PROJECT 1 DATA607

OMER OZEREN

February 23, 2019

Table of Contents

Get Data from Github

```
raw_data <-
read.csv(paste0("https://raw.githubusercontent.com/omerozeren/DATA607/master/
PROJECT_1/tournamentinfo.txt"))
head(raw_data)
```

```
##
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  ------------------------------------------------------------------------
------------------
```

Removing Column Headers

```
raw_data <- raw_data[-c(1:2),]
head(raw_data,6)
```

```
## [1] ----------------------------------------------------------------------
------------------
## [2]     1 | GARY HUA                          |6.0  |W  39|W  21|W  18|W
14|W   7|D  12|D   4|
## [3]    ON | 15445895 / R: 1794   ->1817      |N:2  |W    |B    |W    |B
|W    |B    |W    |
## [4] ----------------------------------------------------------------------
------------------
## [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    |
## 131 Levels: ----------------------------------------------------------------
-------------------------- ...
```

Cleaning the Data

```
clean_data =raw_data[seq(2,length(raw_data),3)]
knitr::kable(head(clean_data,5), digits=2, align=c(rep("l", 4) ) )
```

x
1 | GARY HUA |6.0 |W 39|W 21|W 18|W 14|W 7|D 12|D 4|
2 | DAKSHESH DARURI |6.0 |W 63|W 58|L 4|W 17|W 16|W 20|W 7|
3 | ADITYA BAJAJ |6.0 |L 8|W 61|W 25|W 21|W 11|W 13|W 12|
4 | PATRICK H SCHILLING |5.5 |W 23|D 28|W 2|W 26|D 5|W 19|D 1|
5 | HANSHI ZUO |5.5 |W 45|W 37|D 12|D 13|D 4|W 14|W 17|

Players ID

```
id <-
as.numeric(str_extract(substr(unlist(clean_data),1,length(raw_data)),"\\d+"))
id
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
## [47] 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
```

Player Names

```
names <- str_trim(unlist(str_extract_all(raw_data, "[A-Z]+ [A-Z]+ ([A-Z-]+)?
([A-Z]+)?")))
names
```

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

```
##  [9] "STEFANO LEE"                "ANVIT RAO"
## [11] "CAMERON WILLIAM MC LEMAN"   "KENNETH J TACK"
## [13] "TORRANCE HENRY JR"          "BRADLEY SHAW"
## [15] "ZACHARY JAMES HOUGHTON"     "MIKE NIKITIN"
## [17] "RONALD GRZEGORCZYK"         "DAVID SUNDEEN"
## [19] "DIPANKAR ROY"               "JASON ZHENG"
## [21] "DINH DANG BUI"              "EUGENE L MCCLURE"
## [23] "ALAN BUI"                   "MICHAEL R ALDRICH"
## [25] "LOREN SCHWIEBERT"           "MAX ZHU"
## [27] "GAURAV GIDWANI"             "SOFIA ADINA STANESCU-BELLU"
## [29] "CHIEDOZIE OKORIE"           "GEORGE AVERY JONES"
## [31] "RISHI SHETTY"               "JOSHUA PHILIP MATHEWS"
## [33] "JADE GE"                    "MICHAEL JEFFERY THOMAS"
## [35] "JOSHUA DAVID LEE"           "SIDDHARTH JHA"
## [37] "AMIYATOSH PWNANANDAM"       "BRIAN LIU"
## [39] "JOEL R HENDON"              "FOREST ZHANG"
## [41] "KYLE WILLIAM MURPHY"        "JARED GE"
## [43] "ROBERT GLEN VASEY"          "JUSTIN D SCHILLING"
## [45] "DEREK YAN"                  "JACOB ALEXANDER LAVALLEY"
## [47] "ERIC WRIGHT"                "DANIEL KHAIN"
## [49] "MICHAEL J MARTIN"           "SHIVAM JHA"
## [51] "TEJAS AYYAGARI"             "ETHAN GUO"
## [53] "JOSE C YBARRA"              "LARRY HODGE"
## [55] "ALEX KONG"                  "MARISA RICCI"
## [57] "MICHAEL LU"                 "VIRAJ MOHILE"
## [59] "SEAN M MC CORMICK"          "JULIA SHEN"
## [61] "JEZZEL FARKAS"              "ASHWIN BALAJI"
## [63] "THOMAS JOSEPH HOSMER"       "BEN LI"
```

