

# Data 607 Assignment 9

Joao De Oliveira

October 26, 2025

## Overview

Use the New York Times Top Stories API to read JSON data and transform it into an R DataFrame, then save it to a CSV next to this Rmd.

## Diagnostics

```
## API key present: YES
```

```
## Key length: 32
```

```
## Document directory (doc_dir): /Users/joaodeoliveira/Documents/Data607/Assignment9
```

```
## Working directory during knit (getwd): /Users/joaodeoliveira/Documents/Data607/Assignment9
```

## Fetch & Transform JSON into DataFrame

```
nyt_top_stories <- function(section = "science", api_key = Sys.getenv("NYT_API_KEY")) {  
  api_key <- trimws(api_key)  
  if (!nzchar(api_key)) stop("NYT_API_KEY is empty. Put your key in .Renviron or Sys.setenv().")  
  
  base_url <- paste0("https://api.nytimes.com/svc/topstories/v2/", section, ".json")  
  
  req <- httr2::request(base_url) |>  
    httr2::req_url_query(`api-key` = api_key) |>  
    httr2::req_retry(max_tries = 3, backoff = ~ .x * 0.5) |>  
    httr2::req_error(is_error = function(resp) FALSE)  
  
  resp <- httr2::req_perform(req)  
  status <- httr2::resp_status(resp)  
  
  if (status == 401) {  
    body <- tryCatch(httr2::resp_body_string(resp), error = function(e) "")  
    stop("HTTP 401 Unauthorized. Check the key (no trailing spaces/newlines) and that the Top Stories API key is valid.  
        if (nzchar(body)) paste0("Server said: ", substr(body, 1, 200), "...") else """)  
  }  
  if (status >= 400) stop(paste("HTTP", status, "error from NYT endpoint."))  
}
```

```

obj <- jsonlite::fromJSON(httr2::resp_body_string(resp), flatten = TRUE)
if (is.null(obj$results) || length(obj$results) == 0) stop("No results returned by the API.")

articles <- tibble::as_tibble(obj$results)

articles |>
  dplyr::select(
    section, subsection, title, abstract, byline, published_date, url,
    des_facet, org_facet, per_facet, geo_facet, multimedia
  ) |>
  dplyr::mutate(
    published_date = as.POSIXct(published_date, tz = "UTC"),
    des_facet = purrr::map_chr(des_facet, ~ if (is.null(.x)) NA_character_ else paste(.x, collapse = ", ")),
    org_facet = purrr::map_chr(org_facet, ~ if (is.null(.x)) NA_character_ else paste(.x, collapse = ", ")),
    per_facet = purrr::map_chr(per_facet, ~ if (is.null(.x)) NA_character_ else paste(.x, collapse = ", ")),
    geo_facet = purrr::map_chr(geo_facet, ~ if (is.null(.x)) NA_character_ else paste(.x, collapse = ", ")),
    image_url = purrr::map_chr(multimedia, ~ {
      if (is.null(.x) || length(.x) == 0) return(NA_character_)
      idx <- tryCatch(which.max(.x$width), error = function(e) 1L)
      .x$url[idx] %||% NA_character_
    })
  ) |>
  dplyr::select(-multimedia)
}

```

```

section_chosen <- "science"
top_stories <- nyt_top_stories(section_chosen)

```

## Preview the Resulting DataFrame

```
glimpse(top_stories)
```

```

## Rows: 27
## Columns: 12
## $ section      <chr> "science", "admin", "climate", "climate", "science", "h~
## $ subsection   <chr> "", "", "", "", "", "", "", "politics", "", "", "", ~
## $ title        <chr> "Sign Up for the Science Times Newsletter", "", "Exxon ~
## $ abstract     <chr> "Every week, we'll bring you stories that capture the w~
## $ byline       <chr> "", "", "By Karen Zraick", "By Karen Zraick and Lisa Fr~
## $ published_date <dtm> 2016-02-05, 2015-04-16, 2025-10-25, 2025-10-25, 2025-1~
## $ url          <chr> "null", "", "https://www.nytimes.com/2025/10/25/climate~
## $ des_facet    <chr> "", "", "Law and Legislation, Oil (Petroleum) and Gasol~
## $ org_facet    <chr> "", "", "Exxon Mobil Corp, California Air Resources Boa~
## $ per_facet    <chr> "", "", "Bonta, Rob (1972- )", "", "", "", "Bean, Alan ~
## $ geo_facet    <chr> "", "", "California", "", "New Zealand", "Uganda", "", ~
## $ image_url    <chr> "https://static01.nyt.com/images/2016/02/06/science/sci~

```

```
head(dplyr::select(top_stories, title, byline, published_date, url))
```

```
## # A tibble: 6 x 4
```

```
##   title                               byline published_date      url
##   <chr>                               <chr>   <dtm>             <chr>
## 1 "Sign Up for the Science Times Newsletter" ""      2016-02-05 00:00:00 "nul~
## 2 ""                                ""      2015-04-16 00:00:00 ""
## 3 "Exxon Sues California Over New Climate Disc~ "By K~ 2025-10-25 00:00:00 "htt~
## 4 "An E.P.A. Plan to Kill a Major Climate Rule~ "By K~ 2025-10-25 00:00:00 "htt~
## 5 "Take a Look at Rare Photos of Red Lightning~ "By F~ 2025-10-25 00:00:00 "htt~
## 6 "In Fight Against Malaria, an Unexpected - a~ "By S~ 2025-10-25 00:00:00 "htt~
```

Save to CSV (next to this Rmd)

```
csv_path <- file.path(doc_dir, "nyt_top_stories.csv")
write.csv(top_stories, csv_path, row.names = FALSE)
cat("Saved CSV to:", normalizePath(csv_path, winslash = "/"), "\n")
```

```
## Saved CSV to: /Users/joaodeoliveira/Documents/Data607/Assignment9/nyt_top_stories.csv
```

```
cat("Files in doc_dir now include:", paste(list.files(doc_dir), collapse = ", "), "\n")
```

```
## Files in doc_dir now include: data607_assignment9_FIXED.pdf, data607_assignment9_FIXED.Rmd, data607_
```