DATA607: Project 1

James Williams

Import Libraries

```
library(tidyverse)
## -- Attaching packages --
## v ggplot2 3.2.1 v purrr
                             0.3.3
## v tibble 2.1.3 v dplyr 0.8.4
## v tidyr 1.0.2 v stringr 1.4.0
## v readr 1.3.1
                   v forcats 0.4.0
## -- Conflicts ------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
library(dplyr)
library(tidyr)
library(stringr)
library(stringi)
library(splitstackshape)
library(reshape)
## Attaching package: 'reshape'
## The following object is masked from 'package:dplyr':
##
##
      rename
## The following objects are masked from 'package:tidyr':
##
##
      expand, smiths
Import Data
df <- read.delim("https://raw.githubusercontent.com/jambawilliams/DATA607Project1/master/tournamentinfo
##
                                                                                    V1
## 2 Pair | Player Name
                                        |Total|Round|Round|Round|Round|Round|
```

```
## 3 Num | USCF ID / Rtg (Pre->Post) | Pts | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
## 4 -----
                                                                      7|D 12|D
## 5
        1 | GARY HUA
                                        |6.0 |W 39|W 21|W 18|W 14|W
                                                                                 41
      ON | 15445895 / R: 1794 ->1817
                                       |N:2 |W
## 6
                                                |B
                                                     l W
                                                             lΒ
                                                                ١w
                                                                     lΒ
                                                                                  -
                                                                            ١w
Clean Data
df 2 \leftarrow df[-(1:4),]
df_3 <- str_split(df_2, "\\|")</pre>
df_4 <- gsub("-", "", df_3)
df_5 <- gsub("\"","", df_4, fixed = TRUE)</pre>
df_6 <- stri_remove_empty_na(df_5)</pre>
head(df_6)
## [1] "c( 1 , GARY HUA
                                            , 6.0 , W 39, W 21, W 18, W 14, W 7, D 12, D
## [2] "c( ON , 15445895 / R: 1794
                                  >1817
                                          , N:2 , W , B , W
                                                                                , B
                                                                  , B , W
## [3] "c( 2 , DAKSHESH DARURI
                                            , 6.0 , W 63, W 58, L 4, W 17, W 16, W 20, W
## [4] "c( MI , 14598900 / R: 1553
                                           , N:2 , B , W , B , W , B
                                                                                , W
                                  >1663
## [5] "c( 3 , ADITYA BAJAJ
                                           , 6.0 , L 8, W 61, W 25, W 21, W 11, W 13, W
## [6] "c( MI , 14959604 / R: 1384 >1640
                                          , N:2 , W , B , W , B , W
                                                                                , B
Player Names
names <- unlist(str_extract_all(unlist(df_6), "([[:upper:]]+\\s){2,}"))</pre>
head(names)
## [1] "GARY HUA "
                          "DAKSHESH DARURI "
                                             "ADITYA BAJAJ "
## [4] "PATRICK H SCHILLING " "HANSHI ZUO "
                                            "HANSEN SONG "
Player's State
states <- unlist(str_extract_all(unlist(df_6), "(ON|MI|OH)\\s\\,"))</pre>
states <- gsub(" ,", "", states)</pre>
head(states)
## [1] "ON" "MI" "MI" "MI" "MI" "OH"
Player's Total Number of Points
points <- unlist(str_extract_all(unlist(df_6), "\\d\\.\\d"))</pre>
head(points)
## [1] "6.0" "6.0" "5.5" "5.5" "5.0"
Player's Pre-Rating
ratings <- as.integer(str_extract(ratings, "\\d+"))</pre>
head(ratings)
```

[1] 1794 1553 1384 1716 1655 1686

Summary Table

```
tournament <- data.frame(names, states, points, ratings)
head(tournament)</pre>
```

##		names	states	points	ratings
##	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
##	5	HANSHI ZUO	MI	5.5	1655
##	6	HANSEN SONG	OH	5.0	1686

Export Table to CSV

```
write.csv(tournament, file = "tournament.csv")
```