IS607: PROJECT 1

MUSA T. GANIYU

February 21, 2016

```
# Load data into R, and require package is "stringr"
library(stringr);
library(knitr);
library(ggplot2)

tournamentinfo1 <- read.table("https://raw.githubusercontent.com/mascotinme/MSDA-IS607/master/tournamen

head(tournamentinfo1);
```

```
##   X.....................................................................................
## 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  -----------------------------------------------------------------------------------------
## 4     1 | GARY HUA                       |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|
## 5    ON | 15445895 / R: 1794   ->1817    |N:2  |W    |B    |W    |B    |W    |B    |W    |
## 6  -----------------------------------------------------------------------------------------
```

```
str(tournamentinfo1);
```

```
## 'data.frame':    195 obs. of  1 variable:
##  $ X.....................................................................................: Facto
```

```
# Setting Working Direrctory, and read the line (Because its a .txt file)

setwd("C:/Data");
data <- readLines("tournamentinfo.txt");
```

```
## Warning in readLines("tournamentinfo.txt"): incomplete final line found on
## 'tournamentinfo.txt'
```

```
str(data);
```

```
##  chr [1:196] "-----------------------------------------------------------------------------------------
```

```
#A glance at the data shows we have to extract the vector on after the other and latter combined them t



states <- str_trim(unlist(str_extract_all(data, " ON | OH | MI ")))
str(states);
```

```
##  chr [1:64] "ON" "MI" "MI" "MI" "MI" "OH" "MI" "MI" ...
```

```r
outcomes <- unlist(str_extract_all(data, "[:digit:][//.][:digit:]"))
str(outcomes);
```

```
##  chr [1:64] "6.0" "6.0" "6.0" "5.5" "5.5" "5.0" "5.0" ...
```

```r
# Again, we extract all aphabets, integers unlist it and then trim them to get the outcomes, names, bef

names <- (str_trim(unlist(str_extract_all(data, "([[:alpha:] ]-?){15,31}"))))[2:65]
head(names);
```

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

```r
before_the_rating <- as.integer(sub(pattern = 'R: ', replacement = '', x = unlist(str_extract_all(data,

str(before_the_rating);
```

```
##  int [1:64] 1794 1553 1384 1716 1655 1686 1649 1641 1411 1365 ...
```

```r
opponent <- gsub("\\|", " ", str_sub(unlist(str_extract_all(data, "[:digit:][//.][:digit:][|DLWUXBH[:dig
str(opponent);
```

```
##  chr [1:64] "39 W  21 W  18 W  14 W   7 D  12 D   4" ...
```

```r
#lists the opponent based on "W|D|U|H|B|X|L" as | means 'OR' by splitting.



list_of_opponent <- (strsplit(opponent, " W  | D  | U  | H  | B  | X  | L  "));
head(list_of_opponent);
```

```
## [[1]]
## [1] "39" "21" "18" "14" " 7" "12" " 4"
##
## [[2]]
## [1] "63" "58" " 4" "17" "16" "20" " 7"
##
## [[3]]
## [1] " 8" "61" "25" "21" "11" "13" "12"
##
## [[4]]
## [1] "23" "28" " 2" "26" " 5" "19" " 1"
##
## [[5]]
## [1] "45" "37" "12" "13" " 4" "14" "17"
##
## [[6]]
## [1] "34" "29" "11" "35" "10" "27" "21"
```

```
matrix_1 <- sapply(list_of_opponent, as.numeric);

#We convert the the "before_the_rating" to matrix on 7 rows.

matrix_2 <- matrix(before_the_rating[matrix_1], nrow = 7)

#We convert opposition matrix to integer and obtain it averages/means,


opponent_average <- as.integer(format(apply(matrix_2, 2, mean, na.rm = TRUE), digits = 4))
str(opponent_average)
```

```
##  int [1:64] 1605 1469 1564 1574 1501 1519 1372 1468 1523 1554 ...
```

```
#The final tournament outcomes as requested in the project question.


final_tournament_outcomes <- data.frame(names, states, outcomes, before_the_rating, opponent_average);
str(final_tournament_outcomes);
```

```
## 'data.frame':    64 obs. of  5 variables:
##  $ names            : Factor w/ 64 levels "ADITYA BAJAJ",..: 24 12 1 51 28 27 23 21 59 5 ...
##  $ states           : Factor w/ 3 levels "MI","OH","ON": 3 1 1 1 1 2 1 1 3 1 ...
##  $ outcomes         : Factor w/ 11 levels "1.0","1.5","2.0",..: 11 11 11 10 10 9 9 9 9 9 ...
##  $ before_the_rating: int  1794 1553 1384 1716 1655 1686 1649 1641 1411 1365 ...
##  $ opponent_average : int  1605 1469 1564 1574 1501 1519 1372 1468 1523 1554 ...
```

```
kable(head(final_tournament_outcomes));
```

names	states	outcomes	before_the_rating	opponent_average
GARY HUA	ON	6.0	1794	1605
DAKSHESH DARURI	MI	6.0	1553	1469
ADITYA BAJAJ	MI	6.0	1384	1564
PATRICK H SCHILLING	MI	5.5	1716	1574
HANSHI ZUO	MI	5.5	1655	1501
HANSEN SONG	OH	5.0	1686	1519

```
diff <- (final_tournament_outcomes$before_the_rating - final_tournament_outcomes$opponent_average);

diff;
```

```
##  [1]  189   84 -180  142  154  167  277  173 -112 -189  244  157  168   95
## [15] -264  218  130  120  138  184   93  255  149 -128  382   72  330  -15
## [29]  288  378  234   62  172   24  288  -33 -405 -116    6  -43  155  182
## [43]  176 -128   90 -981  -30   26    5 -240 -345 -560   48   64 -220 -261
## [57] -271 -474 -466 -363 -372  344 -175 -100
```

```r
# .csv file that could be imported into a SQL databa

tournament_csv = read.delim("tournamentinfo.txt")
write.table(tournament_csv, file="tournamentinfo.csv",sep=",",col.names=FALSE,row.names=FALSE)

head(tournament_csv);
```

```
##   X.....................................................................................
## 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  -----------------------------------------------------------------------------------------
## 4     1 | GARY HUA                       |6.0  |W  39|W  21|W  18|W  14|W   7|D  12|D   4|
## 5    ON | 15445895 / R: 1794   ->1817    |N:2  |W    |B    |W    |B    |W    |B    |W    |
## 6  -----------------------------------------------------------------------------------------
```

THANKS.