DATA607 - Project 1

Chess Tournament

Donald Butler

9/19/2021

Introduction

Load required R Libraries

```
library(tidyverse)
```

Import and Clean Data

Import Raw tournament information text file

```
options(width = 60)
infoFile = 'https://raw.githubusercontent.com/dab31415/DATA607/main/Projects/Project_1/TournamentInfo.t
raw_tournament <-read.table(file = infoFile,
                            sep = '|',
                            strip.white = TRUE,
                            fill = TRUE)

head(raw_tournament,7)
```

```
##                                                                                            V1
## 1 ------------------------------------------------------------------------------------------
## 2                                                                                        Pair
## 3                                                                                         Num
## 4 ------------------------------------------------------------------------------------------
## 5                                                                                           1
## 6                                                                                          ON
## 7 ------------------------------------------------------------------------------------------
##                            V2    V3    V4    V5    V6    V7
## 1                                                          
## 2                Player Name Total Round Round Round Round
## 3   USCF ID / Rtg (Pre->Post)   Pts     1     2     3     4
## 4                                                          
## 5                   GARY HUA   6.0 W  39 W  21 W  18 W  14
## 6 15445895 / R: 1794   ->1817   N:2     W     B     W     B
## 7
##      V8    V9   V10 V11
## 1                   NA
## 2 Round Round Round  NA
## 3     5     6     7  NA
## 4                   NA
## 5 W   7 D  12 D   4  NA
## 6     W     B     W  NA
## 7                   NA
```

Remove unneeded rows and columns

The first four rows of the text file contains headers and row separators. The last column is created when loading because each row ends with a column separator.

```
tournament <- raw_tournament[-c(1:4),-11]
```

Set Column Names

```
names(tournament)[1] = 'num_state'
names(tournament)[2] = 'name_rating'
names(tournament)[3] = 'points'
names(tournament)[4:10] = paste('R', 1:7, sep = '')
```

Extract player details

```
players <- as_tibble(tournament %>%
  filter(grepl('[0-9]',num_state)) %>%
  transmute(
    playerNum = as.numeric(num_state),
    playerName = name_rating,
    points = as.numeric(points),
  )) %>%
  rowid_to_column('id')
```

Extract results from each round by player

```
matchResults <- as_tibble(tournament %>%
  filter(grepl('[0-9]',num_state)) %>%
  gather(roundNum,roundResult,R1:R7) %>%
  select(playerNum = num_state,
         roundNum,
         roundResult) %>%
  mutate(
    playerNum = as.numeric(playerNum),
    roundNum = as.numeric(str_extract(roundNum,'\\d+')),
    roundOpponent = as.numeric(str_extract(roundResult,'\\d+')),
    roundResult = str_extract(roundResult,'[WLDBHU]')
  ))
```

Extract Player State and Rating Details

```
ratings <- as_tibble(tournament %>%
  filter(grepl('[A-Z]',num_state)) %>%
  separate(name_rating, into = c('USCF_ID','rawRatings'), sep = '/') %>%
  separate(rawRatings, into = c('rawpreRating','rawpostRating'), sep = '->') %>%
  transmute(
    state = num_state,
    USCF_ID = as.numeric(USCF_ID),
    preRating = as.numeric(str_extract(rawpreRating,'\\d{3,4}')),
    postRating = as.numeric(str_extract(rawpostRating,'\\d{3,4}'))
  )) %>%
  rowid_to_column('id')
```

Merge players and ratings

We avoid using player number to merge the two data frames because we'd be dependent on the raw text file being sorted and numbered in order. By implementing a row number, we ensure that consecutive rows in the data file are merged together.

```
playerDetails <- players %>%
  merge(ratings,by = 'id') %>%
  select(playerNum, USCF_ID, playerName, state, preRating, points, postRating)
```

Create CSV files

Generate a csv of the players and another for the results of the tournament. These files can then be used to load into a database. The files will be saved to the default environment path for the project.

```
write_csv(playerDetails,'playersDetails.csv')
write_csv(matchResults,'matchResults.csv')
```