

Using Survey Data as a Predictor of Pandemic Vaccination

4 - NLP with LDA (soft clustering)

Mark Patterson, March 2021

Introduction to 4: NLP Notebook

The purpose of this phase of my Capstone was to identify current topics of misinformation (within current social media) in order to better target information and education campaigns.

Twint was used to "scrape" Twitter for tweets from the past year (Feb. 2020 to Feb. 2021). I started with negative search terms (covid death; vaccine choice; etc.) as I wanted to find tweets that contained negative and misinformation that might prevent people from getting vaccinated. The TextBlob library was used to assign sentiment to each tweet, words, bi-grams, and tri-grams were examined, and then LDA (soft clustering) was used to look for additional topic clusters. A small set of tweets was read for additional context and understanding.

Note that several iterations of this process occurred with different subsets of the data. Eventually, the set of negative sentiment tweets were the focus of the analysis.

Load appropriate libraries and data set(s)

```
In [1]: # Import the relevant libraries and packages.
```

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns

import re
from wordcloud import WordCloud
from textblob import TextBlob
import gensim
from gensim.utils import simple_preprocess
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
import gensim.corpora as corpora
from pprint import pprint
import pyLDAvis.gensim
from nltk.tokenize import word_tokenize, sent_tokenize
nltk.download('punkt')
import string
from nltk import FreqDist

%matplotlib inline

# Increase column width to display df
pd.set_option('display.max_columns', None)
```

```
[nlkt_data] Downloading package stopwords to /Users/markp/nltk_data...
[nlkt_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /Users/markp/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
In [5]: # Import the csv files and then need to stitch them together.
```

```
dfB = pd.read_csv('data/vacchoice2.csv', sep='\t', lineterminator='\r')
print(dfB.shape)
dfB.head()
```

```
(1670, 36)
```

Out[5]:

	id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls	pho
0	\n1363887309100376067	1.363885e+18	2021-02-22 08:24:08 PST	2021-02-22	08:24:08	-800.0	1.249665e+18	livingnothiding	Tom	@nachimzahawi	Can you still confirm that vacci...	en	0	0	['https://pbs.twimg.com/media/Eu2APN1WQAAiTwl...
1	\n1363871758399524868	1.363872e+18	2021-02-22 07:22:21 PST	2021-02-22	07:22:21	-800.0	1.059953e+18	guggsey	⌚	NaN	If you've had your Covid vaccine, which did yo...	en	0	0	
2	\n1363849779378348035	1.363850e+18	2021-02-22 05:55:01 PST	2021-02-22	05:55:01	-800.0	2.610710e+07	kingofsunshine	Sohail Shah	NaN	Lost someone I loved dearly to COVID today. Pl...	en	0	0	
3	\n1363819329435938816	1.363819e+18	2021-02-22 03:54:01 PST	2021-02-22	03:54:01	-800.0	9.153487e+17	fhussain73	Farzana Hussain	NaN	Let's keep having the conversations. GPs &...	en	0	0	
4	\n1363809630682558464	1.363810e+18	2021-02-22 03:15:28 PST	2021-02-22	03:15:28	-800.0	4.089718e+08	rymmomtaz	Rym Momtaz ریم مومتاز	NaN	4 weeks after receiving 1 jab of AstraZeneca c...	en	0	0	

```
In [6]: dfA = pd.read_csv('data/vacdeath2.csv', sep='\t', lineterminator='\r')
print(dfA.shape)
dfa.head()
```

```
(15001, 36)
```

Out[6]:

	id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls	photos	replies_count	retweets_count
0	\n1363895782475243527	1.363888e+18	2021-02-22 08:57:49 PST	2021-02-22	08:57:49	-800.0	2.753211e+09	mrsrm_senta_ey	Ms M	NaN	@NEUnion It's a death sentence for me. If mine...	en	0	0	0	3.0	2.0
1	\n1363895502627078147	1.363886e+18	2021-02-22 08:56:42 PST	2021-02-22	08:56:42	-800.0	1.331901e+18	lizw75477710	LizW	NaN	@BigBrotherWatch It will start off as just a v...	en	0	0	0	4.0	0.0
2	\n1363895430032031749	1.363895e+18	2021-02-22 08:56:25 PST	2021-02-22	08:56:25	-800.0	9.121231e+08	pickledpuffin	Marion	NaN	Happy that the UK is doing well with their vac...	en	0	0	0	2.0	0.0
3	\n1363889576583065600	1.363883e+18	2021-02-22 08:33:09 PST	2021-02-22	08:33:09	-800.0	9.245927e+17	apk_musicreview	Adam Knight	NaN	@UKCovid19Stats Remember folks this virus had ...	en	0	0	0	2.0	2.0
4	\n1363888280597647370	1.363887e+18	2021-02-22 08:28:00 PST	2021-02-22	08:28:00	-800.0	8.767923e+07	alexberenson	Alex Berenson	NaN	4/ So where are these new cases and deaths com...	en	0	0	0	3.0	18.0

```
In [8]: dfD = pd.read_csv('data/vacrefuse2.csv', sep='\t', lineterminator='\r')
print(dfD.shape)
dfD.head()
```

(6260, 36)

Out[8]:

		id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	u
0	\n1363551422650802176	1.363551e+18		2021-02-21 10:09:27 PST	2021-02-21	10:09:27	-800.0	9.492108e+08	msgeekmedia	Ms. Gesk says let's defeat COVID!!!	NaN	COVID-19 vaccination should be MANDATORY. Both...	en	[]	
1	\n1363471388049309700	1.363392e+18		2021-02-21 04:51:25 PST	2021-02-21	04:51:25	-800.0	1.120361e+18	david_j_bond	David_J_Bond	NaN	@MartinW1510691 @tonyboy pierce @BorisJohnson ...	en	[]	
2	\n1363255097724198914	1.363178e+18		2021-02-20 14:31:58 PST	2021-02-20	14:31:58	-800.0	2.319922e+07	seandlynch	Seán D Lynch	NaN	@leolie o'cananbrennan 11 months failure to pro...	en	[]	
3	\n1363182365510524928	1.363106e+18		2021-02-20 09:42:57 PST	2021-02-20	09:42:57	-800.0	1.765562e+07	aviatorwriting	"Heavy Mayday" Airline Pilot	NaN	@Duguru Airline pilots are, by and large, cons...	en	[]	
4	\n1362933877052895233	1.362934e+18		2021-02-19 17:15:33 PST	2021-02-19	17:15:33	-800.0	4.533734e+08	punt_rd	Punt Road	NaN	Australians banned from the PUB if they refuse...	en	[] ['https://apple.news/AC1lymR2iTnO27Ua2Nr01_	

