Assignment 9

Alice Ding

2023-03-21

Overview

Given the list of APIs found here, I've chosen to look at the most popular one and specifically picked the one that shows the most viewed articles for the last seven days.

```
api_link <- "https://api.nytimes.com/svc/mostpopular/v2/viewed/7.json?api-key="

data <- GET(paste(api_link, Sys.getenv("key"), sep=""))
```

Note: I've stored my `api-key` in a system environment variable called `key`. In order to replicate this .rmd, you would need to set your own key in that variable as well before running all of the code chunks.

JSON -> Data Frame

Using `jsonlite`, let's try to turn this into a dataframe.

```
d <- fromJSON(rawToChar(data$content)) # convert raw unicode
json_df <- as.data.frame(d$results) # pull the results into a data frame
head(json_df, 3)
```

```
##                                                   uri
## 1 nyt://article/0a905873-70c3-50a8-94cd-ffe76ffd7181
## 2 nyt://article/2a080dbf-a7b6-5cbf-a6fa-bfdb14530d5c
## 3 nyt://article/7c9a1c63-67a3-57f7-b29d-e033bfa4f862
##                                                                                                url
## 1           https://www.nytimes.com/2023/03/18/us/politics/trump-indictment-arrest-protests.html
## 2                     https://www.nytimes.com/2023/03/17/arts/television/lance-reddick-dead.html
## 3 https://www.nytimes.com/2023/03/18/us/politics/jimmy-carter-october-surprise-iran-hostages.html
##      id asset_id         source published_date             updated section
## 1 1e+14    1e+14 New York Times     2023-03-18 2023-03-19 13:47:23    U.S.
## 2 1e+14    1e+14 New York Times     2023-03-17 2023-03-20 11:12:38    Arts
## 3 1e+14    1e+14 New York Times     2023-03-18 2023-03-19 09:58:59    U.S.
##   subsection nytdsection
## 1   Politics        u.s.
## 2 Television        arts
## 3   Politics        u.s.
##
## 1                                               Presidential Election of 2016;Campaign Finance;Demons
## 2
## 3 Presidential Election of 1980;United States Politics and Government;Kidnapping and Hostages;United
##   column
## 1     NA
## 2     NA
## 3     NA
##                                                                                    byline
## 1 By Maggie Haberman, Jonah E. Bromwich, Ben Protess, Alan Feuer and William K. Rashbaum
## 2                                                By Michael Levenson and Neil Genzlinger
## 3                                                                         By Peter Baker
##      type
## 1 Article
## 2 Article
## 3 Article
##                                                                         title
## 1 Trump Claims His Arrest Is Imminent and Calls for Protests, Echoing Jan. 6
## 2              Lance Reddick, Star of ‘The Wire’ and ‘John Wick,’ Dies at 60
## 3   A Four-Decade Secret: One Man’s Story of Sabotaging Carter’s Re-election
##
## 1                                                                                His indictment by a
## 2 A prolific actor, he also starred in the Amazon series "Bosch" and had recently branched out into
## 3                                         A prominent Texas politician said he unwittingly took part
##
## 1           Presidential Election of 2016, Campaign Finance, Demonstrations, Protests and Riots, Uni
## 2                                                                        Deaths (Obituaries), Actors
## 3 Presidential Election of 1980, United States Politics and Government, Kidnapping and Hostages, Uni
##                             org_facet
## 1
## 2
## 3 Democratic Party, Republican Party
##                                                                                                per_fac
## 1                Bragg, Alvin, Trump, Donald J, Clifford, Stephanie (1979- ), Cohen, Michael D (1966-
## 2                                                                                           Reddick, Lan
## 3 Barnes, Ben, Connally, John B Jr., Casey, William J, Reagan, Ronald Wilson, Carter, Jimmy, Sick, Ga
##         geo_facet
## 1 Manhattan (NYC)
## 2
## 3            Iran
##
## 1
## 2
## 3 image, photo, "History needs to know that this happened," Ben Barnes now says of his trip to the Mi
##   eta_id
## 1      0
## 2      0
## 3      0
```

Looks like we got it in a dataframe!

Just as a quick peek, what sections are most popular according to this dataset?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
sections <- json_df |>
  group_by(section) |>
  summarise(count = n(),
            .groups = 'drop') |>
  arrange(desc(count))

head(sections, 10)
```

```
## # A tibble: 10 x 2
##    section     count
##    <chr>       <int>
##  1 U.S.            6
##  2 Movies          3
##  3 Business        2
##  4 Arts            1
##  5 Magazine        1
##  6 New York        1
##  7 Opinion         1
##  8 Real Estate     1
##  9 Science         1
## 10 Sports          1
```

Looks like U.S. news is the most popular with 6 of the top articles!

Conclusion & Next Steps

The API call fetches a bit more than the request calls for, but once getting into the `results` JSON, things were simple enough to get into a dataframe. The next steps would definitely be to clean this table up and maybe create a relational database structure to connect things like `media` (currently just an array/list of pictures in each article) as well as the different `geo_facet` values it could have.

One thing I would've liked to have known is how popular is one article over another. I'm not sure how popular each publication is, however maybe they purposefully left that information out as that's something they'd like to analyze themselves.