

# Project 1

Chitrarth Kaushik

2/19/2020

```
library(stringr)
tournament <- readLines("https://raw.githubusercontent.com/chitrarth2018/607_Project_1/master/tournament.txt")
head(tournament)
```

```
## [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 |"
```

```
#replacing dash with empty character
tourn_wo_dash <- gsub("-", "", tournament)
head(tourn_wo_dash)
```

```
## [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 |"
```

```
#removing empty rows
tourn_clean <- tourn_wo_dash[sapply(tourn_wo_dash, nchar) > 0]
head(tourn_clean)
```

```
## [1] " Pair | Player Name |Total|Round|Round|Round|Round|Round|Round|Round| "
## [2] " Num | USCF ID / Rtg (Pre>Post) | Pts | 1 | 2 | 3 | 4 | 5 | 6 | 7 | "
## [3] " 1 | GARY HUA |6.0 |W 39|W 21|W 18|W 14|W 7|D 12|D 4|"
## [4] " ON | 15445895 / R: 1794 >1817 |N:2 |W |B |W |B |W |B |W |"
## [5] " 2 | DAKSHESH DARURI |6.0 |W 63|W 58|L 4|W 17|W 16|W 20|W 7|"
## [6] " MI | 14598900 / R: 1553 >1663 |N:2 |B |W |B |W |B |W |B |"
```

```
even_indexes<-seq(2,130,2)
odd_indexes<-seq(1,129,2)
player_det1<-tourn_clean[even_indexes]
player_det2<-tourn_clean[odd_indexes]
head(player_det1)
```

```
## [1] " Num | USCF ID / Rtg (Pre>Post) | Pts | 1 | 2 | 3 | 4 | 5 | 6 | 7 | "
## [2] " ON | 15445895 / R: 1794 >1817 |N:2 |W |B |W |B |W |B |W |"
## [3] " MI | 14598900 / R: 1553 >1663 |N:2 |B |W |B |W |B |W |B |"
## [4] " MI | 14959604 / R: 1384 >1640 |N:2 |W |B |W |B |W |B |W |"
## [5] " MI | 12616049 / R: 1716 >1744 |N:2 |W |B |W |B |W |B |B |"
## [6] " MI | 14601533 / R: 1655 >1690 |N:2 |B |W |B |W |B |W |B |"
```

```
head(player_det2)
```

```
## [1] " Pair | Player Name |Total|Round|Round|Round|Round|Round|Round|Round| "
## [2] " 1 | GARY HUA |6.0 |W 39|W 21|W 18|W 14|W 7|D 12|D 4|"
## [3] " 2 | DAKSHESH DARURI |6.0 |W 63|W 58|L 4|W 17|W 16|W 20|W 7|"
## [4] " 3 | ADITYA BAJAJ |6.0 |L 8|W 61|W 25|W 21|W 11|W 13|W 12|"
## [5] " 4 | PATRICK H SCHILLING |5.5 |W 23|D 28|W 2|W 26|D 5|W 19|D 1|"
## [6] " 5 | HANSHI ZUO |5.5 |W 45|W 37|D 12|D 13|D 4|W 14|W 17|"
```

```
#removing 1st row from both the tables
```

```
player_det1<-player_det1[-1]
```

```
player_det2<-player_det2[-1]
```

```
#extracting the player name
```

```
player_name<- gsub("\\\\| ", "", str_extract(player_det2, "\\| ([a-zA-Z])* ){1,}"))
```

```
head(player_name)
```

```
## [1] "GARY HUA" "DAKSHESH DARURI"
## [3] "ADITYA BAJAJ" "PATRICK H SCHILLING"
## [5] "HANSHI ZUO" "HANSEN SONG"
```

```
#extracting the player state
```

```
player_state<-str_extract(player_det1, "[a-zA-Z]{2}")
```

```
head(player_state)
```

```
## [1] "ON" "MI" "MI" "MI" "MI" "OH"
```

```
#creating the data frame
```

```
chess_results<-data.frame(player_name,player_state)
```

```
#extracting the player score
```

```
chess_results$player_score<-gsub("\\\\|"," ", str_extract(player_det2, "\\| ([0-9])\\.([0-9])"))
```

```
#extracting the player pre-rating
```

```
chess_results$pre_rating<- gsub(":", "", str_extract(player_det1, ":([0-9])"))
```

```
#counting the number of games played by a player
```

```
chess_results$num_games<-str_count(player_det2, "\\| ([a-zA-Z])([0-9]+)")
```

```
#extracting the opponent player numbers
```

```
player_opp<-gsub("\\| ([a-zA-Z])", "", str_extract_all(player_det2, "\\| ([a-zA-Z])([0-9]*)"))
```

```
player_opp<-str_extract_all(player_det2, "\\| ([a-zA-Z])([0-9]*)", simplify = TRUE)
```

```
#replacing the / and W, B etc.
```

```
for (i in 1:64){
```

```
  player_opp[i,]<-str_replace(player_opp[i,], "\\| ([a-zA-Z])", "")
```

```
}
```

```

#accessing the pre-rating of the opponents using the opponent number and the computing the mean pre-rating
for (i in 1:64){

chess_results$opp_R1[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,1])])
chess_results$opp_R2[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,2])])
chess_results$opp_R3[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,3])])
chess_results$opp_R4[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,4])])
chess_results$opp_R5[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,5])])
chess_results$opp_R6[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,6])])
chess_results$opp_R7[i]<-as.integer(chess_results$pre_rating[as.integer(player_opp[i,7])])

x<-c(chess_results$opp_R1[i],chess_results$opp_R2[i],chess_results$opp_R3[i],      chess_results$opp_R4[i],
chess_results$opp_R5[i],chess_results$opp_R6[i],chess_results$opp_R7[i])

chess_results$opp_rate_mean[i]<-round(mean(x, na.rm=TRUE),0)

}

#deleting unwanted columns
chess_results_final<-chess_results[,c(-5:-12)]
View(chess_results_final)

```