

Brexit

Brexit and the news

Content

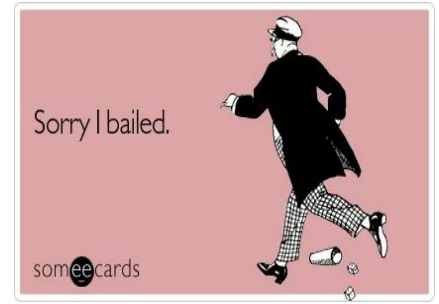
1. Introduction
2. Previous work
3. Data collection
4. Data cleaning
5. Data visualization
 - a. Trend and Date
 - b. Most common words in NY Time
 - c. The opinion of NY Times toward Brexit *
 - d. Difference between different type of article *
6. The results and future work

Introduction

- What is Brexit?
- Why does it matter?
 - Global repercussion
 - Global response
 - Personal interest
- Our Focus
 - How the press react to the brexit?
 - Focus on a single newspaper
 - Focus on New York Times
 - Non-British paper
 - English paper

Initial plan and how it evolved

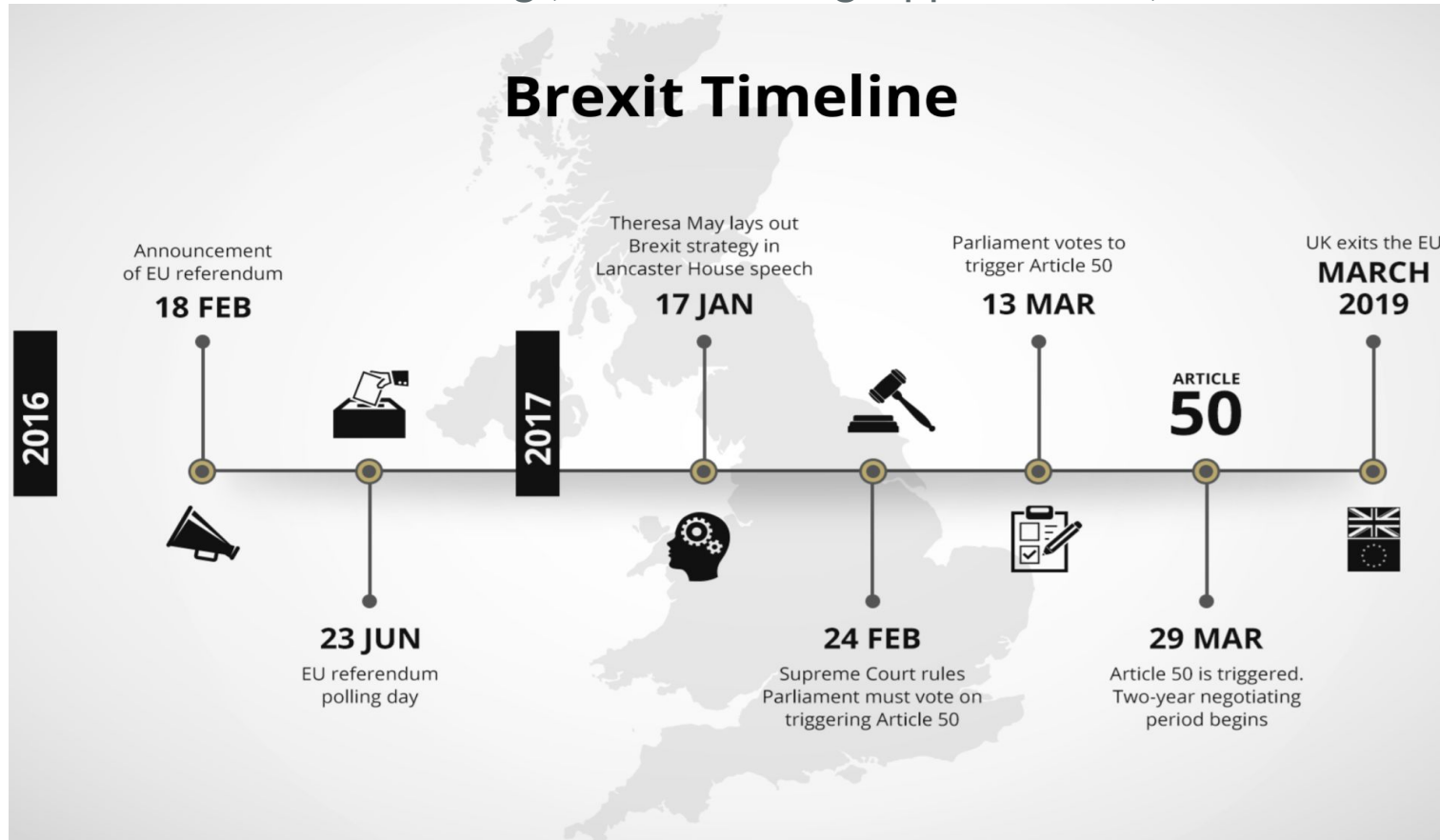
- To tell the tale of Brexit with emojis but Twitter said
- Then we decided to understand the evolution of responses to Brexit from three news sites. -
 - One British
 - One American
 - And a Chinese.
- Then , settled on the American one only. (We speak of the limitations of this at the end)



