Homework3 for STAT535

Jie Wang

September 29, 2017

1.

```r
library(stringr)
nhl1415 <- readLines("http://www.stat.cmu.edu/~cshalizi/statcomp/14/hw/03/NHLHockeySchedule2.html")
#which(nchar(nhl1415)==max((nchar(nhl1415))))
```

a. How many lines does it contain?

```r
length(nhl1415)
```

```
## [1] 3141
```

b. What is the total number of characters in the file?

```r
sum(nchar(nhl1415))
```

```
## [1] 2482700
```

c. What is the maximun number of characters in a line?

```r
max((nchar(nhl1415)))
```

```
## [1] 16047
```

2. Answer: The first game is Montreal versus Toronto. The final game is Edmonton versus Vancouver.

3. Answer: The number of line corresponding to game 1 is

```r
grep("Wed Oct 8, 2014.*Montreal",nhl1415)
```

```
## [1] 480
```

The number of line corresponding to game 1230 is

```r
grep("Sat Apr 11, 2015.*Edmonton",nhl1415)
```

```
## [1] 2950
```

The line corresponding to game 1 begins `</h3>`. The rest of them begin with with `</td>`.

4. Answer: For most URL for purchasing tickets, the link has home team in it.

5. Answer:

```r
gamerows=grep('class="skedStartDateSite">(.*?)</div>',nhl1415)
length(gamerows)
```

```
## [1] 1230
```

6. Answer:

```r
Dates=str_match(nhl1415[gamerows],'class="skedStartDateSite">(.*?)</div>')
date=Dates[,2]
date[1:10]
```

```
##  [1] "Wed Oct 8, 2014" "Wed Oct 8, 2014" "Wed Oct 8, 2014"
##  [4] "Wed Oct 8, 2014" "Thu Oct 9, 2014" "Thu Oct 9, 2014"
##  [7] "Thu Oct 9, 2014" "Thu Oct 9, 2014" "Thu Oct 9, 2014"
## [10] "Thu Oct 9, 2014"
```

7. Answer: Away team:

```
away.pattern='Away.*javascript:void\\(0\\);">(.*?)</a>.*?Home'
away.team=str_match(nhl1415[gamerows],away.pattern)
away=away.team[,2]
away[1:10]
```

```
##  [1] "Montreal"     "Philadelphia" "Vancouver"    "San Jose"
##  [5] "Columbus"     "New Jersey"   "Anaheim"      "Montreal"
##  [9] "Boston"       "Florida"
```

Home team:

```
Home.pattern='Home.*javascript:void\\(0\\);">(.*?)</a>.*?Time'
Home.team=str_match(nhl1415[gamerows],Home.pattern)
Home=Home.team[,2]
Home[1:10]
```

```
##  [1] "Toronto"      "Boston"       "Calgary"      "Los Angeles"
##  [5] "Buffalo"      "Philadelphia" "Pittsburgh"   "Washington"
##  [9] "Detroit"      "Tampa Bay"
```

8. Answer:

```
time.pattern='skedStartTimeEST">(.*?)</div>'
game.time=str_match(nhl1415[gamerows],time.pattern)
Time=game.time[,2]
Time[1:10]
```

```
##  [1] "7:00 PM ET"  "7:30 PM ET"  "10:00 PM ET" "10:00 PM ET" "7:00 PM ET"
##  [6] "7:00 PM ET"  "7:00 PM ET"  "7:00 PM ET"  "7:30 PM ET"  "7:30 PM ET"
```

9. Answer: I am a little confused with this problem. In part 4, it mentions that the goal is to extract the URL for buying ticket. However, the URL for the **TICKET>** button is not for purchasing tickets. Actully, the **BUY/SELL** button is for buying tickets. In this question, I identify these two URL because I am not sure which the right one is.

