New York Times Analysis

Johnny Depp v Amber Heard

Pratiksha Sharma
Daniel Wang

01.

DATA RETRIEVAL

New York Times API + Web Scraping

02.

MOST POPULAR TOPICS

How we explored the most popular articles using topic modeling and sentiment analysis

03. Professional Life

JOHNNY DEPP & AMBER HEARD: MOVIES

How we analyzed reviews of movies starring Depp and Heard

04. Personal Life

JOHNNY DEPP & AMBER HEARD: RELATIONSHIP AND TRIAL

How we analyzed the news coverage of their relationship over time
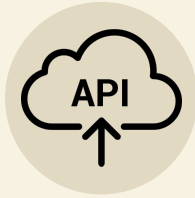
01.

DATA RETRIEVAL
METHOD

New York Times API +
Web Scraping

API & Web Scraping

API

Shortlisting the API & Setting search parameters

Web Scraping

Scraping text content from article HTML

Data Retrieval : Extracting Article Metadata

```
url <-
"https://api.nytimes.com/svc/mostpopular/v2/viewed
/1.json?api-key=ltnSuqlZoDGg6Td519vELgT8Fncs2Bq7"

json_result <- url %>% curl() %>% readLines() %>%
fromJSON()


json_result$results$title
```

→

```
 [1] "Johnny Depp v. Amber Heard: What We Know"
 [2] "A Look Inside the Textbooks That Florida Rejected"
 [3] "McCarthy Said Trump Acknowledged 'Some Responsibility' for Jan. 6"
 [4] "I Lived the #VanLife. It Wasn't Pretty."
 [5] "Bonnie Raitt Faces Mortality With Compassion and Hope"
 [6] "The Former Chancellor Who Became Putin's Man in Germany"
 [7] "Spelling Bee Forum"
 [8] "Suspect in Sniper-Style Shooting in Washington Kills Himself, Police Say"
 [9] "Johnny Depp Confronted With Texts and Audio at Libel Trial"
[10] "Donald Trump's Newest Problem: Elon Musk"
[11] "In Polarized Texas, Rare Accord: A Hispanic Mother Shouldn't Be Executed"
[12] "Orrin Hatch, Seven-Term Senator and a Republican Force, Dies at 88"
[13] "How Much Should You Be Asked to Donate for a Colleague's Gift?"
[14] "Punishing Disney, DeSantis Signals a Lasting G.O.P. Brawl With Business"
[15] "Why the Road Is Getting Even Rockier for First-Time Home Buyers"
[16] "Scientists Find No Benefit to Time-Restricted Eating"
[17] "Meadows Was Warned Jan. 6 Could Turn Violent, House Panel Says"
[18] "Five Action Movies to Stream Now"
[19] "Jeffrey Epstein, a Rare Cello and an Enduring Mystery"
[20] "Charges Dropped Against Tennessee Woman Who Was Jailed Over Voter Fraud"
```

Data Retrieval : Scraping Content from the HTML

```r
extract_text_mp <- function(json_result){
  articles <- json_result$results %>%
    dplyr::select(c('title','url'))
  article_text <-tibble(headline = '',text='')
  for(i in 1:nrow(articles)){
    url_html <- articles[i,2]
    article_text <- article_text %>% add_row(headline =
articles[i,1], text = url_html %>% read_html()
%>% html_nodes("section") %>% html_nodes("p")
%>% html_text2())
  }
  article_text <- article_text %>%
        group_by(headline) %>%
summarise(text = paste(text,collapse = " "))
  article_text <- article_text [-c(1),]
  return(article_text )
}
```

Not Clean!

	headline	text
1	'I've Had It With This Guy': G.O.P. Leaders Privately Blasted T...	In the days after the Jan. 6 attack on the Capitol building, th...
2	A Look Inside the Textbooks That Florida Rejected	After the Florida Department of Education rejected dozens ...
3	A YouTuber Purposely Crashed His Plane in California, F.A.A...	The Federal Aviation Administration has found that Trevor J...
4	CNN+ Streaming Service Will Shut Down Weeks After Its Start	At sunset on the last Monday of March, CNN stars gathered...
5	Disney to Lose Special Tax Status in Florida Amid 'Don't Say ...	Disney employs 38 lobbyists in Florida's capital. Each electio...
6	Does My Mask Protect Me if Nobody Else Is Wearing One?	If I'm the only person wearing a mask on a plane or subway ...
7	Don't Call Her 'Baby.' At 62, Jennifer Grey Is Taking the Lead.	Jennifer Grey arrived at a recent breakfast at the Peninsula ...
8	Effort to Remove Marjorie Taylor Greene From Ballot Can Pr...	A federal judge cleared the way on Monday for a group of ...
9	Federal Judge Strikes Down Mask Mandate for Planes and P...	WASHINGTON — A federal judge in Florida struck down the...
10	Gov. Ron DeSantis clears a way to revoke Disney's special di...	In a move widely seen as retaliation, Gov. Ron DeSantis aske...
11	How a Lone Tenant Is Holding Up a $70 Million Condo Deal	Ahmet Nejat Ozsu won't budge. For 15 years, he has lived in...
12	I Lived the #VanLife. It Wasn't Pretty.	To hear more audio stories from publications like The New Y...
13	Johnny Depp Confronted With Texts and Audio at Libel Trial	Lawyers for Amber Heard worked on Thursday to point out i...
14	Johnny Depp v. Amber Heard: What We Know	The defamation trial in Virginia between the actors Johnny ...