Previous work

- The Emojis of Brexit
 - Focus on two days (the day before the vote and the day after the vote)
- Social factor that may have influenced “Brexit”
- Predicting Brexit
 - Mining Twitter
 - Natural Language Processing

Source: Financial Trading (Brexit: Trading Opportunities)



Data Collection

NY Time API

1. What are the Times APIs?

Our APIs ([application programming interfaces](#)) allow you to programmatically access New York Times data for use in your own applications. Our goal is to facilitate a wide range of uses, from custom link lists to complex visualizations. Why just read the news when you can hack it?

NYT currently has eleven public APIs: [Archive](#), [Article Search](#), [Books](#), [Community](#), [Geographic](#), [Most Popular](#), [Movie Reviews](#), [Semantic](#), [Times Newswire](#), [TimesTags](#), and [Top Stories](#).

2. Who is the intended audience for the Times APIs?

We've designed our APIs for the web developer community, but all noncommercial users are welcome. See our [Terms of Use](#) for more information.

Data Collection

Difficulties

1. **No existing and useful code**

Unlike Twitter API which we learned from Assignment 2

2. **API Call Limit**

The Article Search API is rate limited to 1,000 calls per day, and 1 call per second.

Data Collection

Demo

```
def parse_articles(articles):  
    '''  
    This function takes in a response to the NYT api and parses  
    the articles into a list of dictionaries  
    '''  
    news = []  
    for i in articles['response']['docs']:  
        dic = {}  
        dic['_id'] = i['_id']  
        dic['headline'] = i['headline']['main'].encode("utf8")  
        dic['date'] = i['pub_date'][0:10] # cutting time of day.  
        if i['snippet'] is not None:  
            dic['snippet'] = i['snippet'].encode("utf8")  
        dic['type'] = i['type_of_material']  
  
        news.append(dic)  
    return(news)
```

Data Collection

Demo

```
def get_articles(date,query):  
    '''  
    This function accepts a year in string format (e.g.'1980')  
    and a query (e.g.'Brexit') and it will  
    return a list of parsed articles (in dictionaries)  
    for that year.  
    '''  
  
    import time  
    all_articles = []  
    for i in range(0,100): #NYT limits pager to first 100 pages.  
        articles = api.search(q = query,  
                               fq = {'source':['The New York Times']},  
                               begin_date = date*10000 + 101,  
                               end_date = date*10000 + 1231,  
                               sort='oldest',  
                               page = str(i))  
        articles = parse_articles(articles)  
        all_articles = all_articles + articles  
        time.sleep(1.1)#The Article Search API is rate limited to 1 call per second.  
    return(all_articles)
```