```
In [10]: dfC = pd.read_csv('data/vacfreedom.csv', sep='\t', lineterminator='\r')
print(dfC.shape)
dfC.head()
```

(7089, 36)

Out[10]:

		id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls	photos	replies_count	retweets_count	like
0	\n1363896935237697539	1.363897e+18		2021-02-22 09:02:23 PST	2021-02-22	09:02:23	-800.0	1.456987e+07	channel4news	Channel 4 News	NaN	"A one-way road to freedom." Boris Johnson se...	en	[] [] []		18.0	7.0		
1	\n1363895096664539145	1.363895e+18		2021-02-22 08:55:05 PST	2021-02-22	08:55:05	-800.0	1.337516e+18	frontinef	Frontline Freedom	NaN	I seriously wish anyone who chooses the vaccin...	en	[] [] []		1.0	1.0		
2	\n1363894517020184580	1.363895e+18		2021-02-22 08:52:47 PST	2021-02-22	08:52:47	-800.0	7.938977e+17	johoakes2020	John EFC #KBF	NaN	THAT ISN'T A ROAD MAP OUT THAT'S JUST PROLONGE...	en	[] [] []		3.0	10.0		
3	\n1363894454625665025	1.363894e+18		2021-02-22 08:52:32 PST	2021-02-22	08:52:32	-800.0	1.337516e+18	frontinef	Frontline Freedom	NaN	Shaking with temper. Quizzed on my vaccine sta...	en	[] [] []		11.0	3.0		
4	\n1363890542799708161	1.363886e+18		2021-02-22 08:36:59 PST	2021-02-22	08:36:59	-800.0	1.314295e+18	britinbulgaria	Brit In Bulgaria	NaN	@bbcnickrobinson Most people are not going to ...	en	[] [] []		0.0	7.0		

```
In [11]: dfE = pd.read_csv('data/vacside2.csv', sep='\t', lineterminator='\r')
print(dfE.shape)
dfE.head()
```

(11600, 36)

Out[11]:

		id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls		
0	\n1363896341668184065	1.363896e+18		2021-02-22 09:00:02 PST	2021-02-22	09:00:02	-800.0	2.592825e+07	webmd	WebMD	NaN	Thinking of taking a painkiller before you get...	en	[] ['https://wb.md/3s9DfU'] ['https://pbs.twimg.com/media/Ei...			
1	\n1363895923798024194	1.363896e+18		2021-02-22 08:58:22 PST	2021-02-22	08:58:22	-800.0	6.975819e+17	teamwolfmoon	Fact-Check Yourself Before You Wreck Yourself	NaN	After a week of being thwarted by truck issues...	en	[] []		0.0	'https://pbs.twimg.c...
2	\n1363887777910243333	1.363888e+18		2021-02-22 08:26:00 PST	2021-02-22	08:26:00	-800.0	2.349127e+08	lukeyism	Big Nas X	NaN	All the 60+ people who take the vaccine saying...	en	[]		0.0	'https://pbs.twimg.com/media/Eu2...
3	\n1363873055039127552	1.363873e+18		2021-02-22 07:27:30 PST	2021-02-22	07:27:30	-800.0	7.510217e+17	gerdiebelt	Gerald	NaN	One day, the future HIV+ peeps would wonder wh...	en	[]		0.0	
4	\n1363867677576466432	1.363868e+18		2021-02-22 07:06:08 PST	2021-02-22	07:06:08	-800.0	3.489067e+08	theromancechica	Sabrina Sol	NaN	My mom & 94-year-old abuela get their seco...	en	[]		0.0	

```
In [16]: # Combine the 5 df with concat.
df_all = pd.concat([dfA, dfB, dfC, dfD, dfE], ignore_index=True)
print(df_all.shape)
df_all.tail()

(41620, 36)
```

conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls	photos	replies_count	retweets_count	likes_count	t
069831e+17	2017-09-10 13:50:05 PDT	2017-09-10	13:50:05	-800.0	1.252452e+07	mercola	Dr Joseph Mercola	NaN	Fact: The U.S. military does NOT track any vac...	en	[] ['http://ow.ly/WkFK30eYr9v']	[]	7.0	50.0	42.0		
057689e+17	2017-09-07 05:25:10 PDT	2017-09-07	05:25:10	-800.0	1.252452e+07	mercola	Dr Joseph Mercola	NaN	Fact: The U.S. military does NOT track any vac...	en	[] ['http://ow.ly/3jfp30eOf90']	[]	4.0	41.0	43.0		
050439e+17	2017-09-05 05:24:25 PDT	2017-09-05	05:24:25	-800.0	2.838925e+07	genekerrigan	Gene Kerrigan	NaN	Dear Flinian ...	en	[] ['http://jrnl.ie/2970847']	[]	5.0	11.0	23.0		
048782e+17	2017-09-04 18:25:37 PDT	2017-09-04	18:25:37	-800.0	2.513103e+09	thesciencenpost	The Science Post	NaN	CDC releases new list of vaccine side effects...	en	[{'screen_name': 'docmeehan', 'name': 'jim mee...'} ['http://thesciencenpost.com/?p=70136']]	[]	0.0	7.0	17.0	['v']	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

```
In [18]: df_cut = df_all
```

Clean and Cut the DF down to useful size

The data frame will contain language = en only.

The data frame will contain tweets from 02-22-20 to 02-22-21 (a one year period).

```
In [19]: df_cut.columns
```

```
Out[19]: Index(['id', 'conversation_id', 'created_at', 'date', 'time', 'timezone',
       'user_id', 'username', 'name', 'place', 'tweet', 'language', 'mentions',
       'urls', 'photos', 'replies_count', 'retweets_count', 'likes_count',
       'hashtags', 'cashtags', 'link', 'retweet', 'quote_url', 'video',
       'thumbnail', 'near', 'geo', 'source', 'user_rt_id', 'user_rt',
       'retweet_id', 'reply_to', 'retweet_date', 'translate', 'trans_src',
       'trans_dest'],
      dtype='object')
```

```
In [20]: # Remove unneeded columns
```

```
df_cut1 = df_cut.drop(columns=['conversation_id', 'created_at', 'time', 'timezone',
       'user_id', 'username', 'mentions',
       'urls', 'photos', 'replies_count', 'cashtags', 'link', 'retweet', 'quote_url', 'video',
       'thumbnail', 'source', 'user_rt_id', 'user_rt',
       'retweet_id', 'reply_to', 'retweet_date', 'translate', 'trans_src',
       'trans_dest'], axis=1)
```

```
df_cut1.head()
```

```
Out[20]:
```

	id	date	name	place	tweet	language	retweets_count	likes_count	hashtags	near	geo
0	\n1363895782475243527	2021-02-22	Ms M	NaN	@NEUnion It's a death sentence for me. If mine...	en	2.0	12.0	[] NaN NaN		
1	\n1363895502627078147	2021-02-22	LizW	NaN	@BigBrotherWatch It will start off as just a v...	en	0.0	13.0	[] NaN NaN		
2	\n1363895430032031749	2021-02-22	Marion	NaN	Happy that the UK is doing well with their vac...	en	0.0	14.0	[] NaN NaN		
3	\n136388976583065600	2021-02-22	Adam Knight	NaN	@UKCovid19Stats Remember folks this virus had ...	en	2.0	26.0	[] NaN NaN		
4	\n136388280597647370	2021-02-22	Alex Berenson	NaN	4/ So where are these new cases and deaths com...	en	18.0	83.0	[] NaN NaN		

```
In [21]: print (df_cut1['place'].value_counts())
print (df_cut1['near'].value_counts())
print (df_cut1['geo'].value_counts())
```

```
{'type': 'Point', 'coordinates': [53.13510132, -6.08067508]} 1
{'type': 'Point', 'coordinates': [37.24261475, -8.62509268]} 1
{'type': 'Point', 'coordinates': [33.97125244, -117.75059351]} 1
{'type': 'Point', 'coordinates': [-7.83783784, 110.32798146]} 1
{'type': 'Point', 'coordinates': [40.400895, -86.80653628]} 1
Name: place, dtype: int64
Series([], Name: near, dtype: int64)
Series([], Name: geo, dtype: int64)
```

