Data 607 Project 1

Dilip Ganesan

02/24/2016

Project 1 : Chess Data Transformation.

Introduction.

This Project consist of text file which contains the rankings and players information of a Chess Tournament. The data available should be processed and displayed in the format specified in the project.

Approach

The entire data is divided in to two data frames.

1. *With First data frame contains the player related information.*
2. *Second data frame contains the ranking related information.*
3. *Final result is displayed using the two database as two relational database tables.*

Step 1: Loading the text file and cleansing unwanted elements.

```
datafile=readLines("tournamentinfo.txt")
#Cleaning the unwanted "-----"
datafile=str_replace_all(datafile,"--------------------------------------------------------------------
```

Step 2: Creating a dataframe and loading that file to data frame, cleansing empty rows in dataset.

```
#Creating a dataframe and loading that file to data frame.
df=data.frame()
df=data.frame(datafile)

#Cleaning the unwanted empty rows for clear data set.
df[df==""] <- NA
df = df %>% na.omit()
```

Step 3: Creating 1st Data Frame with Player Information.

```
# 1st Data.Frame.
df5=data.frame()
toDelete <- seq(1, nrow(df), 2)
df5<-data.frame(df[ toDelete ,])
colnames(df5)
```

```
## [1] "df.toDelete..."
```

```
df5=separate(data = df5, col = df.toDelete..., into = c('Pair','Player Name','Total','R1','R2','R3','R4
df5 <- df5[-c(1), ]
```

Step 4: Creating 2nd Data Frame with Player Ranking Information

```
# 2nd Data.Frame.
toDelete2 <- seq(2, nrow(df), 2)
df2<-data.frame(df[ toDelete2 ,])
colnames(df2)
```

```
## [1] "df.toDelete2..."
```

```
df3=data.frame()
df3=separate(data = df2, col = df.toDelete2..., into = c('State','USCF','Pts','1','2','3','4','5','6','
df3<-df3[-c(1),]
```

Step 4: This is a function to calculate Average of Opponents Pre Ratings.

```
# This method is used to calculate the Average Pre Rating for Opponents.

avgrat<-as.vector(c())
for (i in 1:nrow(df5)){
  oc<-str_extract_all(df5[i,],"\\b\\d{1,}")
  oc<-as.matrix(oc[-1:-3])
  oc<-as.matrix(oc[lapply(oc,length)>0])
  total=0
  k=0
  j=0
  for(row in 1:nrow(oc)){
    k=oc[row]
    presub2<- str_extract(df3[as.numeric(k),]$USCF, ': *\\d{2,}')
    presub2<- str_extract(presub2, '\\d{2,}')
    total=total+as.numeric(presub2)
    j=j+1
  }
  avg=total/j
  print(round(avg))
  avgrat[length(avgrat)+1]<-round(avg)
}
```

```
## [1] 1605
## [1] 1469
## [1] 1564
## [1] 1574
## [1] 1501
## [1] 1519
## [1] 1372
## [1] 1468
## [1] 1523
## [1] 1554
## [1] 1468
## [1] 1506
## [1] 1498
## [1] 1515
## [1] 1484
## [1] 1386
## [1] 1499
## [1] 1480
## [1] 1426
## [1] 1411
## [1] 1470
## [1] 1300
## [1] 1214
## [1] 1357
## [1] 1363
## [1] 1507
## [1] 1222
## [1] 1522
## [1] 1314
## [1] 1144
## [1] 1260
## [1] 1379
## [1] 1277
## [1] 1375
## [1] 1150
## [1] 1388
## [1] 1385
## [1] 1539
## [1] 1430
## [1] 1391
## [1] 1248
## [1] 1150
## [1] 1107
## [1] 1327
## [1] 1152
## [1] 1358
## [1] 1392
## [1] 1356
## [1] 1286
## [1] 1296
## [1] 1356
## [1] 1495
## [1] 1345
## [1] 1206
## [1] 1406
## [1] 1414
## [1] 1363
## [1] 1391
## [1] 1319
## [1] 1330
## [1] 1327
## [1] 1186
## [1] 1350
## [1] 1263
```

Step 5: Final Data Cleaning and Resultant Data frame.

```r
# Final Data Cleansing.
prerating <- str_extract(df3$USCF, ': *\\d{2,}')

#Resultant Data Frame.
df6<-data.frame()
df6=data.frame(PlayerName=df5$`Player Name` ,State=df3$State ,TotalPoints=df5$Total ,
                PreRating=as.integer(str_extract(prerating, '\\d{2,}')), AvgOppPreRating=avgrat)

write.csv(df6, "ChessRankingData.csv", row.names=FALSE)
```

References StackOverflow for figuring the Seperate method to split a column in data frame.