Data Retrieval : Cleaning and Preprocessing text for Latent Dirichlet Modelling and Sentiment Analysis

```r
article_text <- extract_text(json_result)
article_text <- article_text %>% mutate(
  clean_text = tolower(text),
  clean_text = removeWords(clean_text,stop_words$word),
  clean_text = str_replace_all(clean_text,"['`$+]", ""),
  clean_text = str_replace_all(clean_text,"[[:punct:]]", " "),
  clean_text = str_replace_all(clean_text,'[[:digit:]]+', " "),
  clean_text = str_replace_all(clean_text,
              "[[:space:]]+", " "),
  clean_text = trimws(clean_text))
```

| text	clean_text
One evening last April, an anxious and free-spirited 13-year-...	evening april anxious free spirited girl suburban minneapoli...
After the Florida Department of Education rejected dozens ...	florida department education rejected dozens math textboo...
Who would expect a Bonnie Raitt song to start like this? "Ha...	expect bonnie raitt song start flu prison infirmary sings hall ...
A Tennessee prosecutor dropped all criminal charges on Fri...	tennessee prosecutor dropped criminal charges friday pame...
WASHINGTON — A Colorado man who set himself on fire in...	washington colorado set fire front supreme court friday app...
BEIJING — Families in Beijing rushed to stock up on food. S...	beijing families beijing rushed stock food supermarkets stay...
"My husband has been sick for a month and can barely walk...	husband sick month barely walk send home woman reserve...
To hear more audio stories from publications like The New Y...	hear audio stories publications york times download audm i...
David Zaslav had been chief executive of Warner Bros. Disco...	david zaslav chief executive warner bros discovery hours lea...
The defamation trial in Virginia between the actors Johnny ...	defamation trial virginia actors johnny depp amber heard fie...

02.

MOST POPULAR
TOPICS

How we explored the most popular articles using topic modeling and sentiment analysis