```
In [29]: df_cut1['date'].value_counts()
```

```
Out[29]: 2021-02-15    608
2020-12-08    562
2020-12-02    536
2021-01-29    532
2021-01-15    528
...
2019-05-23    1
2019-10-30    1
2018-07-21    1
2017-11-01    1
2018-01-09    1
Name: date, Length: 1064, dtype: int64
```

```
In [28]: df_cut1['geo'].value_counts()
```

```
Out[28]: Series([], Name: geo, dtype: int64)
```

```
In [22]: df_cut1['language'].value_counts()

Out[22]:
en    40693
cs     307
fr     244
und    154
tl      66
ja      41
es      21
hi      14
in      11
ca       7
nl       7
it       6
da       6
et       5
pl       5
tr       5
ro       5
zh       3
sv       2
pt       1
mr       1
el       1
de       1
ne       1
si       1
lv       1
ko       1
no       1
th       1
ar       1
ml       1
sl       1
Name: language, dtype: int64
```

```
In [25]: df_cut2 = df_cut1.loc[df_cut1['language'] == 'en']
df_cut2.shape

Out[25]: (40693, 11)
```

```
In [30]: df_cut2 = df_cut2.drop(columns=['place', 'near', 'geo'], axis=1)
df_cut2.shape

Out[30]: (40693, 8)
```

```
In [31]: df_cut2 = df_cut2.loc[df_cut2['date'] > "2020-02-20"]
df_cut2.shape

Out[31]: (38737, 8)
```

```
In [33]: df_cut2.tail()
```

```
Out[33]:
   id      date        name          tweet  language  retweets_count  likes_count  hashtags
40972 \n1233073200025817088  2020-02-27 Freedomfighter @HighWireTalk Thank God. Research the side eff...
40973 \n1232137475646083072  2020-02-24 The Business Source #Breaking: #Tianjin University has successfull...
40974 \n1232118973912584192  2020-02-24 Global Times #Breaking: #Tianjin University has successfull...
40975 \n1231255789588877314  2020-02-22 Blue🌍 -fshai basically visit a doctor and make sure...
40976 \n1231070253863456768  2020-02-21 Dr. Jaime Friedman This is a really important thread highlighting...
```

```
In [35]: # Check for duplicate rows
duplicate = df_cut2[df_cut2.duplicated()]
duplicate.shape

Out[35]: (754, 8)
```

```
In [36]: # Drop duplicate rows
df_cut2.drop_duplicates(keep='first', inplace=True)
df_cut2.shape

Out[36]: (37983, 8)
```

```
In [37]: print("Columns with null values")
display(df_cut2.isnull().sum())

Columns with null values
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags
0	\n1363895782475243527	2021-02-22	Ms M	It's a death sentence for me. If mine go back...	en	2.0	12.0	[]
1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certifica...	en	0.0	13.0	[]
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vac...	en	0.0	14.0	[]
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% surviva...	en	2.0	26.0	[]
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths com...	en	18.0	83.0	[]

```
In [38]: df_cut3 = df_cut2
```

Begin NLP
The process begins with preprocessing steps.

```
In [39]: # First strip out special characters in the tweets
# Strip out @, #, RT, and url from tweets
def cleantext(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = re.sub(r'#', '', text)
    text = re.sub(r'RT[\s]+', '', text)
    text = re.sub(r'https?:\/\/\S+', '', text)
    return text

df_cut3['tweet'] = df_cut3['tweet'].apply(cleantext)
```

```
In [42]: df_cut3.head()
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags
0	\n1363895782475243527	2021-02-22	Ms M	It's a death sentence for me. If mine go back...	en	2.0	12.0	[]
1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certifica...	en	0.0	13.0	[]
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vac...	en	0.0	14.0	[]
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% surviva...	en	2.0	26.0	[]
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths com...	en	18.0	83.0	[]

```
In [43]: # Add subjectivity and polarity scores from TextBlob
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

df_cut3['subjectivity'] = df_cut3['tweet'].apply(getSubjectivity)
df_cut3['polarity'] = df_cut3['tweet'].apply(getPolarity)
df_cut3.head()
```

```
Out[43]:
      id   date       name          tweet language  retweets_count  likes_count hashtags  subjectivity  polarity
0 \n1363895782475243527 2021-02-22     Ms M It's a death sentence for me. If mine go back...      en           2.0        12.0      []  0.250000  0.000000
1 \n1363895502627078147 2021-02-22     LizW It will start off as just a vaccine certifica...      en           0.0        13.0      []  0.712500  0.362500
2 \n1363895430032031749 2021-02-22     Marion Happy that the UK is doing well with their vac...      en           0.0        14.0      []  0.583333  0.533333
3 \n1363889576583065600 2021-02-22 Adam Knight Remember folks this virus had a 99.7% surviva...      en           2.0        26.0      []  0.684091 -0.052273
4 \n1363888280597647370 2021-02-22   Alex Berenson 4/ So where are these new cases and deaths com...      en          18.0        83.0      []  0.404242  0.020606
```

```
In [44]: df_cut3['polarity'].describe()
```

```
Out[44]:
count    37983.000000
mean     0.065572
std      0.227399
min     -1.000000
25%      0.000000
50%      0.005000
75%      0.187555
max      1.000000
Name: polarity, dtype: float64
```

```
In [45]: df_cut3['subjectivity'].describe()
```

```
Out[45]:
count    37983.000000
mean     0.396192
std      0.269657
min     -0.000000
25%      0.200000
50%      0.427778
75%      0.576627
max      1.000000
Name: subjectivity, dtype: float64
```

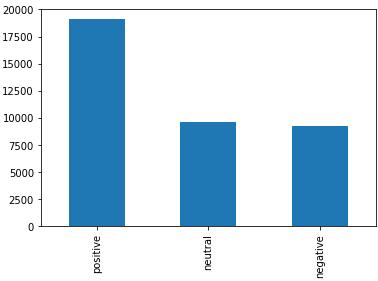
```
In [48]: # Group the tweets by positive, negative, or neutral based on Polarity score
```

```
def getanalysis(score):
    if score <0:
        return "negative"
    elif score ==0:
        return "neutral"
    else:
        return "positive"

df_cut3['Analysis'] = df_cut3['polarity'].apply(getanalysis)
df_cut3.head()
```

```
Out[48]:
      id   date       name          tweet language  retweets_count  likes_count hashtags  subjectivity  polarity  Analysis
0 \n1363895782475243527 2021-02-22     Ms M It's a death sentence for me. If mine go back...      en           2.0        12.0      []  0.250000  0.000000  neutral
1 \n1363895502627078147 2021-02-22     LizW It will start off as just a vaccine certifica...      en           0.0        13.0      []  0.712500  0.362500  positive
2 \n1363895430032031749 2021-02-22     Marion Happy that the UK is doing well with their vac...      en           0.0        14.0      []  0.583333  0.533333  positive
3 \n1363889576583065600 2021-02-22 Adam Knight Remember folks this virus had a 99.7% surviva...      en           2.0        26.0      []  0.684091 -0.052273  negative
4 \n1363888280597647370 2021-02-22   Alex Berenson 4/ So where are these new cases and deaths com...      en          18.0        83.0      []  0.404242  0.020606  positive
```

```
In [49]: # Get percentage of tweets by polarity bucket (aka analysis)
df_cut3['Analysis'].value_counts().plot(kind='bar')
plt.show()
```



```
In [51]: df_cut3['Analysis'].value_counts(normalize=True)*100
```

```
Out[51]:
positive    50.346208
neutral     25.245505
negative    24.408288
Name: Analysis, dtype: float64
```