Player States

```
region <- unlist(str_extract_all(raw_data, "[[:upper:]]. \\|"))
region <- str_trim(str_replace_all(region, " \\|", ""))
region
```

```
##  [1] "ON" "MI" "MI" "MI" "MI" "OH" "MI" "MI" "ON" "MI" "MI" "MI" "MI" "MI"
## [15] "MI" "MI" "MI" "MI" "MI" "MI" "ON" "MI" "ON" "MI" "MI" "ON" "MI" "MI"
## [29] "MI" "ON" "MI" "ON" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI"
## [43] "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI"
## [57] "MI" "MI" "MI" "MI" "ON" "MI" "MI" "MI"
```

Player Points

```
points <- as.numeric(unlist(str_extract_all(raw_data, "(\\d)\\.(\\d)")))
points
```

```
##  [1] 6.0 6.0 6.0 5.5 5.5 5.0 5.0 5.0 5.0 5.0 4.5 4.5 4.5 4.5 4.5 4.0 4.0
## [18] 4.0 4.0 4.0 4.0 4.0 4.0 4.0 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.5
## [35] 3.5 3.5 3.5 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 2.5 2.5 2.5 2.5 2.5
## [52] 2.5 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 1.5 1.5 1.0 1.0 1.0
```

Player Pre Ratings

```
pre_rating  <- unlist(str_extract_all(raw_data, "R:\\s+?(\\d)+"))
pre_rating  <- str_replace_all(pre_rating , "R:\\s+", "")
pre_rating  <- as.integer(pre_rating )
pre_rating
```

```
##  [1] 1794 1553 1384 1716 1655 1686 1649 1641 1411 1365 1712 1663 1666 1610
## [15] 1220 1604 1629 1600 1564 1595 1563 1555 1363 1229 1745 1579 1552 1507
## [29] 1602 1522 1494 1441 1449 1399 1438 1355  980 1423 1436 1348 1403 1332
## [43] 1283 1199 1242  377 1362 1382 1291 1056 1011  935 1393 1270 1186 1153
## [57] 1092  917  853  967  955 1530 1175 1163
```

Player Opponents

```
opponents <- str_extract_all(str_extract_all(clean_data, "\\d+\\|"), "\\d+")
head(opponents,5)
```

```
## [[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"
```

Create DataFrame

```
df <- data.frame(id, names, region, points, pre_rating)
colnames(df)<- c("Player Number","Player Names","Player State","Total
Points","Pre Rating")
df[,"Average_Pre_Chess_Rating"] <- NA

for (i in 1:nrow(df)){
  df$Average_Pre_Chess_Rating[i] <-
round(mean(pre_rating[as.integer(unlist(opponents[id[i]]))]), digits = 0)
}
knitr::kable(df, digits=2, align=c(rep("l", 4) ) )
```

Player Number	Player Names	Player State	Total Points	Pre Rating	Average_Pre_Chess_Rating
1	GARY HUA	ON	6.0	1794	1605
2	DAKSHESH DARURI	MI	6.0	1553	1469
3	ADITYA BAJAJ	MI	6.0	1384	1564