Most Popular Topics : Latent Dirichlet allocation

```r
dtm_mp <-
DocumentTermMatrix(VCorpus(VectorSource(mp_text
$clean_text)),control=list(wordLengths=c(1,Inf)
))
mp_lda <- LDA(dtm_mp, k = 2)
mp_topics <- tidy(mp_lda)
mp_topics %>% filter(topic==2) %>%
   mutate(beta_rank = min_rank(desc(beta))) %>%
   filter(beta_rank <= 10) %>%
   arrange(beta_rank) %>%
   mutate(term = reorder(term, beta)) %>%
   ggplot(aes(beta, term)) +
   geom_col(show.legend = FALSE)
```
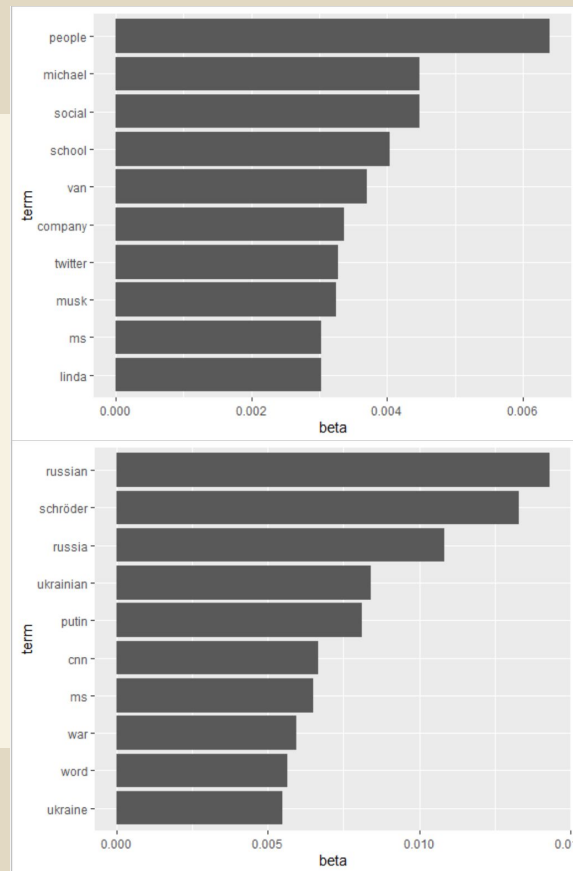
Most Popular Topics : Latent Dirichlet allocation
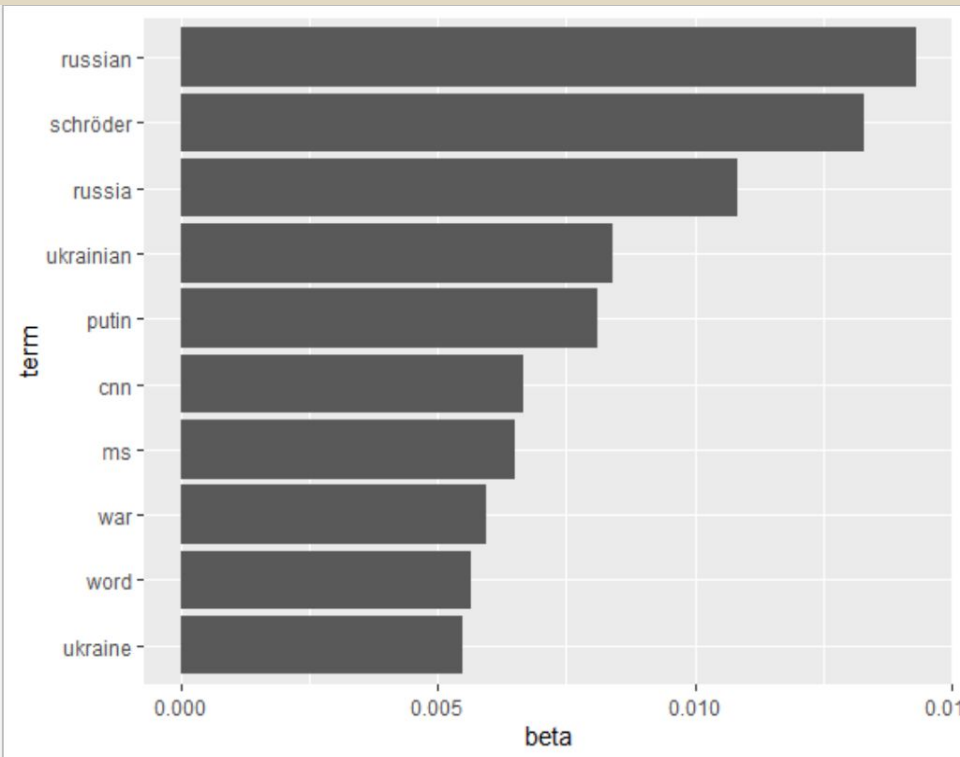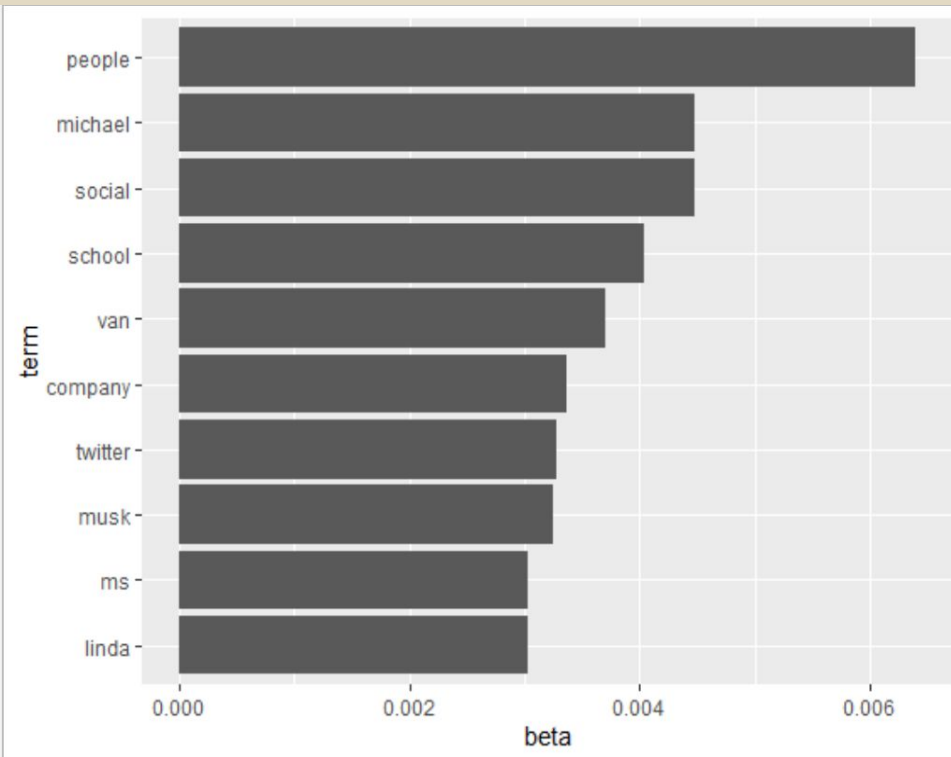
Most Popular Topics : Latent Dirichlet allocation

```
beta_wide <- mp_topics %>%
  mutate(topic = paste0("topic", topic)) %>%
  pivot_wider(names_from = topic, values_from = beta) %>%
  filter(topic1 > 0.001 | topic2 > 0.001) %>%
  mutate(log_ratio = log2(topic2 / topic1))

beta_wide %>% arrange(desc(log_ratio)) %>%
  mutate(term = reorder(term, log_ratio)) %>%
  head(10) %>%
  ggplot(aes(log_ratio, term)) +
  geom_col(show.legend = FALSE)
```
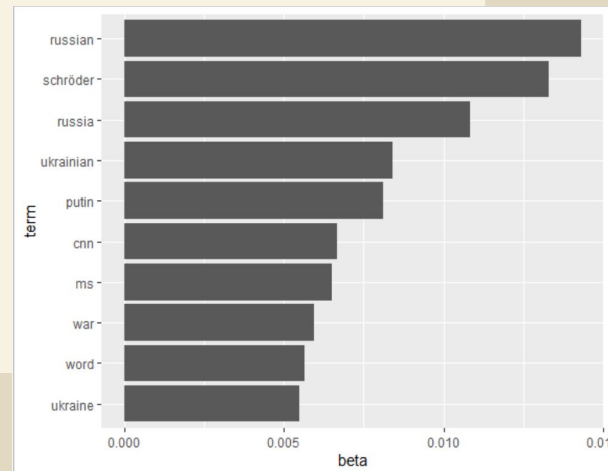
Most Popular Topics : Latent Dirichlet allocation
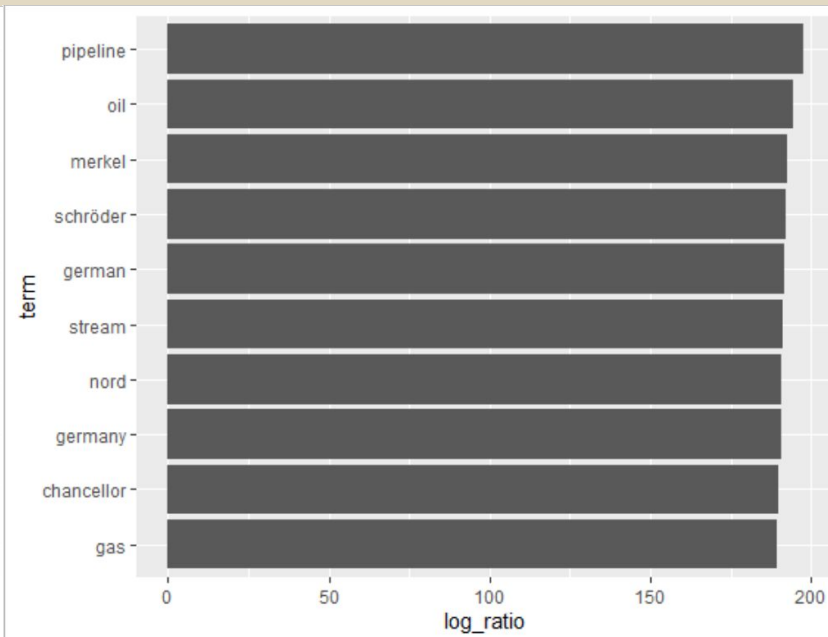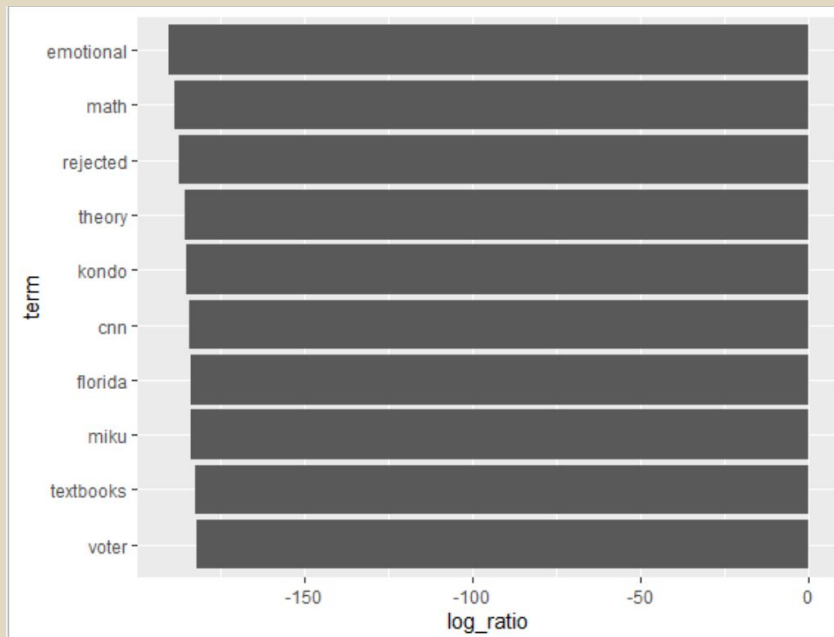
Most Popular Topics : Sentiment Analysis

```
sa_text =
article_text %>% group_by(headline)
%>% unnest_tokens(word, clean_text)
```