Data cleaning

Date

headline	locations	snippet	url	subjects	date	type	id
What to Watch for in European Elections	[Europe']	The economy, €	https://rend	[Elections',	41275	Blog	50e2ea6100:
David Cameron's Losing Battle	[Great Britain',	The British prim	https://www	[Elections',	41815	Editorial	53aaa68638f
Keeping Britain in the E.U.	[Great Britain']	There are strong	https://www	[]	41820	News	53afed6a38f
If Scotland Goes, Bankers Fear	[Great Britain',	As Scotland pre	https://dealk	[Referendu	41893	Blog	54123eea38f
Morning Agenda: Banks Bo	[]	Banks eager to	https://dealk	[]	41894	Blog	5412d74238
For Europe, a Year of Politic	[]	Expect the unex	https://www	[]	42038	News	54cfbbcd38f
4 British Politicians Stumble	[Great Britain']	Three months b	https://www	[]	42046	News	54da3b4638
The Suicide of Britain	[Great Britain',	David Cameron	https://www	[Elections']	42133	Op-Ed	554e523738
Can the Union Hang On?	[Czechoslovakia',	Why the United	https://dout	[Referendu	42135	Blog	5551005138
In Post-Election Britain, Tall	[Great Britain']	Prime Minister	https://www	[Referendu	42137	News	5552816838
Britain to Curb Who Can Vo	[Great Britain']	The governmen	https://www	[Referendu	42150	News	5562fd1638f
Cameron Presses E.U. for ü	[Great Britain']	The British prim	https://www	[Internation	42153	News	55676e0938
Why Germany Can't Lead I	[Germany', Euro	It is the most po	https://www	[Politics and	42172	Op-Ed	557ff1eb38f
Far-Right Parties Form Coa	[Europe']	The new bloc, w	https://www	[Legislature	42172	News	55808c1538f
In English, French Paper Ur	[France', Great	In an unusual st	https://www	[Internation	42175	News	55844c8e38f
Why I'm Still Bullish on Eur	[]	The European U	https://www	[Immigratio	42187	Op-Ed	5593bac838f
Advice to Build New Heath	[Great Britain',	The contretemp	https://www	[Airports']	42187	News	559413c838f
Deal on Greek Debt Crisis E	[Greece']	In the name of	https://www	[Euro (Curr	42198	News Analys	55a3b9a438

Data cleaning

Date

- In the Excel, 1900-01-01 in date format equals to 1 in number format; 2013-01-01 in date format equals to 41275 in number format.
- According to this rule, we make a new table.
- And use Table.join() function to convert number to date format.

41275	2013	1	201301
41276	2013	1	201301
41277	2013	1	201301
41278	2013	1	201301

Data cleaning

Type (briefing)



Worldcloud for briefing

Data cleaning

Type (briefing)

```
brexit_date.where("Type", "briefing").select("Snippet")
```

Snippet

Here's what you need to know to start your day.

Here's what you need to know to start your day.

Here's what you need to know to start your day.

Here's what you need to know to start your day.

Here's what you need to know to start your day.

Here's what you need to know to start your day.

Monday: Congressional races, üBrexitü in New York, and M ...

Here's what you need to know to start your day.

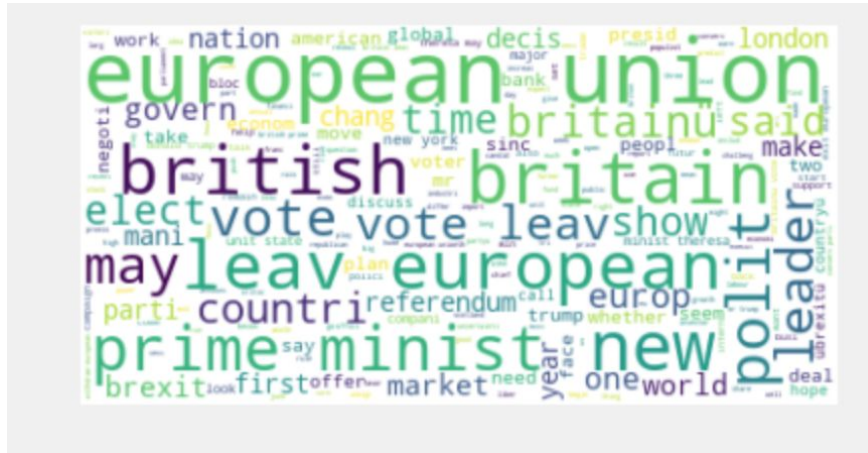
Here's what you need to know to start your day.

Here's what you need to know to start your day.

