lottoDataFrame

Chris Richardson

Oct 11, 2018

```
library(stringi)
library(stringr)
```

```
substrRight <- function(x, n){
  substr(x, nchar(x)-n+1, nchar(x))
}
```

```
# Main Clean -----
df <-  read.table('C:/Users/crich/Downloads/DownloadAllNumbers.txt', sep = '\t')
```

```
# REMOVE FIRST 3 UNNECESSARY ROWS
lotto <-  data.frame(df[-(1:3),])

# Rename column
colnames(lotto) <-  'Record'

# remove white space
lotto$strip <-  gsub("\\s", "", lotto$Record)

lotto$digitsOnly <-  gsub('\\D','', lotto$strip)

# this is built on the constant variable of digits always ending in order 'Serial #, date, and winning
lotto$Serial <-  as.integer(substr(lotto$digitsOnly, 1, nchar(lotto$digitsOnly)-9))
```

```
# Extract DaY as %b%d%Y
lotto$stringDate <- str_extract(lotto$Record, '(\\D{9}).+(\\d{2})')
lotto$stringDate <- substring(lotto$stringDate, 10)

# Remove whitespace from Date
lotto$stringDate <- gsub("\\s", "", lotto$stringDate)

# insert , between month & day for regex purposes
lotto$stringDate <-  sub('^(.{3})(.*)', "\\1,\\2", lotto$stringDate)

lotto$stringDate <-  gsub(",", "/", lotto$stringDate)

# convert stringDate into international date standard
lotto$Date <-  as.Date(lotto$stringDate, format = '%b/%d/%Y' )

# grab string containing date
lotto$Day <-  gsub('\\d','', lotto$strip)
# grab exact day
lotto$Day <-  stri_sub(lotto$Day, 1, 3)

# label rows according to draw time
lotto$Time[lotto$Serial %% 2 == 1] <-  'Evening'
lotto$Time[lotto$Serial %% 2 == 0] <- 'Midday'
```

```r
# grab winning #s
lotto$Winning <-  as.character(substrRight(lotto$strip,3))

lotto <-  lotto[,c(1,4,6:9)]

lotto$One <- as.integer(stri_sub(lotto$Winning, 1, -3))
lotto$Two <-  as.integer(stri_sub(lotto$Winning, 2, -2))
lotto$Three <-  as.integer(stri_sub(lotto$Winning, 3, -1))

# column for the total sum of winning numbers
lotto$WinningSum <- rowSums(lotto[,7:9])
```

```r
lottoExtended = lotto

# split winning numbers into two columns consisting of 1&2 and 2&3
lottoExtended$FrontTwo <- stri_sub(lotto$Winning, 1,2)
lottoExtended$LastTwo <-  stri_sub(lotto$Winning,2,3)

lottoExtended$FrontTwoSum <-  lotto$One + lotto$Two
lottoExtended$LastTwoSum <- lotto$Two + lotto$Three
```

```r
write.table(lotto, file = 'C:/Users/crich/Desktop/Googz/lottoResearch/Data/cleanedLotto.csv', sep = ','

# contains 3
write.table(lottoExtended, file = 'C:/Users/crich/Desktop/Googz/lottoResearch/Data/cleanedLottoExtended

# tableau friendly}
lottoExp = lotto[,c(2:9)]
write.table(lottoExp, file = 'C:/Users/crich/Desktop/Googz/lottoResearch/Data/cleanedLottoTabl.csv', se
rm(lottoExp)
```