→

headline <chr>	word <chr>
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	evening
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	april
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	anxious
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	free
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	spirited
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	girl
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	suburban
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	minneapolis
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	sprang
'It's Life or Death': The Mental Health Crisis Among U.S. Teens	furious

```
sa_processed_text =
sa_text %>% inner_join(get_sentiments("nrc"))
%>% group_by(headline)
%>% count(sentiment, sort = TRUE)
```

→

|	headline <chr>	sentiment <chr>	n <int>
1	'It's Life or Death': The Mental Health Crisis Among U.S. Teens	negative	237
2	The Former Chancellor Who Became Putin's Man in Germany	positive	207
3	'It's Life or Death': The Mental Health Crisis Among U.S. Teens	positive	194
4	I Lived the #VanLife. It Wasn't Pretty.	positive	181
5	'It's Life or Death': The Mental Health Crisis Among U.S. Teens	fear	174
6	'It's Life or Death': The Mental Health Crisis Among U.S. Teens	sadness	155

Most Popular Topics : Sentiment Analysis

```
sa_processed_text %>% ggplot(aes(reorder
(sentiment,-n), n, color = sentiment))
+ geom_bar(stat = "identity") +
theme(axis.text.x = element_text
(angle = 45, hjust = 1))+
xlab("Sentiment")
+ ylab("Number of Words") + ggtitle(
"Sentiments in Most Popular Articles") +
theme(legend.position="none")
```
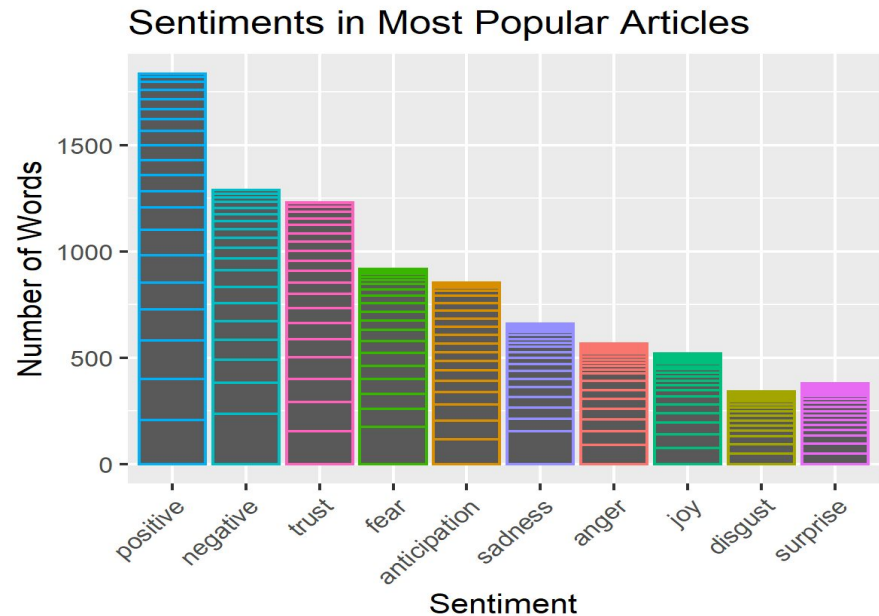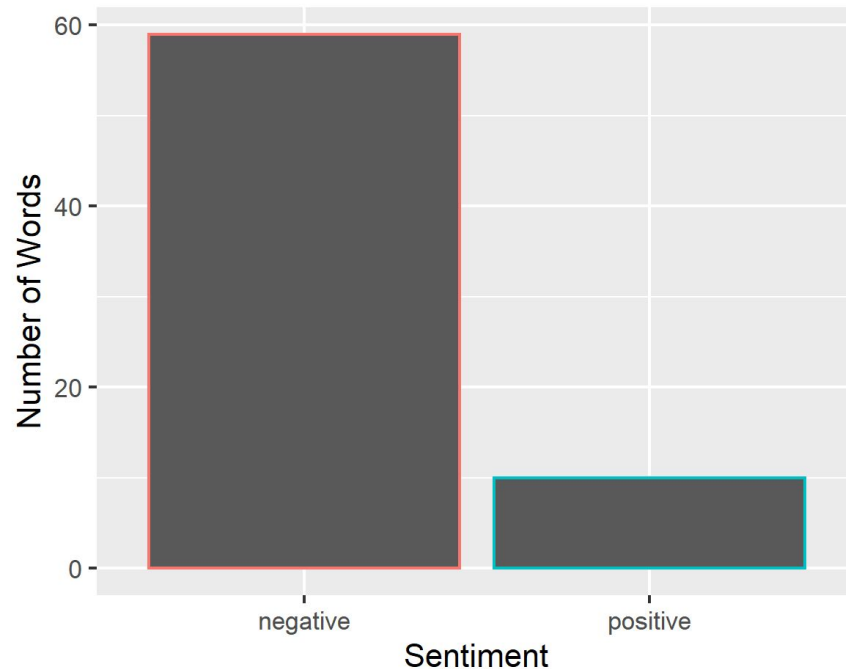