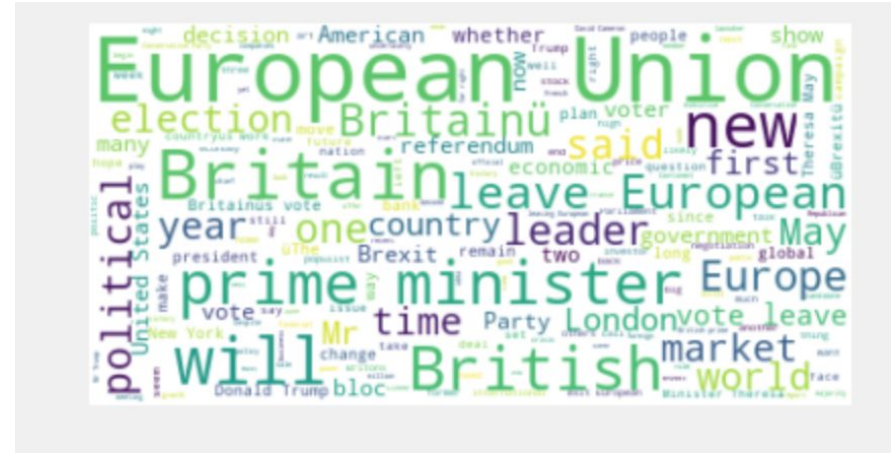
... (170 rows omitted)

