

DATA 607 - Project 1

Jose Mawyin

9/11/2019

```
library(magrittr)
```

```
raw_data1 <- read.delim("https://raw.githubusercontent.com/JMawyin/MSDS2019/master/tournamentinfo.txt",
head(raw_data1,4)
```

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

```
raw_data2 <- subset(raw_data1, Pair != "-----")
head(raw_data2,4)
```

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

```
raw.array.dimension <- dim(raw_data2)
endrow <- raw.array.dimension[1:1]
prune1 <- raw_data2[2:endrow, 1:10]
head(prune1,4)
```

```
##      Pair                                     Player.Name Total Round Round.1 Round.2
## 3 1      GARY HUA                          6.0  W  39   W  21   W  18
## 4 ON    15445895 / R: 1794  ->1817        N:2  W      B      W
## 6 2      DAKSHESH DARURI                    6.0  W  63   W  58   L   4
## 7 MI    14598900 / R: 1553  ->1663        N:2  B      W      B
```

```
##      Round.3 Round.4 Round.5 Round.6
## 3      W 14      W 7      D 12      D 4
## 4      B      W      B      W
## 6      W 17      W 16      W 20      W 7
## 7      W      B      W      B
```

*#Splitting cleaned data into two frames each of 64 rows. Each row contains
#different bits of player information.*

```
Odd.Row <- prune1[ c(TRUE,FALSE), ]
Even.Row <- prune1[!c(TRUE,FALSE), ]
```

```
head(Odd.Row,5) ;head(Even.Row,5)
```

```
##      Pair      Player.Name Total Round Round.1 Round.2
## 3      1      GARY HUA      6.0      W 39      W 21      W 18
## 6      2      DAKSHESH DARURI      6.0      W 63      W 58      L 4
## 9      3      ADITYA BAJAJ      6.0      L 8      W 61      W 25
## 12     4      PATRICK H SCHILLING      5.5      W 23      D 28      W 2
## 15     5      HANSHI ZUO      5.5      W 45      W 37      D 12
##      Round.3 Round.4 Round.5 Round.6
## 3      W 14      W 7      D 12      D 4
## 6      W 17      W 16      W 20      W 7
## 9      W 21      W 11      W 13      W 12
## 12     W 26      D 5      W 19      D 1
## 15     D 13      D 4      W 14      W 17
```

```
##      Pair      Player.Name Total Round Round.1 Round.2
## 4      ON      15445895 / R: 1794 ->1817      N:2      W      B      W
## 7      MI      14598900 / R: 1553 ->1663      N:2      B      W      B
## 10     MI      14959604 / R: 1384 ->1640      N:2      W      B      W
## 13     MI      12616049 / R: 1716 ->1744      N:2      W      B      W
## 16     MI      14601533 / R: 1655 ->1690      N:2      B      W      B
##      Round.3 Round.4 Round.5 Round.6
## 4      B      W      B      W
## 7      W      B      W      B
## 10     B      W      B      W
## 13     B      W      B      B
## 16     W      B      W      B
```

*#Code below separate the Player's Pre-Rating information from a string mixing
#integers and character.*

```
even.prune <- t(as.data.frame(strsplit(as.character(Even.Row$Player.Name), "\\->")))
row.names(even.prune) <- c()
```

```
even.prune2 <- t(as.data.frame(strsplit(as.character(even.prune[,1]), " / R: ")))
row.names(even.prune2) <- c()
```

```
Even.prune3 <- t(as.data.frame(strsplit(as.character(even.prune2[,2]), "P")))
row.names(Even.prune3) <- c()
head(Even.prune3,5)
```

```
##      [,1]      [,2]
```

```
## [1,] "1794" " " "1794" " "
## [2,] "1553" " " "1553" " "
## [3,] "1384" " " "1384" " "
## [4,] "1716" " " "1716" " "
## [5,] "1655" " " "1655" " "
```

```
Pre.Rating <- as.data.frame(as.numeric(trimws(Even.prune3[,1])))
```

```
#Code below separate players Total Number of Points and the opponents ID for each  
#of the 7 games played
```

```
Odd.prune <- as.data.frame(lapply(Odd.Row[,4:11], function(x) as.numeric(gsub("[a-zA-Z ]", "", x))))
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
row.names(Odd.prune) <- c()
head(Odd.prune,5)
```

```
##   Pair Player.Name Total Round Round.1 Round.2 Round.3 Round.4 Round.5
## 1    1          NA   6.0    39     21     18     14      7     12
## 2    2          NA   6.0    63     58      4     17     16     20
## 3    3          NA   6.0     8     61     25     21     11     13
## 4    4          NA   5.5    23     28      2     26      5     19
## 5    5          NA   5.5    45     37     12     13      4     14
##   Round.6
## 1         4
## 2         7
## 3        12
## 4         1
## 5        17
```

```
#dim(test4)
```

```
##Joining columns to form a new data frame containing all the information of  
#interest for us.
```

```
united <- as.data.frame(cbind(Odd.Row$Player.Name, Even.Row$Pair, Odd.prune$Total, Pre.Rating, Odd.prune$Round.6, Odd.prune$Round.7, Odd.prune$Round.12, Odd.prune$Round.1, Odd.prune$Round.2, Odd.prune$Round.3, Odd.prune$Round.4, Odd.prune$Round.5))
colnames(united) <- c("Name", "State", "Total Points", "Pre-Rating", "Oppo1", "Oppo2", "Oppo3", "Oppo4", "Oppo5", "Oppo6", "Oppo7")
head(united,4)
```

```
##           Name State Total Points Pre-Rating Oppo1
## 1  GARY HUA      ON     6.0    1794      39
## 2 DAKSHESH DARURI  MI     6.0    1553      63
## 3 ADITYA BAJAJ     MI     6.0    1384       8
## 4 PATRICK H SCHILLING MI     5.5    1716      23
##   Oppo2 Oppo3 Oppo4 Oppo5 Oppo6 Oppo7
## 1    21    18    14     7    12     4
## 2    58     4    17    16    20     7
## 3    61    25    21    11    13    12
## 4    28     2    26     5    19     1
```

```
write.csv(united, file = "/Users/josemawyin/Library/Mobile Documents/com~apple~CloudDocs/Data Science M
```

```
Average.Rating <- 0
```

```
Average.Rating <- c(united[as.integer(united[1,5]),4] , united[as.integer(united[1,6]),4] , united[as.in  
str(Average.Rating)
```

```
## num 1605
```

```
Average.Rating <- c()
```

```
for(i in 1:64) {  
Average.Rating <- rbind(Average.Rating, mean( c(united[as.integer(united[i,5]),4] , united[as.integer(u  
}  
head(Average.Rating,4)
```

```
##           [,1]  
## [1,] 1605.286  
## [2,] 1469.286  
## [3,] 1563.571  
## [4,] 1573.571
```

```
str(Average.Rating)
```

```
## num [1:64, 1] 1605 1469 1564 1574 1501 ...
```

```
Final.Table <- cbind(united[,1:4], round(Average.Rating,0))  
colnames(Final.Table) <- c("Player Name", "State", "Total Points", "Pre-Rating", "Average Pre-Rating of  
head(Final.Table,4)
```

```
##           Player Name State Total Points Pre-Rating  
## 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  
## Average Pre-Rating of Opponents  
## 1 1605  
## 2 1469  
## 3 1564  
## 4 1574
```

```
write.csv(Final.Table, file = "/Users/josemawyin/Library/Mobile Documents/com~apple~CloudDocs/Data Scien
```