Sentiments in Most Popular Articles
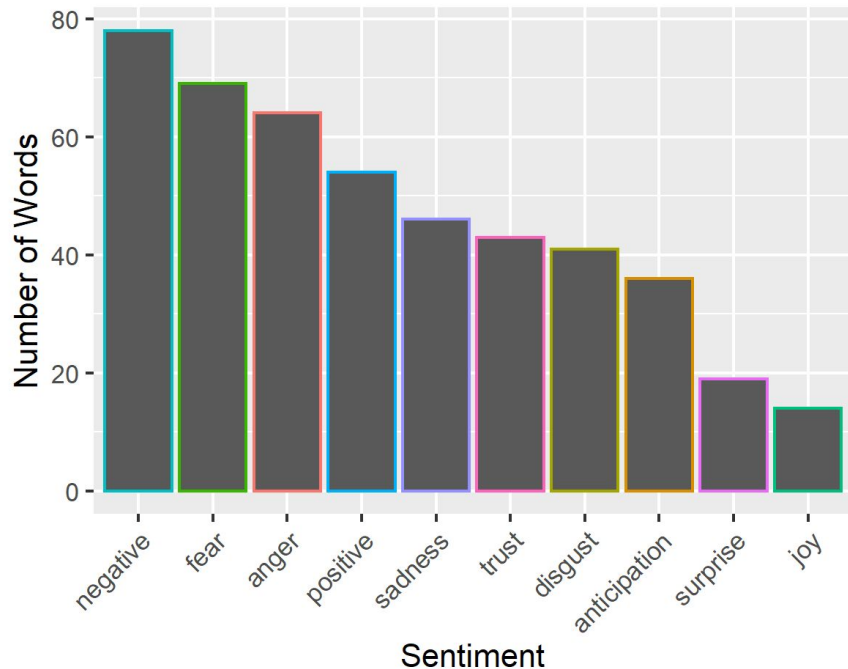
Most Popular Topics : Sentiment Analysis

Most Popular Topics : Sentiment Analysis

Step 1: Calculate affinity for each word in an article by matching with `get_sentiments('afinn')`

Step 2: Calculate overall article affinity using `aggregate()` function

Step 3: Plot the scores!

Afinnity of Most Popular Articles
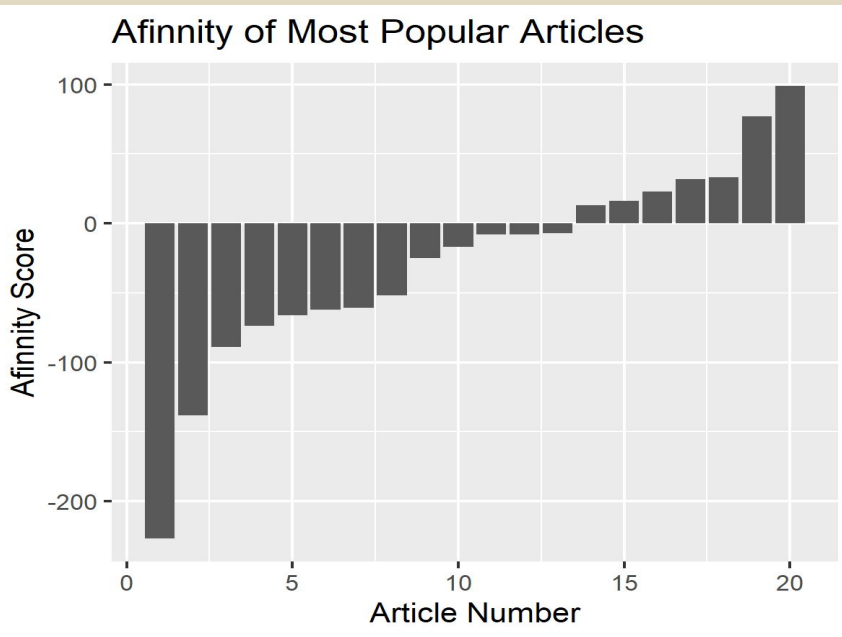
03. PROFESSIONAL
JOHNNY DEPP and
AMBER HEARD
MOVIES

How we analyzed reviews of movies starring Depp and Heard

DATA PROCESSING FOR MOVIE REVIEWS

DATA RETRIEVAL

Retrieved 53 reviews from Movie Reviews API querying for the keywords "johnny+depp" and "amber+heard"

