

KSooklall_Homework9 DATA607

Kenan Sooklall

4/9/2021

```
library(httr)
library(jsonlite)
library(tidyverse)
library(ggplot2)
```

The api key would normally be in a hidden file and read here with readLine

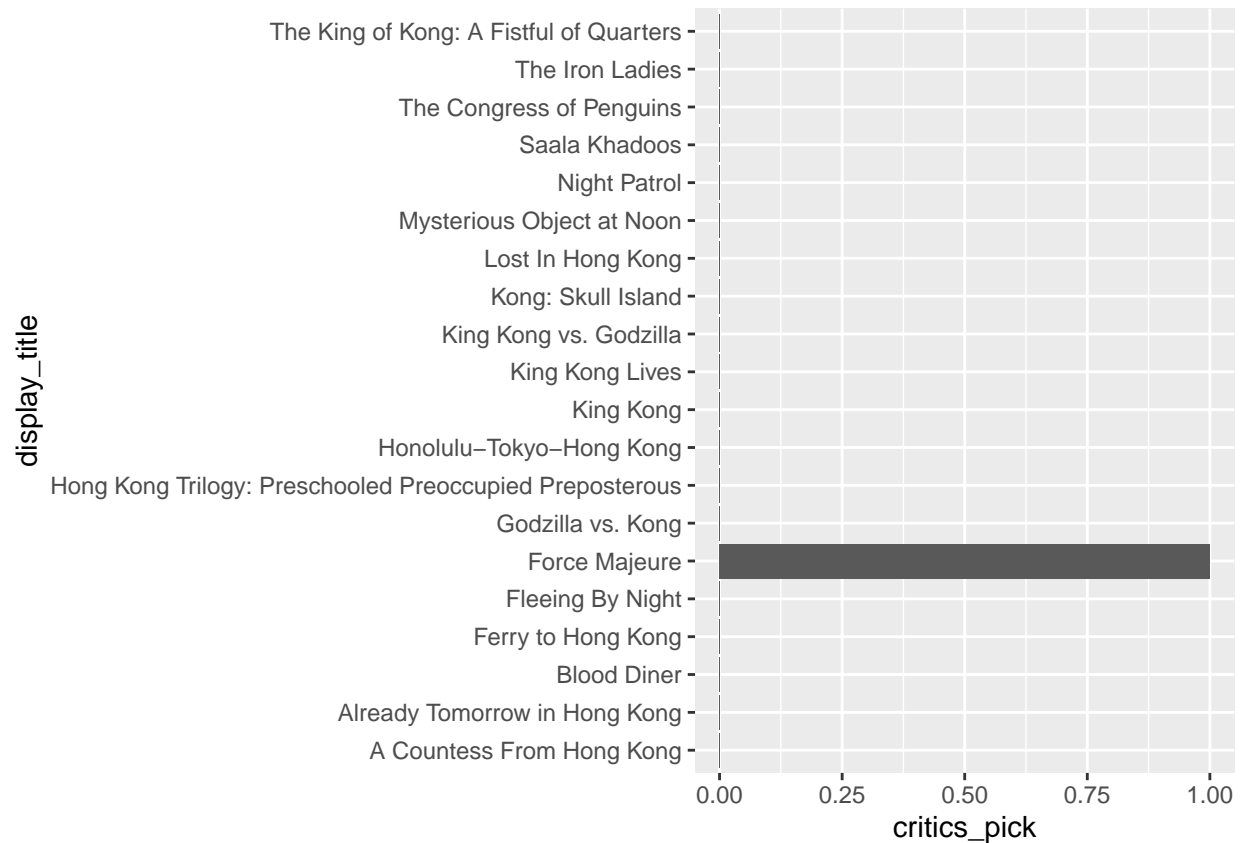
```
api_key = 'qdUXTcfRvGEgKiHVfrG80c3vwWW8YJZG'
base_url <- "https://api.nytimes.com/svc/movies/v2/reviews/search.json?query=kong"
url <- paste0(base_url, "&api-key=", api_key)
response <- GET(url)
```

```
json <- fromJSON(content(response, as="text"))
df <- data.frame(json$results)
glimpse(df)
```

```
## Rows: 20
## Columns: 11
## $ display_title    <chr> "Godzilla vs. Kong", "Hong Kong Trilogy: Preschooled ~
## $ mpaa_rating      <chr> "PG-13", "", "PG-13", "", "", "", "R", "PG-13", "", "~
## $ critics_pick     <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ byline           <chr> "A.O. Scott", "BEN KENIGSBURG", "MANOHLA DARGIS", "NE~
## $ headline         <chr> "'Godzilla vs. Kong' Review: Let's You and Him Fight"~
## $ summary_short    <chr> "Ape and lizard go toe-to-toe, with a cast of talente~
## $ publication_date  <chr> "2021-03-30", "2017-09-21", "2017-03-09", "2016-02-11~
## $ opening_date     <chr> "2021-03-31", NA, "2017-03-10", "2016-02-12", "2016-0~
## $ date_updated      <chr> "2021-03-30 16:16:05", "2017-11-02 04:16:44", "2017-1~
## $ link              <df[,3]> <data.frame[20 x 3]>
## $ multimedia       <df[,4]> <data.frame[20 x 4]>
```

EDA

```
df %>% select(display_title, critics_pick) %>% ggplot(aes(x=display_title, y=critics_pick)) + geom_col()
```



Tough critics, only Force Majeure was picked

```
library(wordcloud)
library(tm)

docs <- Corpus(VectorSource(df$summary_short))

toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords, c("blabla1", "blabla2"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
```