4	PATRICK H SCHILLING	MI	5.5	1716	1574
5	HANSHI ZUO	MI	5.5	1655	1501
6	HANSEN SONG	OH	5.0	1686	1519
7	GARY DEE SWATHELL	MI	5.0	1649	1372
8	EZEKIEL HOUGHTON	MI	5.0	1641	1468
9	STEFANO LEE	ON	5.0	1411	1523
10	ANVIT RAO	MI	5.0	1365	1554
11	CAMERON WILLIAM MC LEMAN	MI	4.5	1712	1468
12	KENNETH J TACK	MI	4.5	1663	1506
13	TORRANCE HENRY JR	MI	4.5	1666	1498
14	BRADLEY SHAW	MI	4.5	1610	1515
15	ZACHARY JAMES HOUGHTON	MI	4.5	1220	1484
16	MIKE NIKITIN	MI	4.0	1604	1386
17	RONALD GRZEGORCZYK	MI	4.0	1629	1499
18	DAVID SUNDEEN	MI	4.0	1600	1480
19	DIPANKAR ROY	MI	4.0	1564	1426
20	JASON ZHENG	MI	4.0	1595	1411
21	DINH DANG BUI	ON	4.0	1563	1470
22	EUGENE L MCCLURE	MI	4.0	1555	1300
23	ALAN BUI	ON	4.0	1363	1214
24	MICHAEL R ALDRICH	MI	4.0	1229	1357
25	LOREN SCHWIEBERT	MI	3.5	1745	1363
26	MAX ZHU	ON	3.5	1579	1507
27	GAURAV GIDWANI	MI	3.5	1552	1222
28	SOFIA ADINA STANESCU-BELLU	MI	3.5	1507	1522
29	CHIEDOZIE OKORIE	MI	3.5	1602	1314
30	GEORGE AVERY	ON	3.5	1522	1144

	JONES				
31	RISHI SHETTY	MI	3.5	1494	1260
32	JOSHUA PHILIP MATHEWS	ON	3.5	1441	1379
33	JADE GE	MI	3.5	1449	1277
34	MICHAEL JEFFERY THOMAS	MI	3.5	1399	1375
35	JOSHUA DAVID LEE	MI	3.5	1438	1150
36	SIDDHARTH JHA	MI	3.5	1355	1388
37	AMIYATOSH PWNANANDAM	MI	3.5	980	1385
38	BRIAN LIU	MI	3.0	1423	1539
39	JOEL R HENDON	MI	3.0	1436	1430
40	FOREST ZHANG	MI	3.0	1348	1391
41	KYLE WILLIAM MURPHY	MI	3.0	1403	1248
42	JARED GE	MI	3.0	1332	1150
43	ROBERT GLEN VASEY	MI	3.0	1283	1107
44	JUSTIN D SCHILLING	MI	3.0	1199	1327
45	DEREK YAN	MI	3.0	1242	1152
46	JACOB ALEXANDER LAVALLEY	MI	3.0	377	1358
47	ERIC WRIGHT	MI	2.5	1362	1392
48	DANIEL KHAIN	MI	2.5	1382	1356
49	MICHAEL J MARTIN	MI	2.5	1291	1286
50	SHIVAM JHA	MI	2.5	1056	1296
51	TEJAS AYYAGARI	MI	2.5	1011	1356
52	ETHAN GUO	MI	2.5	935	1495
53	JOSE C YBARRA	MI	2.0	1393	1345
54	LARRY HODGE	MI	2.0	1270	1206
55	ALEX KONG	MI	2.0	1186	1406
56	MARISA RICCI	MI	2.0	1153	1414
57	MICHAEL LU	MI	2.0	1092	1363
58	VIRAJ MOHILE	MI	2.0	917	1391
59	SEAN M MC CORMICK	MI	2.0	853	1319

60	JULIA SHEN	MI	1.5	967	1330
61	JEZZEL FARKAS	ON	1.5	955	1327
62	ASHWIN BALAJI	MI	1.0	1530	1186
63	THOMAS JOSEPH HOSMER	MI	1.0	1175	1350
64	BEN LI	MI	1.0	1163	1263

Writing results in csv file

```
write.csv(df, file = "tournament_results.csv")
```