

Project1_607

Alvaro Bueno

9/22/2017

build the data frames

extract the data from the file

```
library('stringr')
tmp <- url('https://github.com/delagroove/dataScience/raw/master/tournamentinfo.txt')
open(tmp, 'r')
count = 0
header <- readLines(tmp, n=4)
trim <- function(x) gsub("^\\s+|\\s+$", "", x)

while (length(data <- readLines(tmp, n=3)) > 0){
  count = count + 1
  #extract id
  id <- unlist(str_extract_all(data[1], "(?<= {3,4})[0-9]{1,2}(?= {1}| [a-zA-Z ]{6,})", FALSE))

  name <- unlist(str_extract_all(data[1], "(?<=[0-9]{1,2} {1}| ) [a-zA-Z ]{6,}", FALSE))
  name <- trim(name)

  #extract states
  state <- unlist(str_extract_all(data[2], "[a-zA-Z]{2}(?= | [0-9]{8})", FALSE))

  #extract points
  points <- unlist(str_extract_all(data[1], "(?<=[0-9]{1,2} {1}| [a-zA-Z ]{6,32}|)[0-9].[0-9]", FALSE))

  #extract pre-score
  pres <- unlist(str_extract_all(data[2], "(?<= R: {1,2})[0-9]{3,4}(?= |P)", FALSE))
  df <- data.frame(id,name,state,points,pres, 0)
  names(df) <- c('id', 'name', 'state', 'points', 'prescore', 'calcscore')
  results <- rbind(results, df)

  #extract opponents
  opp <- unlist(str_extract_all(data[1], "(?<=[0-9]{1,2} {1}| [a-zA-Z ]{6,32}|[0-9].[0-9] | (W|D|L) {", FALSE))

  df2 <- data.frame(id, opp)
  names(df2) <- c('id', 'opponents')
  opponents <- rbind(opponents, df2)
}
```

```
## Warning in readLines(tmp, n = 3): incomplete final line found on 'https://
## github.com/delagroove/dataScience/raw/master/tournamentinfo.txt'
```

Fill out the calculated store

```
for (id in 1:max(as.numeric(opponents$id))){
  sum <- 0
  for(opponent in opponents[opponents$id==id,'opponents']){
    pr <- results[results$id==opponent,'prescore']
    num_pr <- as.numeric(levels(pr))[pr]
    sum = sum + num_pr
  }

  calcscore = sum/length(opponents[opponents$id==id,'opponents'])
  results[results$id==id,'calcscore'] <- calcscore
}

close(tmp)
View(results)

#write results to CSV
write.csv(results, file = "results.csv")
```