SCRAPED DATA FROM HTML

Extracted html links from JSON and scraped the contents from each review

PREPARED DATA FOR ANALYSIS

Cleaned text, removed stop words, and formatted into a DTM

Movies : Latent Dirichlet allocation
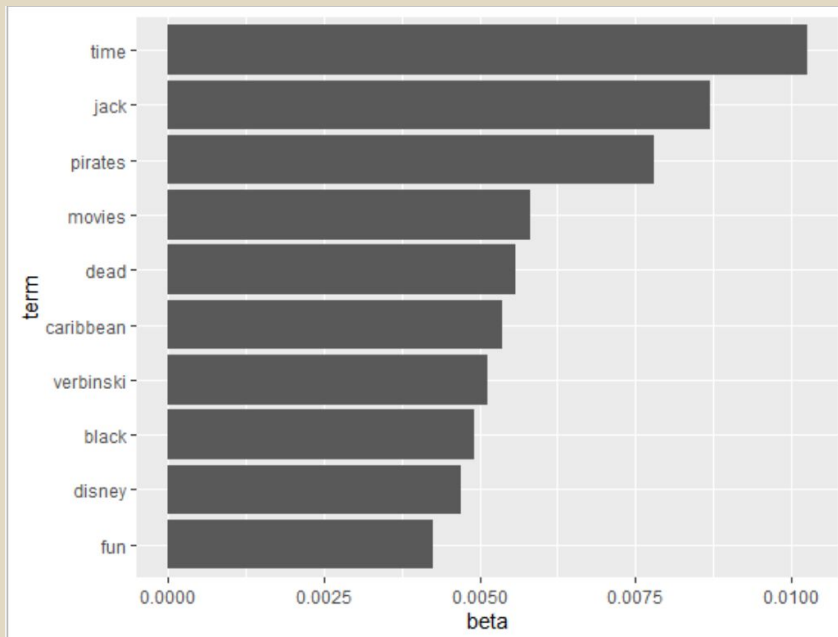
"City of Lies"
"The Professor"
"Transcendence"
"Pirates of the Caribbean: On Stranger Tides"
"Alice in Wonderland"
"Public Enemies"
"Sweeney Todd: The Demon Barber of Fleet Street"
"The Secret Window"
"The Ninth Gate"
"Fear and Loathing in Las Vegas"
"Arizona Dream"
"Don Juan Demarco"
"What's Eating Gilbert Grape"
"Cry Baby"
"Aquaman"
"One More Time"

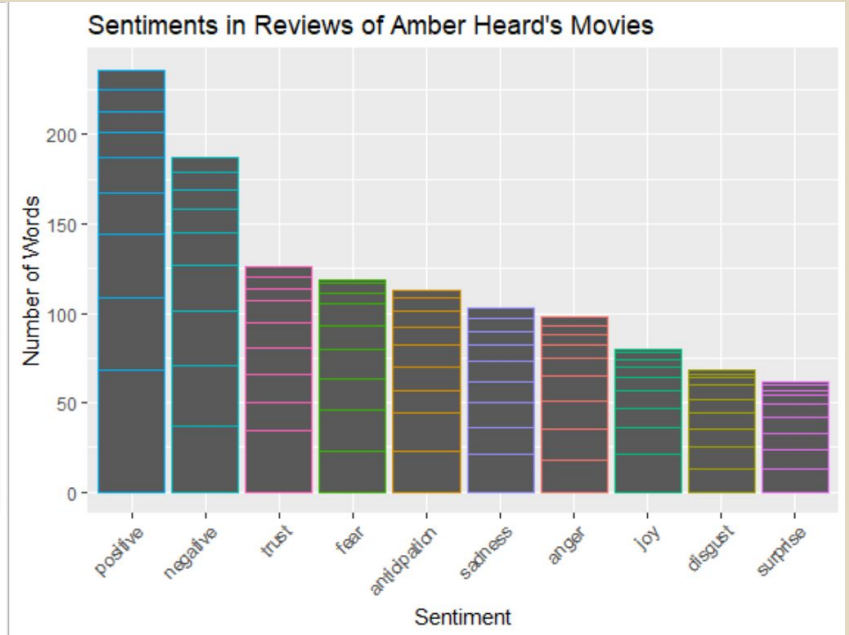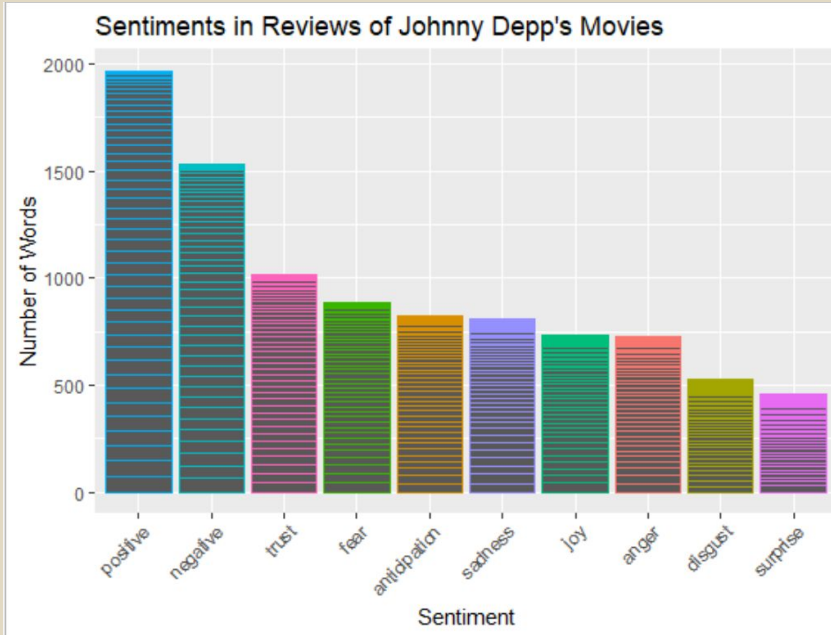Topic 3   Movies in Topic 3

