

HW 3

Salim M'jahad msm2243

February 15, 2018

Part 1

i.

```
nets1718 = readLines("netsschedule.html", warn=FALSE)
summary(nets1718)
```

```
##      Length      Class      Mode
##      828 character character
```

```
sum(nchar(nets1718))
```

```
## [1] 129185
```

```
max(nchar(nets1718))
```

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

It has 828 lines It has 129185 character The maximum number of characters in a single line is 9736

ii.

They are playing Indiana first on Wed, Oct 18 They are playing Boston last on Wed, Apr 11

iii.

```
#nets1718[321]
#nets1718[402]
```

First game is in line 321 Last game is in line 402

iv.

```
index.nets1718 <- grep("[A-Z] [a-z]{2}, [A-Z] [a-z]{2} [0-9]{1,2}", nets1718)
length(index.nets1718)
```

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

v.

```
games.nets1718 <- nets1718[index.nets1718]
matches <- regexpr("[A-Z] [a-z]{2}, [A-Z] [a-z]{2} [0-9]{1,2}", games.nets1718)
```

```

dates <- regmatches(games.nets1718, matches)
head(dates)

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

```

vi.

```

matches <- regexpr("[0-9]{1,2}:[0-9]{2} [A-Z]{2}", games.nets1718)
time <- regmatches(games.nets1718, matches)
head(time)

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

```

vii.

```

status <- grepl(">0<", games.nets1718)
status <- as.integer(!status)
head(status)

## [1] 0 1 1 0 1 0

```

viii.

```

matches <- regexpr(">[a-zA-Z ]+</a></li></ul></td><td>", games.nets1718)
opponent <- regmatches(games.nets1718, matches)
opponent <- substring(opponent, 2, nchar(opponent)-23)
head(opponent)

## [1] "Indiana" "Orlando" "Atlanta" "Orlando" "Cleveland" "NY Knicks"

```

ix.

```

dafa <- data.frame("DATE"=dates, "TIME"=time, "OPPONENT"=opponent, "HOME"=status)
head(dafa)

##      DATE      TIME OPPONENT HOME
## 1 Wed, Oct 18 7:00 PM  Indiana    0
## 2 Fri, Oct 20 7:30 PM  Orlando    1
## 3 Sun, Oct 22 3:30 PM  Atlanta    1
## 4 Tue, Oct 24 7:00 PM  Orlando    0
## 5 Wed, Oct 25 7:30 PM Cleveland    1
## 6 Fri, Oct 27 7:30 PM NY Knicks    0

```