] "Anvit Rao"          ""
```

Extracting and organizing state info

```
## grab the state information for all players
# create an empty vector to store player states
```

```
player_states <- vector(mode = "character")
# start on the 6th line
l <- 6
# while l is smaller than the number of lines in the text file
while(l < length(txt_file2)) {
  # extract 4th and 5th elements of the current line and add to vector player_states
  player_states <- c(player_states, substr(txt_file2[l], 4, 5))
  # increase the line number by 3
  print(substr(txt_file2[l], 4, 5))
  l <- l + 3
}
```

```

## [1] "MI"
## [1] "ON"
## [1] "MI"
## [1] "MI"
## [1] "MI"

```

Cleaning up the data for the states

```

# remove leading and trailing whitespace. trimws returns a character string with leading and/or trailing
player_states <- trimws(player_states)

# preview the first ten player states. we use this head function as reference to see our result so far
head(player_states, 10)

```

```

## [1] "ON" "MI" "MI" "MI" "MI" "OH" "MI" "MI" "ON" "MI"

```

Total Points Extracting information

```

## extract Total Number of Points for all players
# create an empty vector to store player points
player_points <- vector(mode = "character")
# start on the 5th line
l <- 5
# while l is smaller than the number of lines in the text file
while(l < length(txt_file2)) {
  # extract 42nd to 44th elements of the current line and add to vector player_points
  player_points <- c(player_points, substr(txt_file2[l], 42, 44))
  # increase the line number by 3
  print(substr(txt_file2[l], 42, 44))
  l <- l + 3
}

```

```
## [1] "6.0"
## [1] "6.0"
## [1] "6.0"
## [1] "5.5"
## [1] "5.5"
## [1] "5.0"
## [1] "5.0"
## [1] "5.0"
## [1] "5.0"
## [1] "5.0"
## [1] "5.0"
## [1] "5.0"
## [1] "4.5"
## [1] "4.5"
## [1] "4.5"
## [1] "4.5"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "4.0"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.5"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "3.0"
## [1] "2.5"
## [1] "2.5"
## [1] "2.5"
## [1] "2.5"
## [1] "2.5"
## [1] "2.0"
## [1] "2.0"
```

```

## [1] "2.0"
## [1] "2.0"
## [1] "2.0"
## [1] "2.0"
## [1] "2.0"
## [1] "1.5"
## [1] "1.5"
## [1] "1.0"
## [1] "1.0"
## [1] "1.0"

# as.numeric function to convert to numeric.
player_points <- as.numeric(player_points)
# preview the first five player points
head(player_points, 5)

## [1] 6.0 6.0 6.0 5.5 5.5

```

Pre-rating Extracting information

```

## below was created to extract the players pre-rating
# create an empty vector to store player ratings
player_rating <- vector(mode = "character")
# start on the 6th line
l <- 6
# while l is smaller than the number of lines in the text file
while(l < length(txt_file2)) {
  # extract 23rd to 26th elements of the current line and add to vector player_rating
  player_rating <- c(player_rating, substr(txt_file2[l], 23, 26))
  # increase the line number by 3
  print(substr(txt_file2[l], 23, 26))
  l <- l + 3
}

## [1] "1794"
## [1] "1553"
## [1] "1384"
## [1] "1716"
## [1] "1655"
## [1] "1686"
## [1] "1649"
## [1] "1641"
## [1] "1411"
## [1] "1365"
## [1] "1712"
## [1] "1663"
## [1] "1666"
## [1] "1610"
## [1] "1220"
## [1] "1604"
## [1] "1629"

```

```

## [1] "1600"
## [1] "1564"
## [1] "1595"
## [1] "1563"
## [1] "1555"
## [1] "1363"
## [1] "1229"
## [1] "1745"
## [1] "1579"
## [1] "1552"
## [1] "1507"
## [1] "1602"
## [1] "1522"
## [1] "1494"
## [1] "1441"
## [1] "1449"
## [1] "1399"
## [1] "1438"
## [1] "1355"
## [1] " 980"
## [1] "1423"
## [1] "1436"
## [1] "1348"
## [1] "1403"
## [1] "1332"
## [1] "1283"
## [1] "1199"
## [1] "1242"
## [1] " 377"
## [1] "1362"
## [1] "1382"
## [1] "1291"
## [1] "1056"
## [1] "1011"
## [1] " 935"
## [1] "1393"
## [1] "1270"
## [1] "1186"
## [1] "1153"
## [1] "1092"
## [1] " 917"
## [1] " 853"
## [1] " 967"
## [1] " 955"
## [1] "1530"
## [1] "1175"
## [1] "1163"

# convert to numeric
player_rating <- as.numeric(player_rating)
# preview the first five player ratings
head(player_rating, 5)

```

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

Opponents Extracting information

```
## used to extract which opponents each player played against
# create an empty list to store each player's opponents
player_opponents <- vector("list", length(player_names))
# start on the 5th line
l <- 5
# list position
p <- 1
# while l is smaller than the number of lines in the text file
while(l < length(txt_file2)) {
  # for i in 51, 57, 63, 69, 75, 81, 87
  for(i in seq(51, 87, by = 6)) {
    # extract ith and (i+1)th elements of the current line and add it to the list for pth player
    player_opponents[[p]] <- c(player_opponents[[p]], substr(txt_file2[l], i, i+1))
  }
  # increase the line number by 3
  l <- l + 3
  # increase list position by 1
  p <- p + 1
}

# convert each element of list to numeric vector
player_opponents <- lapply(player_opponents, as.numeric)
head(player_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
```

Avg rating calculation

```
## used to calculate the average pre chess rating of Opponents for all players
# create an empty vector to store average opponent ratings
avg_rating <- vector(mode = "numeric")
# for each player
```

```

for(i in 1:length(player_opponents)) {
  # find which opponents the player had
  opponents_ind <- player_opponents[[i]]
  # remove missing values if there are any
  opponents_ind <- na.omit(opponents_ind)
  # find the average rating of opponents
  avg_rating[i] <- mean(player_rating[opponents_ind])
}
# round the average rating to nearest integer
avg_rating <- round(avg_rating)
# preview the first four player average rating of opponents
head(avg_rating, 4)

## [1] 1605 1469 1564 1574

# create the data frame
df_tournament <- data.frame(Player_Name = player_names,
                             Player_State = player_states,
                             Total_Points = player_points,
                             Player_Pre_Rating = player_rating,
                             Avg_Opponents_Pre_Rating = avg_rating,
                             stringsAsFactors = FALSE)
# preview the first four lines of the data frame
head(df_tournament, 4)

##      Player_Name Player_State Total_Points Player_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
##   Avg_Opponents_Pre_Rating
## 1                      1605
## 2                      1469
## 3                      1564
## 4                      1574

```

Writing data to CSV [absolute path]

```

# As instructed in the Project 1 rubric, we did NOT include a relative path. Instead we used our absolute path
write.csv(df_tournament, "C:\\\\Users\\\\localadmin\\\\Documents\\\\Data607\\\\DATA 607\\\\project1\\\\final\\\\tournament.csv")

```