Movies : Latent Dirichlet allocation

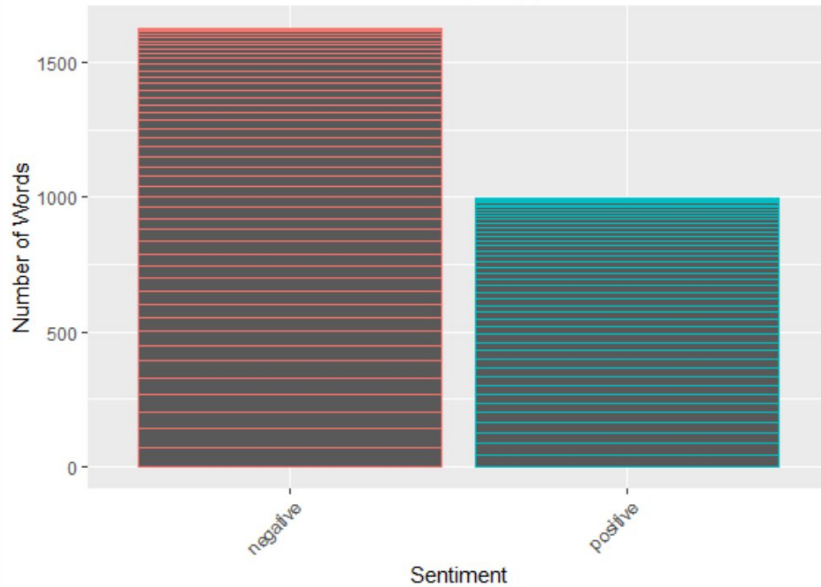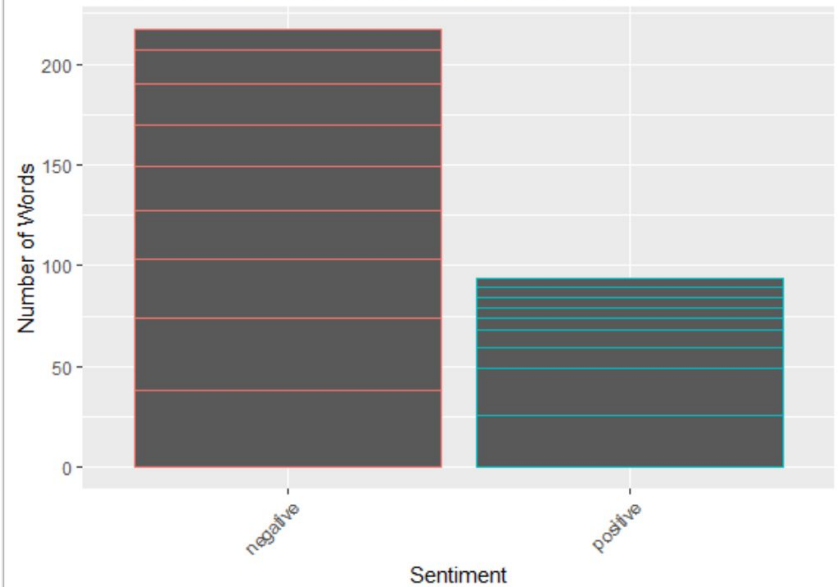| document <dbl>	topic <int>	gamma <dbl>
53	4	0.2641326
53	6	0.2641326
52	2	0.2300195
52	5	0.2300195
51	4	0.3279857
50	2	0.2068376
49	4	0.4364364
48	4	0.2655367
47	6	0.2006803
46	3	0.4603989

Amber Heard's Topics
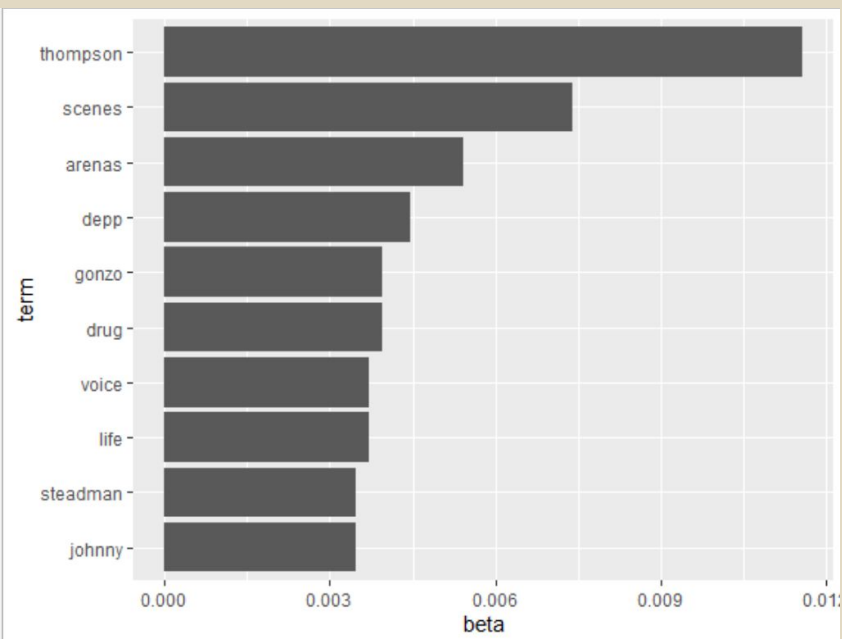
Topic 1

Movies : Sentiment Analysis

Sentiments in Reviews of Johnny Depp's Movies

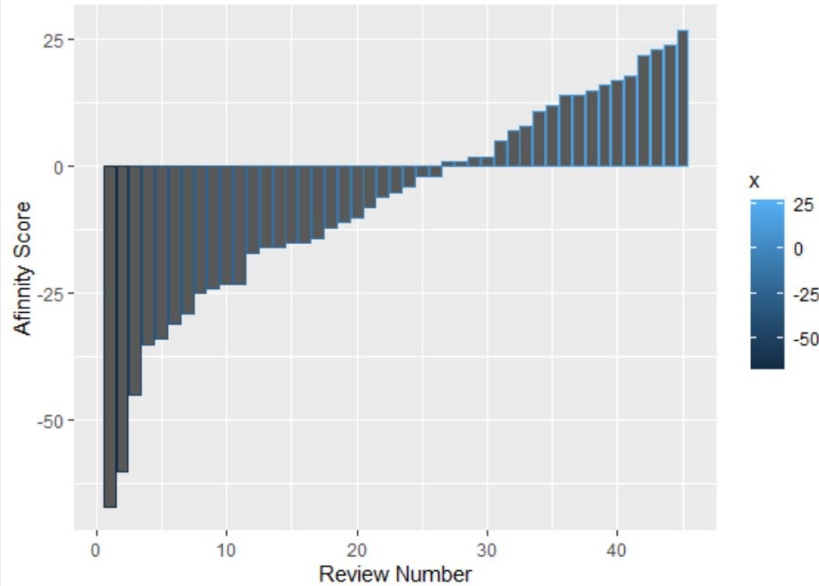Sentiments in Reviews of Amber Heard's Movies

