

Homework APIs

Shane Hylton

10/23/2021

NYT Bestselling Books API

I chose to work with the New York Times Bestseller List. I loaded the data into R using the api-key I requested, then I collected the raw data from the JSON data. I then took the raw data, selected the results subsection, and created a dataframe based on books, where the data is stored. From there, I performed some minor cleaning procedures to improve the appearance of the dataframe.

```
res <- GET("https://api.nytimes.com/svc/books/v3/lists/current/hardcover-fiction.json?api-key=YWPApnXSK")

raw <- jsonlite::fromJSON(unlist(rawToChar(res$content)))

df <- data.frame(raw$results$books)
dftrim <- subset(df, select = -c(4,5,14, 19:23))
books <- dftrim %>% relocate(c("title", "author", "publisher"), .after = "weeks_on_list")

books_quick <- subset(books, select = c(1:6))
colnames(books_quick) <- c("rank", "prev_rank", "weeks_on_list", "title", "author", "publisher")
books_quick
```

	rank	prev_rank	weeks_on_list	title	
## 1	1	0	1	STATE OF TERROR	
## 2	2	2	3	THE WISH	
## 3	3	1	2	THE LINCOLN HIGHWAY	
## 4	4	3	3	CLOUD CUCKOO LAND	
## 5	5	5	5	APPLES NEVER FALL	
## 6	6	0	1	SILVERVIEW	
## 7	7	8	24	THE LAST THING HE TOLD ME	
## 8	8	0	1	THE BOOK OF MAGIC	
## 9	9	6	5	HARLEM SHUFFLE	
## 10	10	7	2	THE BUTLER	
## 11	11	9	11	BILLY SUMMERS	
## 12	12	4	2	CROSSROADS	
## 13	13	0	45	THE MIDNIGHT LIBRARY	
## 14	14	13	6	BEAUTIFUL WORLD, WHERE ARE YOU	
## 15	15	10	4	THE JAILHOUSE LAWYER	
##				author	publisher
## 1				Hillary Rodham Clinton and Louise Penny	Simon & Schuster, St. Martin's
## 2				Nicholas Sparks	Grand Central
## 3				Amor Towles	Viking
## 4				Anthony Doerr	Scribner
## 5				Liane Moriarty	Holt

## 6	John Le Carré	Viking
## 7	Laura Dave	Simon & Schuster
## 8	Alice Hoffman	Simon & Schuster
## 9	Colson Whitehead	Doubleday
## 10	Danielle Steel	Delacorte
## 11	Stephen King	Scribner
## 12	Jonathan Franzen	Farrar, Straus & Giroux
## 13	Matt Haig	Viking
## 14	Sally Rooney	Farrar, Straus & Giroux
## 15	James Patterson and Nancy Allen	Little, Brown

Closing Thoughts

The workflow for loading the JSON data into a dataframe was somewhat straightforward after taking the time to discover in which subsection the data was stored. Using APIs should enable me to collect up to date data from a variety of sources. The Books API alone has an immense amount of data to pull from, and just like a SQL relational database queries are the way to obtain relevant data.