Project 1

Benjamin Bravo

2026-02-20

```r
install.packages("stringr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```r
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```r
url <- "https://raw.githubusercontent.com/bb2955/Data-607/main/tournamentinfo.txt"
lines <- readLines(url)
```

```
## Warning in readLines(url): incomplete final line found on
## 'https://raw.githubusercontent.com/bb2955/Data-607/main/tournamentinfo.txt'
```

```r
head(lines)
```

```
## [1] "---------------------------------------------------------------------------"
## [2] " Pair | Player Name                     |Total|Round|Round|Round|Round|Round|Round|Round| "
## [3] " Num  | USCF ID / Rtg (Pre->Post)       | Pts |  1  |  2  |  3  |  4  |  5  |  6  |  7  | "
## [4] "---------------------------------------------------------------------------"
## [5] "    1 | GARY HUA                        |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|"
## [6] "   ON | 15445895 / R: 1794   ->1817     |N:2  |W    |B    |W    |B    |W    |B    |W    |"
```

```r
player_lines  <- lines[grep("^\\s*[0-9]+\\s+\\|", lines)]
rating_lines  <- lines[grep("R:", lines)]
```

```r
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
players <- data.frame(
  PairNum = as.numeric(str_trim(sapply(str_split(player_lines, "\\|"), `[`, 1))),
  Name = str_trim(sapply(str_split(player_lines, "\\|"), `[`, 2)),
  TotalPts = as.numeric(str_extract(player_lines, "\\d+\\.\\d")),
  stringsAsFactors = FALSE
```

```r
)

players$State <- str_trim(substr(rating_lines, 1, 2))

players$PreRating <- as.numeric(
  gsub(".*R:\\s*([0-9]+).*", "\\1", rating_lines)
)

get_opponents <- function(line) {
  matches <- str_extract_all(line, "[WLD]\\s+[0-9]+")[[1]]
  as.numeric(str_remove(matches, "[WLD]\\s+"))
}

opponent_list <- lapply(player_lines, get_opponents)

avg_opp_rating <- sapply(1:length(opponent_list), function(i) {
  opp_nums <- opponent_list[[i]]
  opp_ratings <- players$PreRating[match(opp_nums, players$PairNum)]
  mean(opp_ratings, na.rm = TRUE)
})

players$AvgOppPreRating <- round(avg_opp_rating)

final_df <- players %>%
  select(Name, State, TotalPts, PreRating, AvgOppPreRating)

head(final_df)
```

```
##                 Name State TotalPts PreRating AvgOppPreRating
## 1           GARY HUA              6.0      1794            1605
## 2      DAKSHESH DARURI           6.0      1553            1469
## 3         ADITYA BAJAJ           6.0      1384            1564
## 4 PATRICK H SCHILLING           5.5      1716            1574
## 5           HANSHI ZUO           5.5      1655            1501
## 6          HANSEN SONG           5.0      1686            1519
```