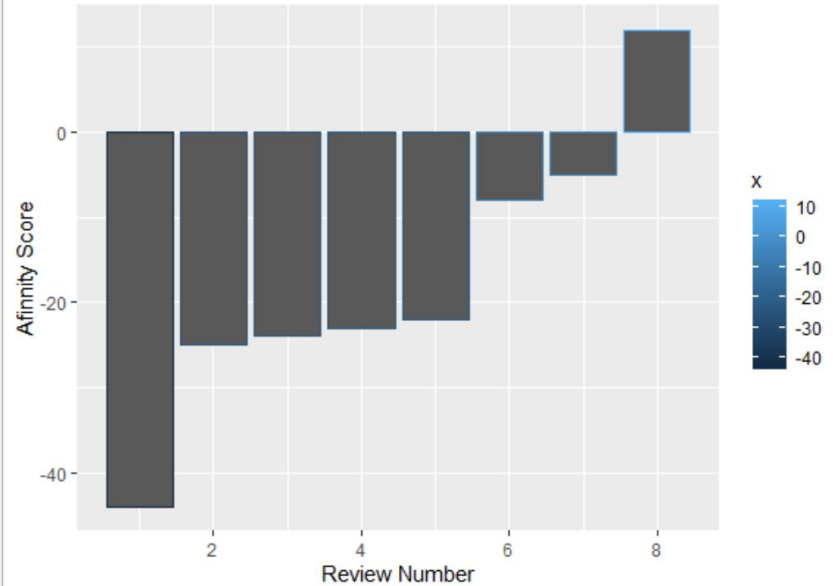# Movies : Sentiment Analysis

Movies : Sentiment Analysis

04. PERSONAL

JOHNNY DEPP and AMBER HEARD
RELATIONSHIP & TRIAL

How we analyzed the news coverage of their relationship over time

DATA PROCESSING FOR Johnny Depp & Amber Heard Articles

DATA RETRIEVAL

Retrieved 35 articles from Article Search API querying for the keywords "johnny+depp+amber+heard" & "johnny+depp+amber+heard+trial" for their relationship since 2015

SCRAPED DATA FROM HTML

Extracted html links from JSON and scraped the contents from each review

PREPARED DATA FOR ANALYSIS

Cleaned text, removed stop words, and unnested tokens

SENTIMENT ANALYSIS FOR Johnny Depp & Amber Heard Articles

COMPUTING SENTIMENTS

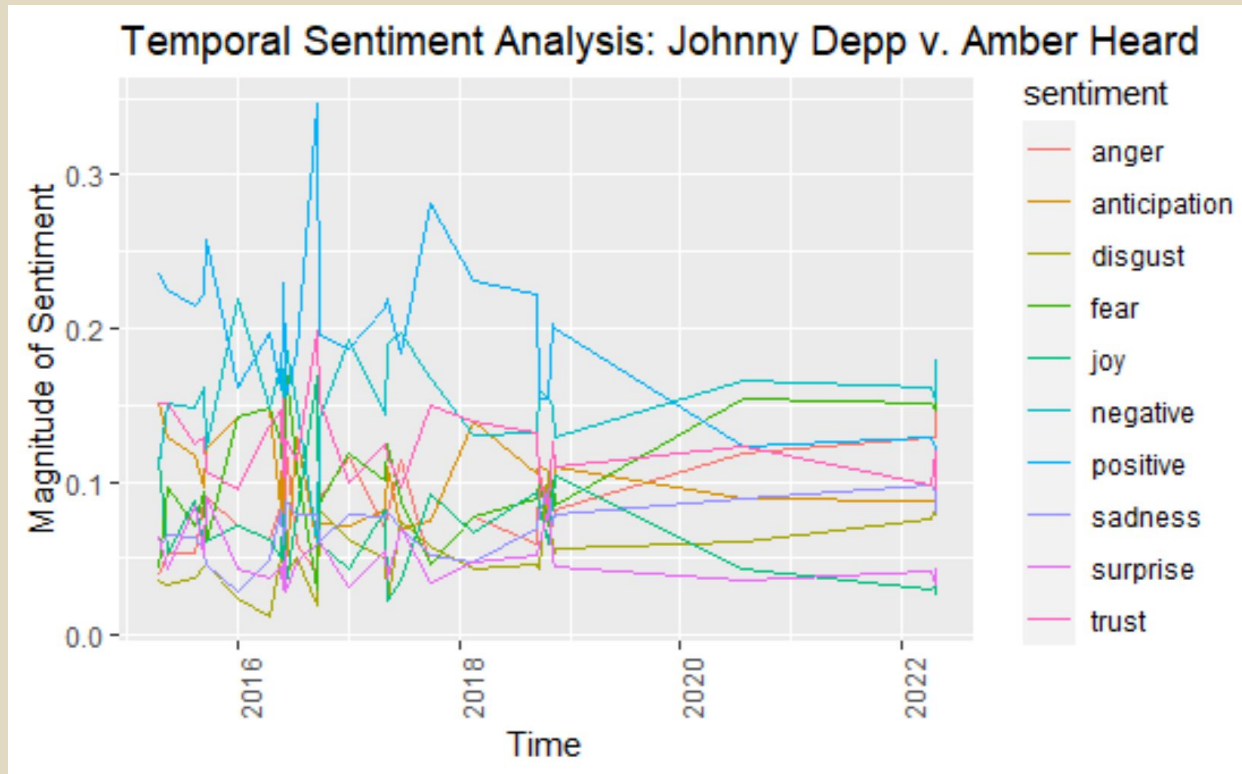Matched NRC and Bing Sentiments with Johnny Depp + Amber Heard articles

THE TIME FACTOR

Merged dataframes with article publication dates and sorted to track sentiment changes progressively

PLOTTING

Plotted proportion for NRC and actual counts for Bing SAs

Relationship & Trial : Sentiment Analysis (NRC)

Temporal Sentiment Analysis: Johnny Depp v. Amber Heard

Relationship & Trial : Sentiment Analysis (Bing)

Temporal Sentiment Analysis: Johnny Depp v. Amber Heard

THANKS!

Any Questions?