```
In [53]: view_neg = df_cut3.loc[df_cut3['Analysis'] == 'negative']
```

```
In [59]: view_neg(tweet.sample(20))

Out[59]: 24050 US taxpayer money funded a controversial healt...
13238 If you advocate for herd immunity without a v...
8387 That depends heavily on the adverse effects a...
32009 Hmm.. vaccine with minimal side effects that c...
11603 Covid related hospitalizations and deaths are ...
7798 Guys this is scary : Norway not the first rec...
7157 This vaccine programme in Scotland needs a se...
5933 just a reminder that yesterday California had ...
39830 Before we ask "what are the health impacts of ...
21048 Listening to callers rn: Host: "What's an is...
1762 _ns You'd think the sheer number of deaths and...
16562 Tough choice between the Coronavirus, Trump o...
16941 March 2022 - everyone's gotten 2 doses, at le...
14168 The UK can't do mass testing, failed everybody...
20717 Vaccine? Anyone would think he had shares in ...
34764 A vaccine with potentially bad side effects or...
9925 There is a choice. We can allocate vaccines ac...
32018 I had zero side effects fr Pfizer vaccine I de...
31572 I've gotten both my doses, second one was wor...
1544 It's just so weird, like, our economy is do...
Name: tweet, dtype: object

In [70]: list(view_neg(tweet.sample(20)))

Out[70]: ['I am finding it difficult to discern what may be COVID vaccine side effects and what is just a normal part of being a 32 year old in Ohio, who hasn't seen the sun for 7+ days, during a national coup, a national&gt;global pandemic, in the healthcare field. Ya know?", "the thalidomide tragedy in the 50s/60s led to thousands of deaths and birth defects. it's the reason why the vaccine development and administration process today is so heavily regulated", "British Army's Information Warfare Unit will be deployed to tackle anti-vaccine propaganda ahead of jab rollout Bring it on, I'm not anti vax, just anti covid poison, my body my choice ", "Anti-vaccine pseudoscience peddler Robert F. Kennedy, Jr. used baseball legend Hank Aaron's death in a desperate attempt to fan hysteria over the coronavirus vaccine ", "Never heard of Nipah virus? Well, read this from the : 'The death rate for Nipah virus is up to 75% & it has no vaccine. While the world focuses on Covid-19, scientists are working hard to ensure it doesn't cause the next pandemic.' via Future", "'Vaccination (immediate) side-effects - UK Of 112,807 vaccinated, 3150 people (2.8%) unable to perform normal daily activities, unable to work, required care from doctor or health care professional having received the vaccine. Only the first dose. ', "State vaccination priorities will shape whether vaccines go to Hispanic, Black, & American Indian people disproportionately affected by COVID deaths. Prioritizing essential workers will be key Vaccine priorities: ", "Seems the people who shouted loudest for 9mths about how unfair it's been having their freedoms taken away are now also the ones shouting loudest about not wanting any part of the vaccine designed to give freedom back. All based on their vast scientific qualifications, naturally.", "We prioritized giving the vaccine to the very oldest people. So yeah, some of them died. There's been no link between the deaths of these people and the shot. This guy spent almost a year saying COVID is nbd but now the vaccines are catastrophic? What a fucking lunatic. ", "'Biggest risks to markets in 2021+ IMO: 1) Inflation rears its ugly head and starts printing 4%. Markets will be down 30-40%. 2) Vaccine roll-outs- esp with elderly being targeted first -lead to unexpected deaths/illnesses leading to roll-out suspension. Markets down 20-30%.', "ethics Q: When u go for vaccination, u don't get choice of vaccine. So if they give you unproven vaccine & u refuse, u get nothing. Your choice is take bad science vaccine or risk covid. If u refuse, ur number may not come again. So get pressurized for vaccine without trials", "Mike thinks these children are throbbing bellends who deserve to suffer. Mike. They are suffering. Missed school. Lack of healthcare. A future riddled with corrupt politicians hellbent on filling their little bodies with an experimental vaccine or lose their freedom. ", "'If everyone else gets vaccine, there is no reason to wait for kids to open schools. 60M kids under 15 in the U.S. During the 10m, March - Dec 2020: 26,214 tragic deaths from all causes 2,384, from accidental injury 257, from flu 109, from covid', "Why the anti-lockdown "herd immunity" argument has aged so poorly: 1. Sweden's U-turn (after a death rate *ten times* higher than its neighbours). 2. The vaccine rollout 500,000 people vaccinated in the UK already. This is no time to "let the virus rip".', "Can't get a vaccine down here in Matt Gaetz country, but they're lining up for Presidential Medals of Freedom at Publix ", "I INSIST that our "leaders" here in Australia get the vaccine first. After all what's the loss if it's no good for you and creates all sorts of side effects?", "How it says on the package it causes cancer and death! Not sure how the vaccine side effects could be worse ", "MEK Iran Maryam Rajavi called the delay in purchasing and importing the COVID19 vaccine the continuation of the regime's criminal policy, which has so far resulted in the death of 193,000 of our compatriots due to coronavirus. ", "Please don't believe that COVID only attacks people with weak immunity. Ask doctors who have seen the peak of this pandemic. If vaccine prevents COVID and has no side-effects, my opinion is to go for it. ", "Why isn't son-of-a-bus-driver demanding vaccine priority for drivers? Data suggests driving a London bus is *the riskiest profession in the country*, with 51 deaths from Covid so far, 20 since July. Is he scared that would draw attention to his failure to protect them?']

In [ ]:

In [194]: # Take a look at a NEUTRAL set
view_neutral = df_cut3.loc[df_cut3['Analysis'] == 'neutral']
list(view_neutral(tweet.sample(25)))

Out[194]: ['Doctors Say CDC Should Warn People the Side Effects From COVID Vaccine Shots Won't Be 'a Walk in the Park' ", "UPDATE The PfizerBioNTech Covid-19 vaccine began leaving the company's Michigan factory on Sunday, set to be injected into the arms of millions of the most vulnerable Americans as the global death toll topped 1.6 million Kalamazoo ", "The Bill Gates funded World Health Organisation (WHO) has told South Africa to use the Ineffective AstraZeneca Vaccines. Earlier Mkhize said SA won't return these vaccines, which have undisclosed side effects. Trust Western Puppets, Cyril and Mkhize to gamble with Citizens Lives.", "'Bird86 Prevent any restrictions on those who refuse a Covid-19 vaccination ', "Pfizer coronavirus vaccine side effects and safety for COVID19Vaccine ", "' If you CHOOSE to have the vaccine then you are protected. Let people CHOOSE!! If the choice is not to, that's their choice! It's called freedom of choice, another freedom we will lose. Where does it end if they are allowed to force something into your body against your will???'", "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ", "' Manufacturing back. 3. Cut red tape with the FDA to get therapeutics and vaccines approved in record time. What else would Biden have done different? They y should not impose mask mandates. This is AMERICA. Where your fear begins, my rights don't end.', "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ", "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ", "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ", "Thank you. You can definitely have my vaccine 🌟", "'A truck came and announced that anyone who wanted the COVID vaccine and Rs 750 for taking it could go to People's University Hospital and take it. I needed the money:" Man Singh Parihar, 74. reports from Bhopal after a trial participant's death ", "'ALERT! Scamdemic Plans Leaked' The Commies plan on using Covid to totally lockdown countries and offer debt relief & freedom to anyone who uses vaccines. Anyone who disagrees is later moved to a permanent lockdown facility. ", "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ", "'Staffer That's what the national guards are for. Vaccines are life and death so access to them should be treated as such.', "You're advocating a health passport Mark. It's discriminatory. Nothing to do with the freedom to have the vaccine mate.", "'me, yesterday wanting to wait a year or two to get the covid vaccine so I can suss out any side effects vs me, today seriously considering taking the vaccine so I can see Louis, when I say I would risk it all for him, I really meant it', "If you have a death wish, take the vaccine. ", "Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination Come on people of Port Talbot...sign it!!!", "Covid vaccine: You can't sue Pfizer or Moderna over side effects ", "'Zimbabwe plans to buy sufficient coronavirus vaccines to inoculate about two-thirds of its population, amid a surge in infections and deaths ', "'If you die, for whatever reason, within 28 days of having the Covid vaccine, what is the cause of death?', "' THIS is what we will be seeing in the US after Biden is sworn in for those who REFUSE to take the COVID vaccination!', "Get my Covid vaccine shot 1 today. No side effects so far.....and vaccine given by my friend Laurel! ']
```

