

CUNY Fall 2016 Class 607 Project 1

*Alex Low*

September 25, 2016

```
library(stringr)
```

I started by trying to create some sort of file from the original text file saving it as a csv file.

```
tournamentinfo <- read.csv("c:/tournamentinfo.csv")
test_row <- as.character(tournamentinfo)
```

Next I tried to extract each of the valid data elements as strings using their patterns from the file. Needless to say I'm only sharing the final code. There was a *huge* amount of trial and error!

```
player_states <- unlist(str_extract_all(test_row, "\\b[:upper:]{2}\\\\"))
player_states <- str_replace_all(player_states, pattern = "\\|", replacement = "")
player_names <- c(unlist(str_extract_all(test_row, "[[:alpha:]]+.[[:alpha:]]+.[[:alpha:]]+.[[:alpha:]]*"),
player_names <- player_names[-c(1)]
total_points <- c(unlist(str_extract_all(test_row, "[:digit:]\\.[[:digit:]]"))
total_points <- as.numeric(total_points)
pre_rating <- c(unlist(str_extract_all(test_row, "R:..[[:digit:]]{3,4}P?"))
pre_rating <- c(unlist(str_replace_all(pre_rating, pattern = "R: ", replacement = "")))
pre_rating <- c(unlist(str_replace_all(pre_rating, pattern = "P", replacement = "")))
pre_rating <- c(unlist(str_replace_all(pre_rating, pattern = "P", replacement = "")))
pre_rating <- as.numeric(pre_rating)
```

Of all the things I did, creating the average opponent score was the most difficult. I first found a way to extract all the player results and create a data frame. I next converted it into a matrix and then a data frame of those opponents ratings before calculating an average.

```
player_results <- c((str_extract_all(test_row, "([WDL]{1,4}[[:digit:]]{1,2}\\|)|([HBUX]{2,5}))){7}")
player_results <- c(unlist(str_replace_all(player_results, pattern = "[[:upper:]]", replacement = "")))
player_results <- c(unlist(str_replace_all(player_results, pattern = " ", replacement = "")))
player_results <- c(unlist(str_replace_all(player_results, pattern = ",$", replacement = "")))
player_results <- c(split(player_results, ceiling(seq_along(player_results)/8)))
df_trial <- data.frame(player_results)
df_trial <- t(df_trial)
opponent_rating <- c(pre_rating[(df_trial)])
opponent_rating_matrix <- matrix(data = opponent_rating, nrow = 8, ncol = 64)
opponent_rating_matrix <- t(opponent_rating_matrix)
opponent_rating_df <- data.frame(opponent_rating_matrix)
opponent_rating_df$avg <- rowMeans(opponent_rating_df, na.rm = TRUE)
```

Finally a bound all these values together into a data frame.

```
df_chess <- cbind.data.frame(player_names,player_states,total_points, round(pre_rating,opponent_rating_c  
write.csv(df_chess)
```

```
## "", "player_names", "player_states", "total_points", "round(pre_rating, opponent_rating_df$avg)"
```