Data cleaning

Tokenization, Removing Punctuation & Stopwords, and Stemming



After

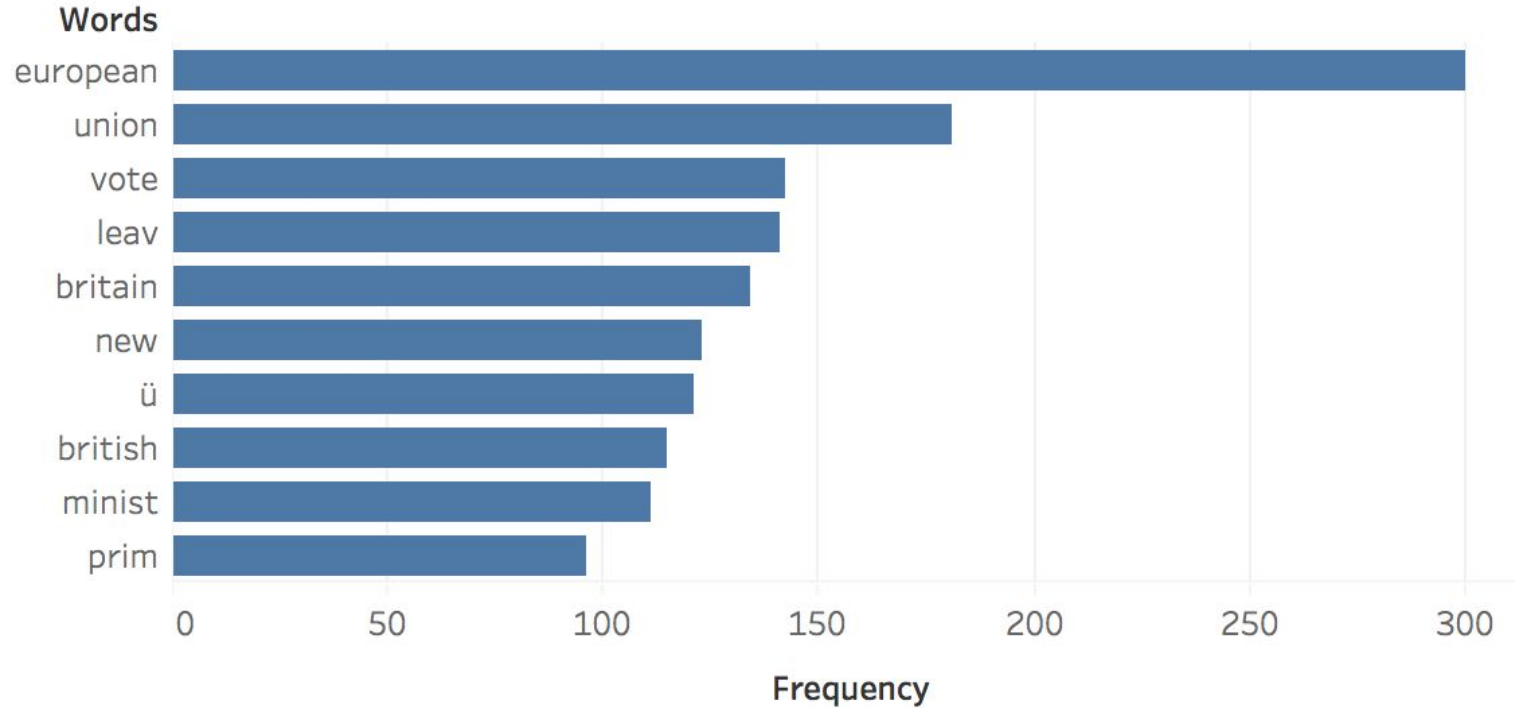


Before

Data visualization

- Tools: Python, Tableau
- Topics:
 - Trend and Date (Python and Tableau)
 - Most common words in NY Time (Python, WordCloud)
 - The Prediction of opinion NY Times held toward Brexit (Python, WorldCloud)
 - Difference between different type of article (Python, WorldCloud)

