NYT API exploratory analysis

Hugo Vega

API Description and Research Question

For the New York Times API:

API Description: The NYT API provides access to a variety of endpoints, such as the Article Search API, which allows users to search for NYT articles based on keywords, dates, and other filters. **Data Content*: This API returns metadata about articles, including titles, abstracts, URLs, and publication dates. **Research Question**: For example, "How has the New York Times covered advancements in AI over the past year?" This question will guide the query, filtering articles related to "artificial intelligence" or "AI."

```r
library(tidyverse)
library(DBI)
library(duckdb)
library(nycflights13)
library(dplyr)
library(ggplot2)
library(maps)
library(tidyr)
library(rvest)
library(robotstxt)
library(scales)
library(stringr)
library(purrr)
library(httr)
library(httr2)
library(jsonlite)
library(lubridate)
```

Using the NY Times API

Problem 6: The New York Times web site provides a rich set of APIs, as described here . You'll need to start by signing up for an API key. Your task is to choose one of the New York Times APIs, construct an interface in R to read in the JSON data, and transform it into an R DataFrame.

Set Up my API Key and Base URL

```
Rows: 1 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): key

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Query the Article Search API Let's make a simple request to the Article Search API for articles related to "technology".

```r
# Define the base URL for the Article Search API
base_url <- "https://api.nytimes.com/svc/search/v2/articlesearch.json"

# Define the query parameters (e.g., search for 'technology' articles)
query <- "technology"

# Build and perform the request using httr2
request <- request(base_url) %>%
  req_url_query("q" = query, "api-key" = api_key) %>%
  req_perform()

# Check if the request was successful
if (resp_status(request) == 200) {

  # Parse the JSON response
  json_data <- resp_body_json(request, simplifyVector = TRUE)

  # Extract the articles section of the response
  articles <- json_data$response$docs

  # Convert articles to a data frame
  articles_df <- as.data.frame(articles)
```

```
  # Print the first few rows of the data frame
  print(head(articles_df))

} else {
  cat("Error: Failed to retrieve data. Status code:", resp_status(request), "\n")
}
```

```
1
2                                                 Sales for the iPhone maker were up 6 percent to to S
3                                                  The world's richest men have the
4 The U.S. government has tried to keep Chinese companies from obtaining certain advanced tec
5
6                                                  Suchir Balaji helped gather and

1 https://www.nytimes.com/2024/11/01/podcasts/billionaire-game-theory-we-are-not-ready-for-ag
2                                     https://www.nytimes.com/2024/10
3                                https://www.nytimes.com/2024/10/30/tech
4                               https://www.nytimes.com/2024/10/29
5                        https://www.nytimes.com/2024/10/25/podcasts
6                               https://www.nytimes.com/2024/1

1
2                                                 Sales for the iPhone maker were up 6 percent to to S
3                                                  The world's richest men have the
4 The U.S. government has tried to keep Chinese companies from obtaining certain advanced tec
5
6                                                  Suchir Balaji helped gather and

1 Last week, Jeff Bezos canceled the Washington Post editorial board's plan to endorse Kamala
2
3
4
5
6
              source
1 The New York Times
2 The New York Times
3 The New York Times
4 The New York Times
5 The New York Times
```

```
6 The New York Times


1
2
3
4
5 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
6
                                                                   headline.main
1 Billionaire Game Theory + We Are Not Ready for A.G.I. + Election Betting Markets Get Weird
2                                 Apple's Quarterly Profit Down Because of Tax Payment in Europe
3                       Jeff Bezos, Elon Musk and the Billions of Ways to Influence an Election
4              TSMC Chips Ended Up in Devices Made by China's Huawei Despite U.S. Controls
5                                 The Elon-ction + Can A.I. Be Blamed for a Teen's Suicide?
6                         Former OpenAI Researcher Says the Company Broke Copyright Law
  headline.kicker headline.content_kicker
1            NA                      NA
2            NA                      NA
3            NA                      NA
4            NA                      NA
5            NA                      NA
6            NA                      NA
                         headline.print_headline headline.name headline.seo
1                                          <NA>            NA           NA
2        Tax Bill Makes Apple's Profit Take Big Hit            NA           NA
3              Billionaires Make Bets On Control            NA           NA
4          Despite U.S., Huawei Gets TSMC Chips            NA           NA
5                                          <NA>            NA           NA
6 Seeing Potential for Harm, OpenAI Researcher Quit            NA           NA
  headline.sub
1          NA
2          NA
3          NA
4          NA
5          NA
6          NA


1
2
3 persons, persons, subject, subject, subject, subject, subject, organizations, subject, orga
4
5
6
```

4

```
                 pub_date document_type news_desk section_name
1 2024-11-01T10:00:16+0000       article  Podcasts     Podcasts
2 2024-10-31T21:10:57+0000       article  Business   Technology
3 2024-10-30T15:42:42+0000       article  Business   Technology
4 2024-10-29T16:43:42+0000       article  Business Business Day
5 2024-10-25T10:01:07+0000       article  Podcasts     Podcasts
6 2024-10-23T15:35:46+0000       article  Business   Technology