```
In [195]: # Take a look at a POSITIVE set
view_positive = df_cut3.loc[df_cut3['Analysis'] == 'positive']
list(view_positive.tweet.sample(25))
```

Out[195]: ["This is not the first time my church has failed to protect it's members. I won't be silent and allow such misinformation to spread leading to more deaths. factsmatter Bishop tells Catholics not to jump on 'vaccine bandwagon' - Los Angeles Times ",
'The flu vaccine is different every year but it still works and side effects are minimal. They know how to make these things. Spreading skepticism is really not helping right now when we're already going to have problems with distribution and vaccination.' ,
"So, I'm about 36 hours out from my 2nd vaccine. Side effects: incredibly sore arm, achy muscles in chest, back, & arms. Joint pain & stiffness in my fingers. Some fatigue. The second shot was definitely worse than the first, but still not horrible. I've had worse flu shots.",
'Two lines I see every morning on my ride into work: COVID testing and a food pantry. The vaccine is here and that's fantastic but this pandemic is still far from over. Cases still skyrocketing and record number of deaths across the country.',
'"The three vaccines that are approved in Ireland are all really safe" - Dr Yvonne Williams There have already been reports of deaths & severe reactions around the world from the 3 vaccines that the Doctor is talking about Pfizer/Moderna/AstraZeneca. Is she unaware or lying?!'",
"_tilson That's not at all what this is dude. This is weaponizing religion to convince others not to take the vaccine that will prevent so many deaths.",
'_ also if the vaccine is 85% effective and we have 90% take up then 24% could still get it of the most vulnerable which is 3.4million, lets say 0.5% death rate that 16,800 extra dead in that category alone let alone the hospital bed take up etc',
'Big pharma wants to make money and they won't be denied. Also Bill Gates said "life will never be the same." Unless we promise to be good and take his rushed, experimental, untested, vaccine with the tracking chip. He says we can have our freedom back if we do that.',
'CDC: *recommends COVID-19 vaccine* Me: "I just don't understand how people in the US can ignore science and public health recommendations" CDC: *recommend s against eating raw cookie dough* Me: "I can make my own choices THANKYOUVERYMUCH",
'Some people on here think Trump conceding & the start of the formal transition is some interesting parlor game. It's not. It scares me to death that the Chucklehead crew at the White House aren't giving Biden's team access to info they'll need to make vaccine distribution smooth.',
'So. Covid vaccine update. No side effects so far. Almost 24 hrs post-shot. Arm wasn't particularly sore yesterday until right before i went to sleep, it is now currently VERY sore and tender.',
'If you are worried about the side effects of the vaccine, let me tell you about the side effects of naturally acquired immunity... Oh, and these numbers are made up. Here's the truth: "4 per cent of people reported fatigue and 2 per cent a headache." ',
" My wife's uncle had 1st jab three weeks ago. Developed nausea and headaches the day after. Had a huge stroke last Sunday and died on Tuesday. Coincidence? Don't know and will never know. Perhaps should be counted as a vaccine death given he died within 28 days of the jab!!",
' Might not prevent transmission but the whole point of the vaccine is to teach your body how to defend itself against the antigens of COVID-19. Vaccine = increased protection/less deaths! Less strain on NHS, more chance of actually hugging your nan without the fear of killing her',
'Vaccine Passports: Is Showing Your "Papers" the "New Normal?" ...individuals who refuse the COVID-19 vaccination might "face some limits"...,
' ER, I am a healthy 53 yr old and I am not going near that vaccine. You watch, there will be some profound side effects to this thing they rushed through.',
'A forced vaccine would make me feel unsafe in my own country. Without freedom of choice there is nothing. I'm already beginning to regret bringing my beautiful daughter into this f*cked up world.',
'The powers that be are lying to us. The so called wonder vaccines will not give us our freedom back. We will never be allowed to go back to the old normal... Governments have lied to us from the beginning when they said give us two weeks to flatten the curve.',
'2021 will be exactly like 2020 just with fewer restaurants and small businesses, more foreclosures, more homeless, more repossessions, more suicides, more deaths from treatable cancers and a vaccine that will be found to be ineffective.',
' How can you justify the deployment of a vaccine that hasn't been trialled for long term side effects? The "running tests in parallel" narrative won't work because no matter what, the virus hasn't been around for more than a year.',
'The good news piles up - Vaccines are working as over-80s deaths drop twice as quickly as under-65s ' ,
'UK Column News - 12th February 2021. So far in uk: 143 deaths straight after the Pfizer vaccine, spontaneous miscarriages and care home deaths are not even counted. ',
' Governor Believes in Jesus Christ yet is wanting to not take a vaccine, probably doesn't wear a mask either. love thy neighbor unless it involves a vaccine or a piece of fabric over their mouth. Then it's "muh freedom! Ugh China!"Hypocritical trash.',
'There is a lot more detailed information about COVID19 vaccines in this letter from the UK Medical Freedom Alliance to , the JCVI and , listing serious issues regarding the safety of the COVID19 Vaccine. Please RT. ',
'Less than half the vaccines distributed to states have been used. US about to surpass 400,000 deaths. The pandemic is far from over. The latest, here. ']

In []:

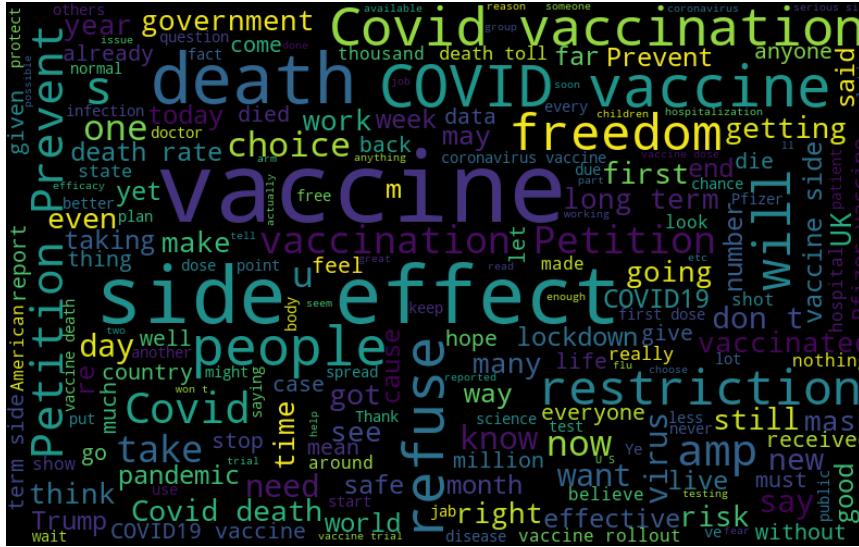
```
In [81]: # ISSUE: Difficulty with text and df flowing off the screen. Tried several possible things... but not able to resolve it  
# pd.set_option('display.max_colwidth', None)
```