Data visualisation



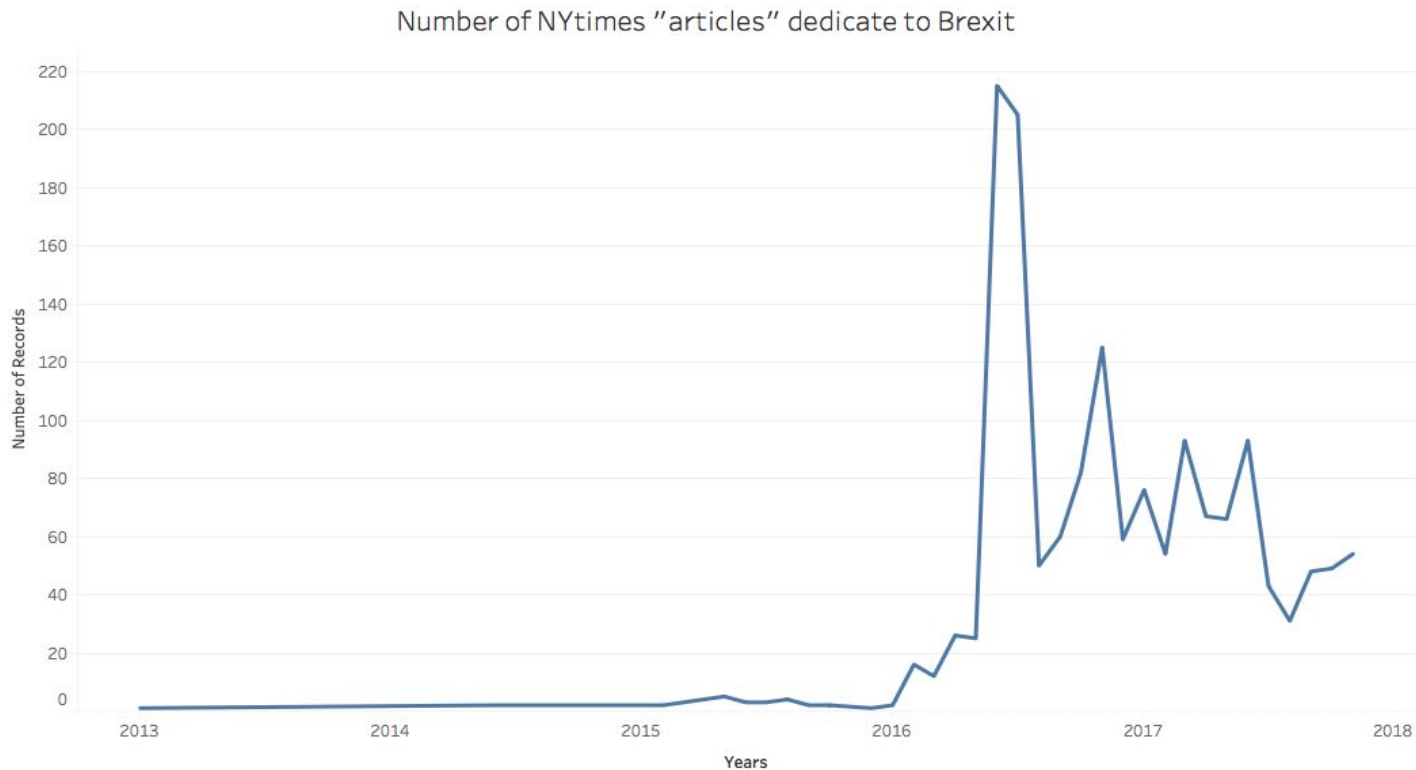
Data Visualization

Most Common Words



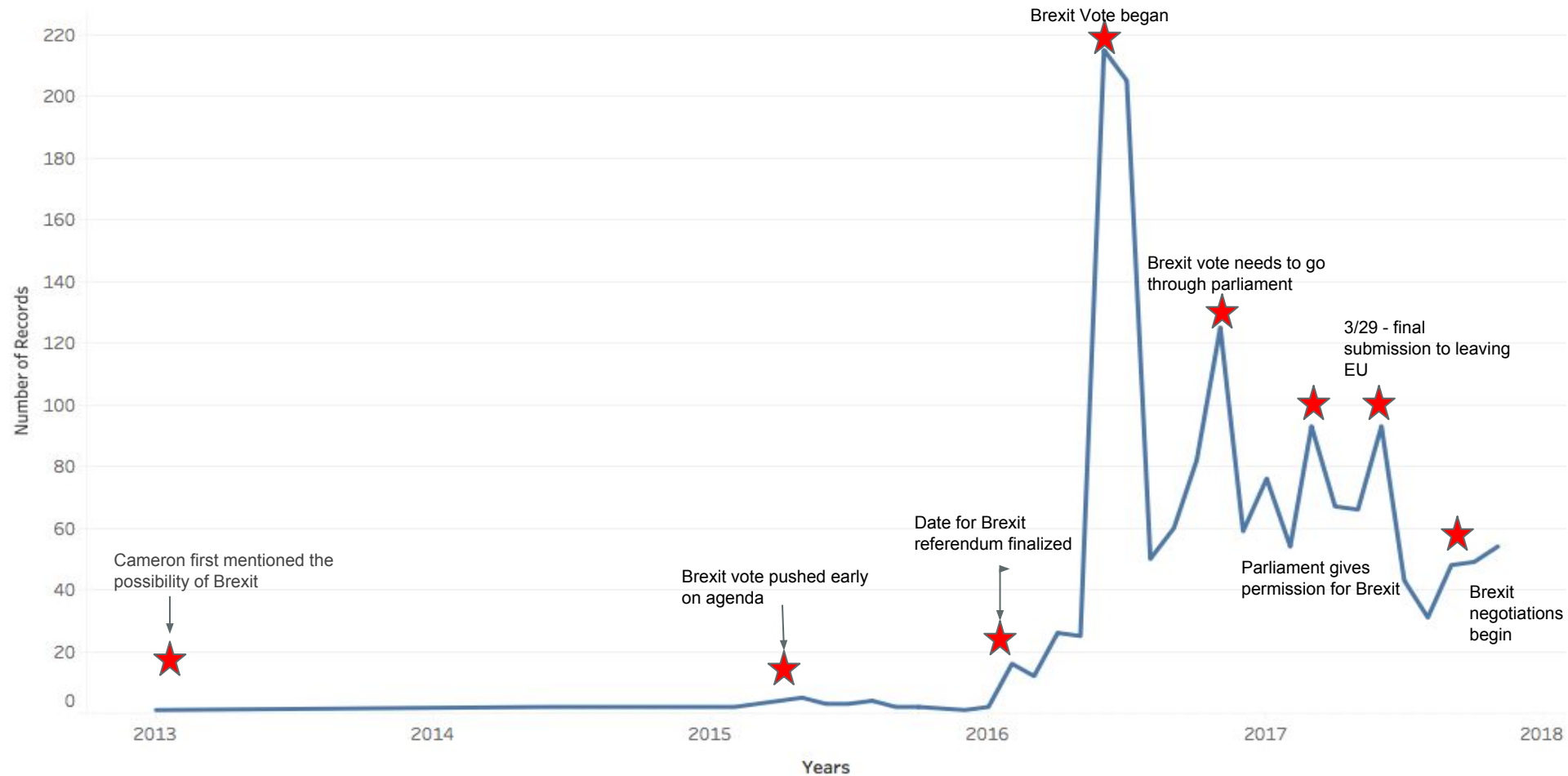
Data Visualization

Trend and Date



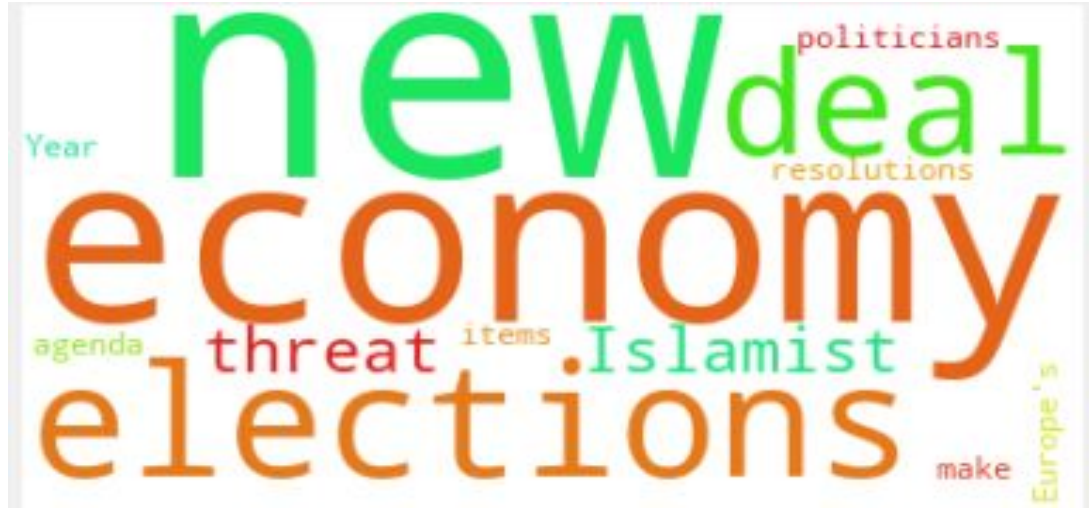
The trend of sum of Number of Records for Date Month.

Number of NYtimes "articles" dedicate to Brexit



Data Visualization

2013



Data Visualization

2014



2015



Data Visualization

2016



2017



Limits

- Data cleaning is an iterative process
- Brexit is a neologism and does not represent the beginning of the talk on UK leaving Europe
- Stemming and Lemmatization are still work in progress
- Limited amount of times of usage
- Most APIs are rate limited to 1,000 calls per day, and 5 calls per second. The Article Search API is rate limited to 1,000 calls per day, and 1 call per second.
- Focus on a larger amount of data will have result in a more accurate result.

Results

- The articles highlight the main concerns around the time of Brexit
- We can see words such as Bankers, Leave, Start, Need , Europe and Britain a lot
- The trend from tableau
- The main thematics are not static and evolve through time

Difficulties

- Struggle with data gathering
- Date conversion (different format)
- Miscommunication in our process to seek help

How to improve?

- Better communication in terms of seeking help
