Project 1

2022-09-25

```
library(readr)
library(stringr)

Url <- 'https://raw.githubusercontent.com/arinolan/Project-1/main/tournamentinfo.txt'

raw_data <- read_csv(file = Url, col_names = FALSE)
```

```
## Rows: 196 Columns: 1
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): X1
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(raw_data)
```

```
## # A tibble: 6 x 1
##   X1
##   <chr>
## 1 -----------------------------------------------------------------------------~
## 2 Pair | Player Name                     |Total|Round|Round|Round|Round|Round|R~
## 3 Num  | USCF ID / Rtg (Pre->Post)       | Pts |  1  |  2  |  3  |  4  |  5  | ~
## 4 -----------------------------------------------------------------------------~
## 5 1 | GARY HUA                         |6.0  |W  39|W  21|W  18|W  14|W   7|D  1~
## 6 ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B  ~
```

```
tail(raw_data)
```

```
## # A tibble: 6 x 1
##   X1
##   <chr>
## 1 63 | THOMAS JOSEPH HOSMER            |1.0  |L   2|L  48|D  49|L  43|L  45|H  ~
## 2 MI | 15057092 / R: 1175   ->1125     |     |W    |B    |W    |B    |B    |   ~
## 3 -----------------------------------------------------------------------------~
## 4 64 | BEN LI                          |1.0  |L  22|D  30|L  31|D  49|L  46|L  ~
## 5 MI | 15006561 / R: 1163   ->1112     |     |B    |W    |W    |B    |W    |B  ~
## 6 -----------------------------------------------------------------------------~
```

```
m1 <- matrix(unlist(raw_data), byrow=TRUE)
m2 <- m1[seq(5, length(m1), 3)]
head(m2)
```

```
## [1] "1 | GARY HUA                       |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|"
## [2] "2 | DAKSHESH DARURI                |6.0  |W  63|W  58|L   4|W  17|W  16|W  20|W   7|"
## [3] "3 | ADITYA BAJAJ                   |6.0  |L   8|W  61|W  25|W  21|W  11|W  13|W  12|"
## [4] "4 | PATRICK H SCHILLING            |5.5  |W  23|D  28|W   2|W  26|D   5|W  19|D   1|"
## [5] "5 | HANSHI ZUO                     |5.5  |W  45|W  37|D  12|D  13|D   4|W  14|W  17|"
## [6] "6 | HANSEN SONG                    |5.0  |W  34|D  29|L  11|W  35|D  10|W  27|W  21|"
```

```
m3 <- m1[seq(6, length(m1), 3)]
head(m3)
```

```
## [1] "ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
## [2] "MI | 14598900 / R: 1553   ->1663     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
## [3] "MI | 14959604 / R: 1384   ->1640     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
## [4] "MI | 12616049 / R: 1716   ->1744     |N:2  |W    |B    |W    |B    |W    |B    |B    |"
## [5] "MI | 14601533 / R: 1655   ->1690     |N:2  |B    |W    |B    |W    |B    |W    |B    |"
## [6] "OH | 15055204 / R: 1686   ->1687     |N:3  |W    |B    |W    |B    |B    |W    |B    |"
```

```
ID <- as.numeric(str_extract(m2, '\\d+'))
Name <- str_extract(m2, '[A-z].{1,32}')
Name <- str_trim(str_extract(Name, '.+\\s{2,}'))
State <- str_extract(m3, '[A-Z]{2}')
TotalNumberofPoints <- as.numeric(str_extract(m2, '\\d+\\.\\d'))
PreRating <- str_extract(m3, 'R:.{8,}-')
PreRating <- as.numeric(str_extract(PreRating, '\\d{1,4}'))
Rounds <- str_extract_all(m2, '[A-Z]\\s{2,}\\d+')
Rounds <- str_extract_all(Rounds, '\\d+')
```

```
## Warning in stri_extract_all_regex(string, pattern, simplify = simplify, :
## argument is not an atomic vector; coercing
```

```
avgPreRating <- c()

for(i in c(1:length(Rounds))){
  avgPreRating[i] <- round(mean(PreRating[as.numeric(Rounds[[i]])]), 0)
}
avgPreRating
```

```
##  [1] 1605 1469 1564 1574 1501 1519 1372 1468 1523 1554 1468 1506 1498 1515 1484
## [16] 1386 1499 1480 1426 1411 1470 1300 1214 1357 1363 1507 1222 1522 1314 1144
## [31] 1260 1379 1277 1375 1150 1388 1385 1539 1430 1391 1248 1150 1107 1327 1152
## [46] 1358 1392 1356 1286 1296 1356 1495 1345 1206 1406 1414 1363 1391 1319 1330
## [61] 1327 1186 1350 1263
```

```
output <- data.frame(ID,Name,State,TotalNumberofPoints,PreRating,avgPreRating)

head(output)
```

```
##   ID               Name State TotalNumberofPoints PreRating avgPreRating
## 1  1          GARY HUA    ON                 6.0      1794         1605
## 2  2   DAKSHESH DARURI    MI                 6.0      1553         1469
## 3  3      ADITYA BAJAJ    MI                 6.0      1384         1564
```

```
## 4  4 PATRICK H SCHILLING    MI                5.5      1716         1574
## 5  5          HANSHI ZUO    MI                5.5      1655         1501
## 6  6         HANSEN SONG    OH                5.0      1686         1519
```

```
write_csv(output, 'tournament.csv', append = FALSE)
```