```
In [196]: # def display_text_max_col_width(df, width):
#     with pd.option_context('display.max_colwidth', width):
#         print(df)

# display_text_max_col_width(df_cut3['Analysis']) == 'negative'
```

In []:

```
In [47]: # Create word cloud of top occurring words - This is the full dataset (all polarities); also prior to stop word  
allwords = ''.join([twts for twts in df_cut3['tweet']])  
  
plt.figure(figsize=(16,10))  
wordcloud = WordCloud(width=800, height=500, random_state=32, max_font_size=90).generate(allwords)  
plt.imshow(wordcloud, interpolation='bilinear')  
plt.axis('off')  
plt.show()
```



```
In [60]: df_cut4 = df_cut3
```

Begin LDA process

```
Out[61]: 0    it's a death sentence for me if mine go back ...
1    it will start off as just a vaccine certifica...
2    happy that the uk is doing well with their vac...
3    remember folks this virus had a 997% survival...
4    4/ so where are these new cases and deaths com...
Name: text_processed, dtype: object
```

```
In [62]: # Prepare the data for LDA analysis - adding in some additional stopwords based on the wordcloud above
```

```

stop_words = stopwords.words('english')
stop_words.extend(['Covid', 'covid', 'vaccine', 'vaccination', 'rt', 'amp', 'http', 'Vaccine', 'Vaccination', 'Covid vaccination', 'Covid vaccine'])

def sent_to_words(sentences):
    for sentence in sentences:
        # deacc=True removes punctuation
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))

def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc))
             if word not in stop_words] for doc in texts]

data = df_cut4.text_processed.values.tolist()
data_words = list(sent_to_words(data))
# remove stop words
data_words = remove_stopwords(data_words)
print(data_words[1][0][:30])

```

```
['death', 'sentence', 'mine', 'go', 'back', 'th', 'march', 'exposed', 'yrr', 'cev', 'th', 'march', 'means', 'go', 'back', 'vulnerable', 'point', 'announced', 'survival', 'elite', 'fittest', 'screw', 'everybody', 'else']
```

```
In [63]: # Create dictionary and corpus
import gensim.corpora as corpora
# Create Dictionary
id2word = corpora.Dictionary(data_words)
# Create Corpus
texts = data_words
# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]
# View
print(corpus[:1][0][:30])
```

```
In [64]: # For the first model, am using 10 topics.
# Train the model
from pprint import pprint
# number of topics
num_topics = 10
# Build LDA model
lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                         id2word=id2word,
                                         num_topics=num_topics)

# Print the Keyword in the 10 topics
pprint(lda_model.print_topics())
doc_lda = lda_model[corpus]

[(0,
  '0.032*"prevent" + 0.032*"restrictions" + 0.031*"refuse" + 0.030*"petition" +
  '0.025*"deaths" + 0.008*"effects" + 0.008*"vaccines" + 0.008*"side" +
  '0.007*"people" + 0.006*"get"),
 (1,
  '0.020*"vaccines" + 0.012*"deaths" + 0.010*"us" + 0.010*"death" +
  '0.009*"side" + 0.009*"effects" + 0.007*"getting" + 0.006*"get" +
  '0.006*"people" + 0.006*"first"),
 (2,
  '0.018*"refuse" + 0.017*"people" + 0.017*"get" + 0.015*"vaccines" +
  '0.015*"side" + 0.011*"effects" + 0.009*"virus" + 0.008*"death" +
  '0.008*"health" + 0.007*"restrictions"),
 (3,
  '0.026*"petition" + 0.023*"prevent" + 0.022*"refuse" + 0.021*"restrictions" +
  '0.020*"people" + 0.012*"vaccines" + 0.011*"deaths" + 0.009*"death" +
  '0.008*"choice" + 0.008*"freedom"),
 (4,
  '0.021*"vaccines" + 0.015*"side" + 0.015*"effects" + 0.015*"people" +
  '0.012*"deaths" + 0.009*"get" + 0.008*"death" + 0.007*"day" + 0.005*"refuse" +
  '0.004*"health"),
 (5,
  '0.093*"effects" + 0.090*"side" + 0.018*"vaccines" + 0.009*"long" +
  '0.008*"term" + 0.006*"people" + 0.006*"know" + 0.005*"immune" + 0.005*"one" +
  '0.004*"get"),
 (6,
  '0.040*"effects" + 0.035*"side" + 0.013*"first" + 0.009*"know" +
  '0.009*"dose" + 0.008*"death" + 0.008*"us" + 0.008*"got" + 0.007*"one" +
  '0.006*"yesterday"),
 (7,
  '0.038*"effects" + 0.036*"side" + 0.013*"people" + 0.011*"vaccines" +
  '0.008*"deaths" + 0.008*"new" + 0.007*"death" + 0.007*"refuse" +
  '0.006*"first" + 0.006*"trials"),
 (8,
  '0.025*"side" + 0.025*"freedom" + 0.021*"effects" + 0.016*"take" +
  '0.011*"get" + 0.010*"vaccines" + 0.010*"like" + 0.009*"people" +
  '0.009*"death" + 0.008*"anti"),
 (9,
  '0.022*"freedom" + 0.015*"deaths" + 0.012*"side" + 0.011*"effects" +
  '0.011*"vaccines" + 0.008*"death" + 0.006*"people" + 0.006*"phase" +
  '0.005*"us" + 0.005*"effective")]
```

```
In [65]: # Code to generate the interactive visualization
pyLDAvis.enable_notebook()