- Determining a wider range of sources to collect from
- We need more practice on gathering data

Work citation

Celli, F., Stepanov, E., Poesio, M., & Riccardi, G. (2016). Predicting Brexit: Classifying agreement is better than sentiment and pollsters. In Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES) (pp. 110-118).

Prismoji. (2016, June 30). <https://prismoji.com/2016/07/01/brexit-emojis/>. Retrieved November 30, 2017, from <https://prismoji.com/2016/07/01/brexit-emojis/>

New York Times (n.d.). The New York Times Developer Network. Retrieved November 24, 2017, from <http://developer.nytimes.com/>

Financial Trading (2017, March 31). Brexit: Trading Opportunities. Retrieved December 01, 2017, from <http://www.financialtrading.com/brexit-trading-opportunities/>

Seaton, J. (2016). Brexit and the media. The Political Quarterly, 87(3), 333-337.

The Guardian (2016, April 22). Barack Obama: Brexit would put UK 'back of the queue' for trade talks . Retrieved December 01, 2017, from <https://www.theguardian.com/politics/2016/apr/22/barack-obama-brexit-uk-back-of-queue-for-trade-talks>

The Guardian. (2017, January 16). Donald Trump's first UK post-election interview: Brexit a 'great thing' . Retrieved from <https://www.theguardian.com/us-news/2017/jan/15/trumps-first-uk-post-election-interview-brexit-a-great-thing>