The URL for the **TICKET>** button:

```
URLgamerows=c(gamerows[2:1230], gamerows[1230]+2)
URL.pattern1='href=\"(.*?)"><span>TICKETS'
URL1=str_match(nhl1415[URLgamerows],URL.pattern1)
T.URL=URL1[,2]
length(T.URL)
```

```
## [1] 1230
```

```
T.URL[1:5]
```

```
## [1] "http://www.ticketmaster.ca/artist/806033?intcmp=tm204081&wt.mc_id=NHL_LEAGUE_TIX_LIST_PR_TOR"
## [2] "http://www.ticketmaster.com/artist/805902?brand=nhlgeneric&intcmp=tm204056&wt.mc_id=NHL_LEAGUE_
## [3] "http://www.ticketmaster.ca/artist/805907?intcmp=tm204058&wt.mc_id=NHL_LEAGUE_TIX_LIST_PR_CAL"
## [4] "http://www.ticketmaster.com/artist/805961?intcmp=tm204067&wt.mc_id=NHL_LEAGUE_TIX_LIST_PR_LAK"
## [5] "http://sabres.nhl.com/club/page.htm?id=82007"
```

The URL for purchasing ticket should be the **BUY/SELL** button:

```
URL.pattern2='TICKETS></span></a><a class="btn" shape="rect" href=\"(.*)"><span>BUY/SELL'
URL2=str_match(nhl1415[URLgamerows],URL.pattern2)
B.URL=URL2[,2]
length(B.URL)
```

```
## [1] 1230
```

```
B.URL[1:5]
```

```
## [1] "http://www.ticketsnow.com/nhl/Toronto-Maple-Leafs-tickets?intcmp=tm204051&wt.mc_id=NHL_LEAGUE_T
## [2] "https://teamexchange.ticketmaster.com/html/eventlist.htmI?l=EN&team=bruins&intcmp=tm204026&wt.mo
## [3] "http://www.ticketsnow.com/calgary-flames-tickets?intcmp=cr200076&wt.mc_id=NHL_TM_CAL_TE__PAGE_L:
## [4] "https://teamexchange.ticketmaster.com/html/eventlist.htmI?l=EN&team=kings&brand=nhl&intcmp=tm20
## [5] "http://www.ticketsnow.com/nhl/buffalo-sabres-tickets?intcmp=tm204027&wt.mc_id=NHL_LEAGUE_TIX_LI
```

For the URL for purchasing ticket button, most links have the home team in it.

10.Answer:

```
df=data.frame(date, away, Home, Time, B.URL)
df[1221:1230,]
```

```
##                   date        away        Home       Time
## 1221 Sat Apr 11, 2015    Montreal     Toronto  7:00 PM ET
## 1222 Sat Apr 11, 2015  New Jersey     Florida  7:00 PM ET
## 1223 Sat Apr 11, 2015    Columbus NY Islanders  7:00 PM ET
## 1224 Sat Apr 11, 2015     Detroit    Carolina  7:00 PM ET
## 1225 Sat Apr 11, 2015      Boston   Tampa Bay  7:30 PM ET
## 1226 Sat Apr 11, 2015   Minnesota   St. Louis  7:30 PM ET
## 1227 Sat Apr 11, 2015   Nashville      Dallas  8:00 PM ET
## 1228 Sat Apr 11, 2015     Chicago    Colorado  9:00 PM ET
## 1229 Sat Apr 11, 2015     Anaheim     Arizona  9:00 PM ET
## 1230 Sat Apr 11, 2015     Edmonton   Vancouver 10:00 PM ET
##
## 1221                                    http://www.ticketsnow.com/nhl/Toronto-Maple-Leafs-tickets?intcmp
## 1222 https://teamexchange.ticketmaster.com/html/eventlist.htmI?l=EN&team=panthersnhl&brand=nhl&intcmp
## 1223                                      http://www.ticketsnow.com/nhl/new-york-islanders-tickets?intcmp
## 1224                                      http://www.ticketsnow.com/nhl/carolina-hurricanes-tickets?intcmp
## 1225                                      http://www.ticketsnow.com/nhl/tampa-bay-lightning-tickets?intcr
## 1226                                        http://www.ticketsnow.com/nhl/st-louis-blues-tickets?intcmp
## 1227                                          http://www.ticketsnow.com/nhl/dallas-stars-tickets?intcmp
## 1228                                      http://www.ticketsnow.com/nhl/colorado-avalanche-tickets?intcmp
## 1229                                        http://www.ticketsnow.com/nhl/arizona-coyotes-tickets/?intcmp
## 1230                                          http://www.ticketsnow.com/vancouver-canucks-tickets?int
```