

Homework #2

Fan Yang (fy2232)

November 14, 2017

i.

```
schedual <- readLines("NetsSchedule.html", warn = FALSE)
length(schedual)
```

```
## [1] 828
```

```
sum(nchar(schedual))
```

```
## [1] 129188
```

```
max(nchar(schedual))
```

```
## [1] 9736
```

- a. There are 828 lines.
- b. Total number of characters is 129188.
- c. The maximum number of characters is 9736.

ii.

They playing with Indiana first at 7:00 PM, Wed, Oct 18. They playing with Boston last at 8:00 PM, Wed, Apr 11.

- iii. The 321st line in the file holds information about the first game of the regular season. The 402st line in the file provides the date, time, and opponent for the final game.

iv.

```
regdate <- ">[A-Z]{2},.*[0-9]<"
length(grep(regdate, schedual))
```

```
## [1] 82
```

```
grep(regdate, schedual)
```

```
## [1] 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337
## [18] 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354
## [35] 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
## [52] 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388
## [69] 389 390 391 392 393 394 395 396 397 398 399 400 401 402
```

v.

```
rdate <- regexpr(regdate, schedual)
date <- regmatches(schedual, rdate)
date <- substring(date, 2, nchar(date) - 1)
head(date)
```

```
## [1] "Wed, Oct 18" "Fri, Oct 20" "Sun, Oct 22" "Tue, Oct 24" "Wed, Oct 25"
## [6] "Fri, Oct 27"
```

vi.

```
regtime <- "<td>[0-9].*M</td>"
rtime <- regexpr(regtime, schedual)
```

```
time <- regmatches(schedual, rtime)
time <- substring(time, 5, nchar(time) - 5)
head(time)
```

```
## [1] "7:00 PM" "7:30 PM" "3:30 PM" "7:00 PM" "7:30 PM" "7:30 PM"
```

vii.

```
reghome <- "<li\\sclass=\"game-status\\>\"."
rhome <- regexpr(reghome, schedual)
home <- regmatches(schedual, rhome)
home <- gsub("<li\\sclass=\"game-status\\>\">@", "1", home)
home <- gsub("<li\\sclass=\"game-status\\>\">v", "0", home)
home <- as.numeric(home)
sum(home)
```

```
## [1] 41
```

viii.

```
regopponent <- "class=\"team-name\\>.*>[A-Z]"
roponent <- regexpr(regopponent, schedual)
opponent <- regmatches(schedual, roponent)
opponent <- gsub("class.*>[a-z]/", "", opponent)
opponent <- gsub("\\>.*", "", opponent)
head(opponent, 10)
```

```
## [1] "indiana-pacers" "orlando-magic" "atlanta-hawks"
## [4] "orlando-magic" "cleveland-cavaliers" "new-york-knicks"
## [7] "denver-nuggets" "phoenix-suns" "los-angeles-lakers"
## [10] "phoenix-suns"
```

ix.

```
nets <- data.frame(date, time, opponent, home)
head(nets, 10)
```

```
##      date      time      opponent home
## 1 Wed, Oct 18 7:00 PM    indiana-pacers  1
## 2 Fri, Oct 20 7:30 PM    orlando-magic   0
## 3 Sun, Oct 22 3:30 PM    atlanta-hawks   0
## 4 Tue, Oct 24 7:00 PM    orlando-magic   1
## 5 Wed, Oct 25 7:30 PM cleveland-cavaliers 0
## 6 Fri, Oct 27 7:30 PM    new-york-knicks  1
## 7 Sun, Oct 29 6:00 PM    denver-nuggets  0
## 8 Tue, Oct 31 7:30 PM    phoenix-suns   0
## 9  Fri, Nov 3 10:30 PM  los-angeles-lakers  1
## 10 Mon, Nov 6 9:00 PM    phoenix-suns   1
```

The results match the first 10 games as website.