vis = pyLDAvis.gensim.prepare(lda_model, corpus, id2word)
vis
```

/Users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/pyLDAvis/_prepare.py:257: FutureWarning: Sorting because non-concatenation axis is not aligned. A future version of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.

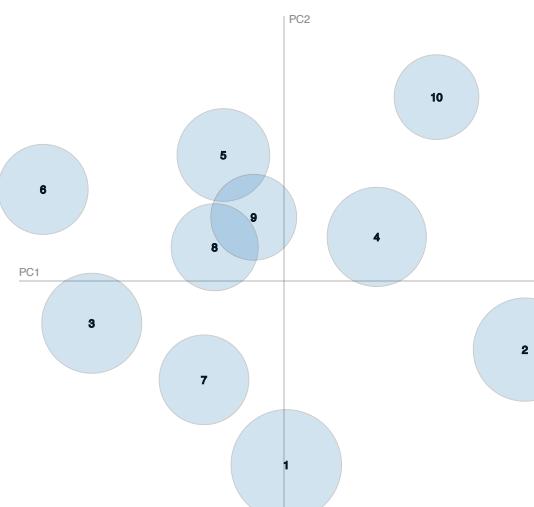
```
return pd.concat([default_term_info] + list(topic_dfs))
```

Out[65]: Selected Topic: 0 | Previous Topic | Next Topic | Clear Topic |

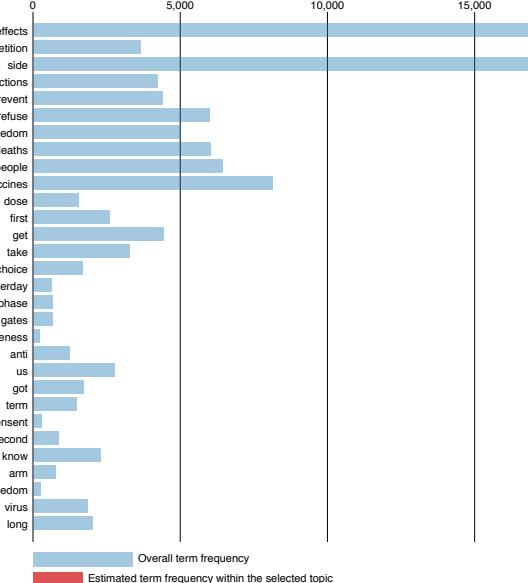
Slide to adjust relevance metric: 0.0 0.2 0.4 0.6 0.8 1

$\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)

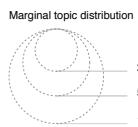


Top-30 Most Salient Terms¹



1. saliency(term w) = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] for topics t; see Chuang et al (2012)

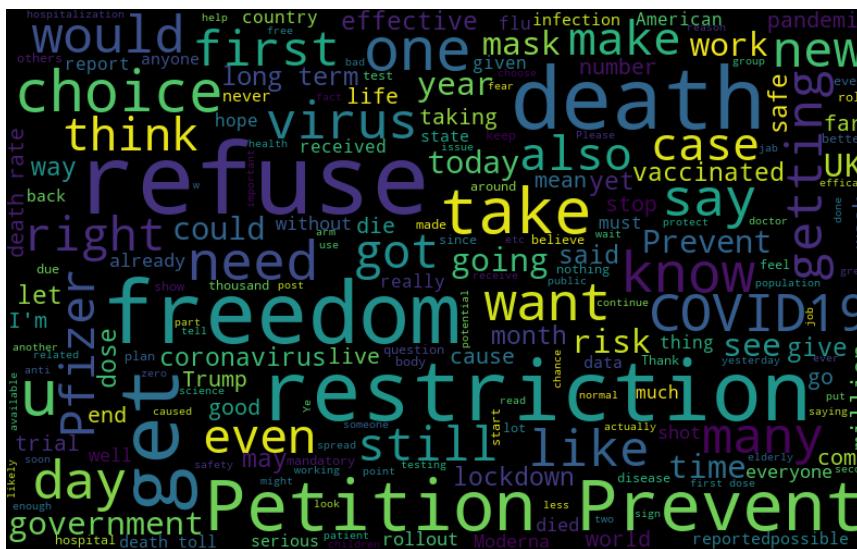
2. relevance(term w | topic t) = $\lambda * p(w|t) + (1 - \lambda) * p(w|t)p(w)$; see Sievert & Shirley (2014).



```
In [68]: # Another word cloud (full dataset) with stop words removed.  
stop_words = stopwords.words('english')  
stop_words.extend(['Covid', 'covid', 'vaccine', 'vaccination', 'rt', 'amp', 'http', 'Vaccine', 'Vaccination', 'Covid vaccination', 'Covid vaccine'])
```

```
In [69]: # Create word cloud of top occurring words
# allwords = ' '.join([twt for twt in df_cut3['tweet']])
# allwords = ' '.join([twt for twt in df_cut3['text']])

plt.figure(figsize=(16,10))
wordcloud = WordCloud(stopwords=stop_words, max_words=200, width=800, height=500, random_state=32, max_font_size=90).generate(allwords)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



Observations on LDA and word cloud - for full dataset

Observations on LDA and word cloud - for full dataset
Can see that with 10 clusters there was some overlap of topics (5,8,9). So fewer topics may be better. The top occurring words overall were no surprise. There is a lot of repetition of words between the clusters so hard to id specific themes for each topic cluster. In the word cloud we see lots of words around a petition (to prevent restrictions): freedom and choice; death (related to side effects).

In [1]:

Conduct the analysis with just the negative set of tweets

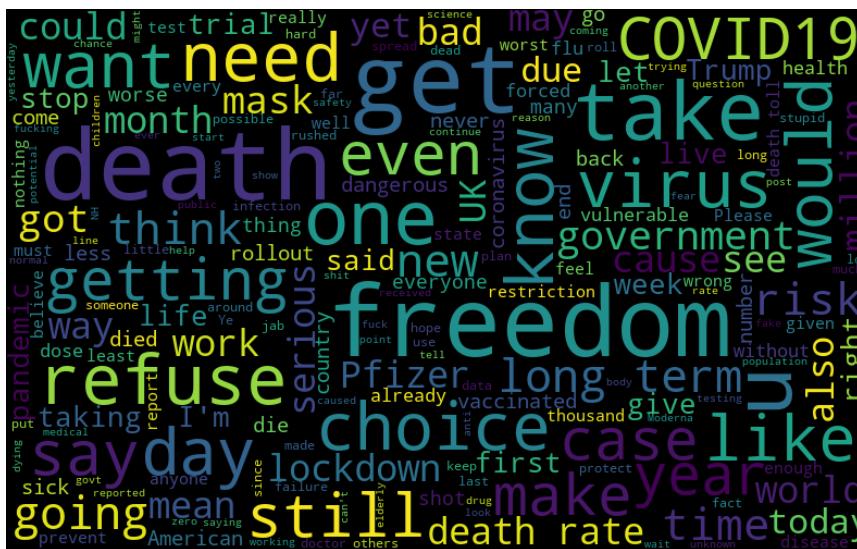
Decided that I should further focus into just the tweets with negative sentiment (according to TextBlob classification).

```
In [82]: df_cut5 = df_cut3.loc[df_cut3['Analysis'] == 'negative']
df_cut5.head()
```

Out[82]:		id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% survival...	en	2.0	26.0]	0.684091	-0.052273	negative	remember folks this virus had a 99.7% survival...
6	\n1363887707076837376	2021-02-22	kerry Hooper	Sick in a bag moment IDS wanting praise &...	en	7.0	14.0	['borisjohnson', 'road2recovery', 'roadmap', ...]]	0.407937	-0.404762	negative	sick in a bag moment ids wanting praise &...
7	\n1363886935052324866	2021-02-22	Lawrence Gilder	DailyCovidUpdate 22nd February 2021 - Cases...	en	70.0	115.0	['dailycovidupdate']]	0.266667	-0.100000	negative	dailycovidupdate 22nd february 2021 - cases...
20	\n1363876534579298316	2021-02-22	The Independent	Pfizer vaccine is cutting infections, hospital...	en	9.0	41.0]	0.900000	-0.600000	negative	pfizer vaccine is cutting infections hospital...
22	\n1363876368044486657	2021-02-22	ITV News	Boris Johnson says Public Health England has f...	en	14.0	32.0]	0.233333	-0.150000	negative	boris johnson says public health england has f...

```
In [83]: # Create word cloud of top occurring words  
allwords = ' '.join([twts for twts in df.cut5['tweet']])
```

```
In [84]: plt.figure(figsize=(16,10))
wordcloud = WordCloud(stopwords=stop_words, max_words=200, width=800, height=500, random_state=32, max_font_size=90).generate(allwords)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```




```
In [89]: # 5 - Code to generate the interactive vizualization
pyLDAvis.enable_notebook()