1         By Kevin Roose, Casey Newton, Whitney Jones, Rachel Cohn, Jen Poyant, Chris Wood, 
2                                                                                          
3                                                                                          
4                                                                             By Meaghan Tobin, A
5 By Kevin Roose, Casey Newton, Rachel Cohn, Whitney Jones, Jen Poyant, Daniel Ramirez, Dan 
6                                                                                          


1         Kevin, Casey, Whitney, Rachel, Jen, Chris, Dan, Marion, Diane, NA, NA, NA, NA, NA, 
2                                                                                          
3                                                                                          
4                                                                                          
5 Kevin, Casey, Rachel, Whitney, Jen, Daniel, Dan, Sophia, Rowan, NA, NA, NA, NA, NA, NA, NA
6                                                                                          
  byline.organization type_of_material
1                  NA             News
2                  NA             News
3                  NA             News
4                  NA             News
5                  NA             News
6                  NA             News
                                              _id word_count
1 nyt://article/430e2cae-5685-53bf-b753-25ae867c553c        218
2 nyt://article/71236860-6ef7-54cc-bda3-561c068c2d3f        841
3 nyt://article/f518411a-a3a1-5a0e-b6fe-53c90b6c9c55       1430
4 nyt://article/b44bbcdd-0f4c-572f-84f0-de1f935ddfff        917
5 nyt://article/a86023a3-81ca-51d9-92c5-61577a568422        201
6 nyt://article/dc193626-329b-5cbb-8c86-1d5a4de9f1e5       1226
                                              uri print_section print_page
1 nyt://article/430e2cae-5685-53bf-b753-25ae867c553c          <NA>       <NA>
2 nyt://article/71236860-6ef7-54cc-bda3-561c068c2d3f             B          3
3 nyt://article/f518411a-a3a1-5a0e-b6fe-53c90b6c9c55             B          1
4 nyt://article/b44bbcdd-0f4c-572f-84f0-de1f935ddfff             B          1
5 nyt://article/a86023a3-81ca-51d9-92c5-61577a568422          <NA>       <NA>
6 nyt://article/dc193626-329b-5cbb-8c86-1d5a4de9f1e5             B          1
```

```r
# Set the base URL and API key
base_url <- "https://api.nytimes.com/svc/search/v2/articlesearch.json"
api_key <- read_csv('api.key.csv') %>%
pull(key)
```

```
Rows: 1 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): key

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Set the query for articles related to "artificial intelligence"
query <- "artificial intelligence"

# Build and perform the request using httr2
request <- request(base_url) %>%
  req_url_query("q" = query, "api-key" = api_key) %>%
  req_perform()

# Check if the request was successful
if (resp_status(request) == 200) {

  # Parse JSON response
  json_data <- resp_body_json(request, simplifyVector = TRUE)

  # Extract articles data
  articles <- json_data$response$docs

  # Convert articles to a data frame and handle fields
  articles_df <- as_tibble(articles) %>%
    mutate(
      # Use `abstract` as `headline` since `headline` doesn't exist
      headline = if ("abstract" %in% names(.)) abstract else NA_character_,
      # Check if `pub_date` exists and parse it
      pub_date = if ("pub_date" %in% names(.)) as.Date(pub_date) else NA_Date_
    ) %>%
    select(pub_date, headline, web_url, snippet)  # Select relevant fields

  # Check the resulting data frame
```

```r
  print(head(articles_df))

  # Visualization: Monthly article counts on 'Artificial Intelligence'
  articles_df %>%
    filter(!is.na(pub_date)) %>%  # Filter out rows with missing dates
    mutate(month = floor_date(pub_date, "month")) %>%
    count(month) %>%
    ggplot(aes(x = month, y = n)) +
    geom_line(color = "blue") +
    labs(title = "Monthly Article Counts on 'Artificial Intelligence'",
         x = "Month", y = "Number of Articles") +
    theme_minimal()

} else {
  print("Failed to retrieve data.")
}
```

```
# A tibble: 6 x 4
  pub_date   headline                                          web_url snippet
  <date>     <chr>                                             <chr>   <chr>
1 2024-11-01 It's already powering remarkable visual innovation~ https:~ It's a~
2 2024-09-29 Artificial intelligence poses unique risks, so the~ https:~ Artifi~
3 2024-09-29 The bill would have been the first in the nation t~ https:~ The bi~
4 2024-09-09 The company also showcased a new Apple Watch and A~ https:~ The co~
5 2024-11-01 The change, starting next Friday, lifts a dominant~ https:~ The ch~
6 2024-10-30 The company's profit increased 11 percent to $24.7~ https:~ The co~
```

Monthly Article Counts on 'Artificial Intelligence'