Project1

2023-09-25

```
chess_data <- read.csv("https://raw.githubusercontent.com/Kingtilon1/DATA607/main/tournamentinfo.txt", s
chess_data2 <- str_split(chess_data[, ], "-", simplify = TRUE)
```

store players info

```
player_names <- unlist(str_extract_all(chess_data2[, ], "\\w+[[:space:]]\\w+([[:space:]]\\w+)*", simpli
player_names <- player_names[!player_names[,] == "", ]

## states
player_states <- unlist(str_extract_all(chess_data2[, ], "[A-Z][A-Z][[:space:]][\\|]"))
player_states <- str_split(player_states, "[[:space:]][\\|]", simplify = TRUE)
player_states <- player_states[, -2]

## total points
totalPoints <- unlist(str_extract_all(chess_data2[, ], "(\\d+)[.](\\d+)", simplify = TRUE))
totalPoints <- totalPoints[!totalPoints[,] == "", ]



## player rating
player_ratings <- unlist(str_extract_all(chess_data2[, ], "[R:]([[:space:]]+)([[:alnum:]]+)([[:space:]]+
player_ratings <- unlist(str_extract_all(player_ratings, "\\d+[[:alnum:]]+", simplify = TRUE))
player_ratings <- unlist(str_extract_all(player_ratings, "\\d\\d\\d+", simplify = TRUE))
player_ratings <- player_ratings[!player_ratings[,] == "", ]
player_ratings <- as.numeric(player_ratings)

## opponent strings
Opponent_Info <- unlist(str_extract_all(chess_data2[, ], "([\\|][A-Z]([[:space:]]+)\\d*[\\|])([A-Z]([[:s
Opponents_Data <- matrix(ncol = 7)
Opponents_Data <- unlist(str_extract_all(Opponent_Info[, ], "\\d+", simplify = TRUE))
Opponents_Data <- Opponents_Data[rowSums(Opponents_Data == "") != ncol(Opponents_Data), ]
```

Now lets calculate the opponents average player ratings and create a dataframe to store the results

```
# Create a matrix of numeric opponent ratings
opponent_ratings <- matrix(NA, nrow = nrow(Opponents_Data), ncol = ncol(Opponents_Data))

# Fill the matrix with numeric opponent ratings
for (row in 1:nrow(Opponents_Data)) {
```

```
  for (col in 1:ncol(Opponents_Data)) {
    if (Opponents_Data[row, col] != "") {
      index <- Opponents_Data[row, col]
      index <- strtoi(index, base = 0L)
      opponent_ratings[row, col] <- player_ratings[index]
    }
  }
}

# Calculate mean while ignoring NAs
averages_of_ratings <- rowMeans(opponent_ratings, na.rm = TRUE)

# Create a data frame with the results
tourn_results <- data.frame(
  "Player Name" = player_names,
  "State" = player_states,
  "Points" = totalPoints,
  "Player_Rating" = player_ratings,
  "Opponent Avg Player_Rating" = round(averages_of_ratings)
)
```

lets view the new data frame holding the organzied data

```
head(tourn_results)
```

```
##           Player.Name State Points Player_Rating Opponent.Avg.Player_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
```

save to csv

```
write.csv(tourn_results, file = "finished_results.csv", row.names = FALSE)
```