# IS 607: Assignment 10

*MUSA T. GANIYU*

*March 29, 2016*

## Contents

Your task is to choose one of the New York Times APIs, construct an interface in R to read in the JSON d ata, and transform it to an R dataframe

Note: Load the require packages for easy accesibility.

```
library(knitr)
library(XML)
library(jsonlite)
library(plyr)
```

Load the data file from the New York Times webpage. And an API key is required for accessing the webpage as a developer.

```
url.times <- ("http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107-113/nominees/updated.

XML_doc   <- htmlParse(url.times)

str(XML_doc)
```

```
## Classes 'HTMLInternalDocument', 'HTMLInternalDocument', 'XMLInternalDocument', 'XMLAbstractDocument'
```

We found out that the dataset isnt in dataframe yet, we will therefore convert if from list to dataframe.

```
nytimes_all <- ldply(xmlToList(XML_doc), data.frame)
```

As you have noticed, we have a whole bunch of "repeating" columns, we will therefore subset 10 of it and rename it.

```
nytimes <- nytimes_all[, 1:10]

names(nytimes)[names(nytimes)==".id"] <- "id"
names(nytimes)[names(nytimes)=="result_set.status"] <- "status"
names(nytimes)[names(nytimes)=="result_set.copyright"] <- "copyright"
names(nytimes)[names(nytimes)=="result_set.results.congress"] <- "congress_results"
names(nytimes)[names(nytimes)=="result_set.results.num_results"] <- "num_results"
names(nytimes)[names(nytimes)=="result_set.results.nominations.nomination.id"] <- "nomination_id"
names(nytimes)[names(nytimes)=="result_set.results.nominations.nomination.uri"] <- "nomination_uri"
names(nytimes)[names(nytimes)=="result_set.results.nominations.nomination.date_received"] <- "date_rece
names(nytimes)[names(nytimes)=="result_set.results.nominations.nomination.description"] <- "description
names(nytimes)[names(nytimes)=="result_set.results.nominations.nomination.nominee_state"] <- "state"

xmlSize(nytimes) #how many children in node, 10
```

```
## [1] 10
```

```
nytimes[[1]]
```

```
## [1] "body"
```

Here is the extracted dataset head

```
names(nytimes)
```

```
##  [1] "id"               "status"           "copyright"
##  [4] "congress_results" "num_results"      "nomination_id"
##  [7] "nomination_uri"   "date_received"    "description"
## [10] "state"
```

```
kable(head(nytimes))
```

| id | status | copyright | congress_results | num_results |
|---|---|---|---|---|
| body | OK | Copyright (c) 2016 The New York Times Company. All Rights Reserved. | 107 | 20 |

lets now try to access thesame dataset from different format called (JSON)

```
url.json <- ("http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107-113/nominees/updated.j

# We have to extract the file from url to json format using fromJson

nytimes.json <- fromJSON(url.json)

# Setting it to dataframe for analysis

nytimes.json <- ldply (nytimes.json[4], data.frame)

kable(head(nytimes.json))
```

| .id | id | uri | date_receiv |
|---|---|---|---|
| results | PN965 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN965.json | 2001-09-04 |
| results | PN851 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN851.json | 2001-09-04 |
| results | PN817 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN817.json | 2001-09-04 |
| results | PN814 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN814.json | 2001-09-04 |
| results | PN923 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN923.json | 2001-09-04 |
| results | PN922 | http://api.nytimes.com/svc/politics/v3/us/legislative/congress/107/nominees/PN922.json | 2001-09-04 |