

Week3_Project1_Chess

Shyam BV

September 17, 2016

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 Excel) 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. 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, and dividing by the total number of games played.

If you have questions about the meaning of the data or the results, please post them on the discussion board. Chess, like science, is a game of back and forth...

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.

Import necessary Library

```
library(stringr)
```

```
chess_data <- read.table("C:/CUNY/Courses/CUNY-repository/607/Project1/Data/tournamentinfo.txt",header=1)
```

```
combined_data <- paste (chess_data[seq(5,196,3),], chess_data[seq(6,196,3),1],sep = "")
```

```
str_location <- str_locate_all(combined_data,"\\|")
```

```
#New version with all columns
```

```
datalist <- vector("list",NROW(combined_data))
```

```
for(i in 1:NROW(combined_data)) {
```

```
  for(j in 1:20) {
```

```
    datalist[[i]][j] <- str_trim(substr(combined_data[i],str_location[[i]][j],str_location[[i]][j+1]),"both")
```

```
  }
```

```
}
```

```
csvdf <- do.call(rbind.data.frame, datalist)
```

```

csvdf <- setNames(csvdf,c(1:20))

#End of new version

#New df without /

without_symbols <- sapply(csvdf,function(x){
  x <- str_replace(x,"\\\\\\\\s?", "")
  x <- str_replace(x,"\\\\s+?\\\\\\\\|", "")
  x <- str_replace(x,"\\\\\\\\|", "")
  x <- str_trim(x,"both")
  return(x)
})

without_symbols_df <- data.frame(without_symbols)

avg_df <- data.frame(without_symbols_df$X1,without_symbols_df$X2,str_extract(without_symbols_df$X3,"\\\\d+"))

avg_df <- setNames(avg_df,c("Player","Points","Play1","Play2","Play3","Play4","Play5","Play6","Play7","Play8","Play9"))

avg_df[is.na(avg_df)] <- 0

#Replace the player data points

i <- 0
j <- 0
avg_df1 <- avg_df
for(i in 1:nrow(avg_df1)) {
  for(j in 3:9) {

avg_df1[i,j] <- avg_df1[avg_df1[i,j],10]

  }
}

avg_df1[is.na(avg_df1)] <- 0

columns <- c(3:10)
avg_df1[,columns] <- apply(avg_df1[,columns],2,function(x){ as.numeric(as.character(x))})

avg_df1[is.na(avg_df1)] <- 0

avg_df1$Sum <- rowMeans(avg_df1[,3:9],1)

remove(chess_data,csvdf,without_symbols,without_symbols_df,columns,combined_data,datalist,i,j,str_locat

```