

Week 4 Assignment

September 25, 2016

```
library(data.table)
library(stringr)
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(plyr)
```

3. Your job is to create an R Markdown file that generates a .CSV file (that could for example be imported into a SQL database)

with the following information for all of the players:

Player's Name, Player's State, Total Number of Points, Player's Pre-Rating, and Average Pre Chess Rating of Opponents

```
#set url
tourneyUrl <- getURL('https://raw.githubusercontent.com/FlOwControl/Data607/master/Proj1/tournamentinfo')

#remove the hyphens
parsed <- gsub('-', '', tourneyUrl)

#create the data frame
tournamentData <- read.table(textConnection(parsed), header = TRUE, sep="|", allowEscapes=TRUE)

#pull out the rows with the wins, rank, country info
metaDataRows1<- seq(1, nrow(tournamentData), by=2)
playerMetaData1 <- data.frame(tournamentData[metaDataRows1, ])

#pull out the rows with the player name, points, round score
metaDataRows2<- seq(2, nrow(tournamentData), by=2)
playerMetaData2 <- tournamentData[metaDataRows2, ]

# remove the header row
playerMetaData1<-playerMetaData1[-1,]
```

Once the data manipulation is done, we can create a fresh data frame with only the variables we want to keep

```
# create a new data frame with all the information in a new column
playerInfo <-data.frame(
  playerMetaData2$Player.Name,
  playerMetaData1$Pair,
  playerMetaData2$Total,
  playerMetaData1$Player.Name,
  playerMetaData2$Round,
```

```

playerMetaData2$Round.1,
playerMetaData2$Round.2,
playerMetaData2$Round.3,
playerMetaData2$Round.4,
playerMetaData2$Round.5,
playerMetaData2$Round.6
)

#rename the columns
playerInfo <- rename(playerInfo, c(
  "playerMetaData2.Player.Name"="playerName",
  "playerMetaData1.Pair"="playerState",
  "playerMetaData2.Total" = "playerTotal",
  "playerMetaData1.Player.Name"="playerPreRating",
  "playerMetaData2.Round"="round1",
  "playerMetaData2.Round.1"="round2",
  "playerMetaData2.Round.2"="round3",
  "playerMetaData2.Round.3"="round4",
  "playerMetaData2.Round.4"="round5",
  "playerMetaData2.Round.5"="round6",
  "playerMetaData2.Round.6"="round7"
))

#strip out the characters in the preRating columns
playerInfo$playerPreRating <- str_trim(unlist(str_replace_all(playerInfo$playerPreRating,
pattern="[:digit:]{8}[:,\\/:[:blank:]]R\\:[:,blank:]*", replacement='')))

playerInfo$playerPreRating <- str_trim(unlist(str_replace_all(playerInfo$playerPreRating,
pattern="([[:blank:]]?([[:alpha:]]*)\\>([[:alpha:]])?.*[:digit:]*", replacement="")))

playerInfo$playerPreRating <- str_trim(unlist(str_replace_all(playerInfo$playerPreRating,
pattern="[:upper:] [:digit:]{1,3}$", replacement="")))

removeCharFromRounds <- function(column){
  column <- str_trim(unlist(str_replace_all(column,
pattern="[L|W|X|U|D|H|B] [:blank:]*", replacement="")))
  return (column)
}

#strip out the characters in each round column
playerInfo$round1 <- removeCharFromRounds(playerInfo$round1)
playerInfo$round2 <- removeCharFromRounds(playerInfo$round2)
playerInfo$round3 <- removeCharFromRounds(playerInfo$round3)
playerInfo$round4 <- removeCharFromRounds(playerInfo$round4)
playerInfo$round5 <- removeCharFromRounds(playerInfo$round5)
playerInfo$round6 <- removeCharFromRounds(playerInfo$round6)
playerInfo$round7 <- removeCharFromRounds(playerInfo$round7)

```

Once the data has been massaged, we want to loop through it to gain the information we want. Basically, we are going to loop through the columns, pull out the data, and then perform calculations on it to find the averages of each player's opponent.

```

#create an empty vector to hold the new rating field
#loop through the rows in the playerInfo data frame
#loop through the 5-11 columns, pulling out the information for opponent rows
#create a variable to hold the opponent, create the rating
#add the rating to the playerInfo dataFrame. As it is a double, need to round up to 0 decimal places
#remove the NA values by setting na.rm to true
rating <- c()
for(i in 1:nrow(playerInfo)){
  for(j in 5:ncol(playerInfo)){
    opponent<-playerInfo[i, j]
    rating <-c(rating,as.numeric(playerInfo[opponent, 4]))
  }
  playerInfo[i, 'AvgOpponentPreChessRating'] <- round(mean(rating, na.rm=TRUE),digits=0)
}

#create a dataframe with only the required info, by removing the round info
playerInfo <- subset(playerInfo, select = -c( 5 : 11 ))

#create a csv of the data frame with a name, and keep the row names
write.csv(playerInfo, 'chessCrossTable.csv', row.names=T)

```