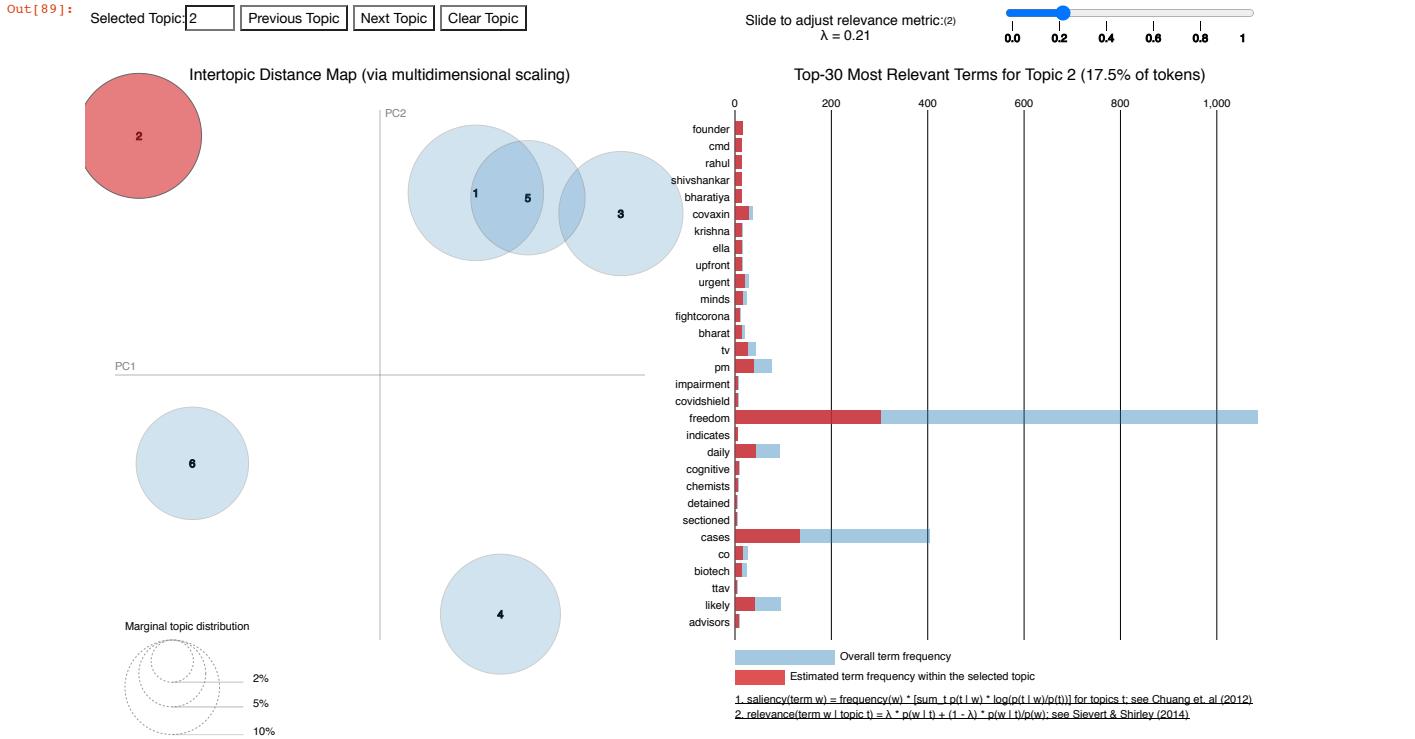
vis = pyLDAvis.gensim.prepare(lda_model, corpus, id2word)

/Users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/pyLDAvis/_prepare.py:257: FutureWarning: Sorting because non-concatenation axis is not aligned. A future version of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.

return pd.concat([default_term_info] + list(topic_dfs))
```



Observations on the negative set of tweets and LDA topics

With 6 topics, we still see a bit of overlap (clusters 1,3,5). The main 4 topic areas seem to be related to the following:

- * news focused - several people mentioned (jan adolan, rahul shivshankar, bill gates, dolly parton)
- * refuse due to bad side effects
- * conspiracy theories - china, evil, scared, defective, plagued
- * action - sign, petition, refuse
- * political - colonial, globalism, extermination

I found that in order to get the more distinct words, it helps to use the slider above the bar chart and slide it to the left. This cuts out some of the more popular words that are repeated in multiple columns.

Examine a set of positive sentiment tweets to see if topics are any different

```
In [117]: # Take a positive set from the combo-DF-of negative topics
df_cut7 = df_cut3.loc[df_cut3['Analysis'] == 'positive']
df_cut7.head()
```

Out[117]:

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed
1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certificate /passport but eventually will be used for so mu...	en	0.0	13.0	[]	0.712500	0.362500	positive	it will start off as just a vaccine certificate /passport but eventually will be used for so mu...
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vaccines & it's great to see hospital admissions ...	en	0.0	14.0	[]	0.583333	0.533333	positive	happy that the uk is doing well with their vaccines & it's great to see hospital admissions ...
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths coming from? The over-60 group is the group at risk. ...	en	18.0	83.0	[]	0.404242	0.020606	positive	4/ so where are these new cases and deaths coming from the over-60 group is the group at risk se...
5	\n1363887730527260674	2021-02-22	Alex Berenson	3/ And the Israeli government claims nearly perfect effectiveness for the vaccine - 99% reductio...	en	18.0	84.0	[]	0.538333	0.153333	positive	3/ and the israeli government claims nearly perfect effectiveness for the vaccine - 99% reductio...
8	\n1363886675844362248	2021-02-22	Alex Berenson	1/ I have to keep saying this: the Israeli's government's claims about how well the vaccine work...	en	172.0	350.0	[]	0.430303	0.075758	positive	1/ i have to keep saying this: the israeli's government's claims about how well the vaccine work...

```
In [126]: df_cut7.shape
```

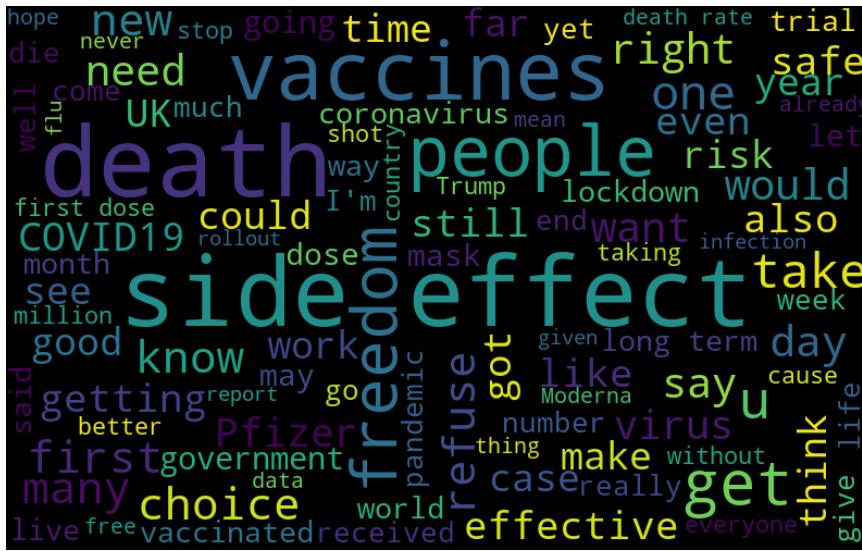
Out[126]: (19123, 12)

```
In [127]: list(df_cut7.tweet.sample(20))
```

Out[127]: ["From my research I've found no account of COVID vaccines tested on the type of people likely to be encountered in care homes. But yet it's being administered to them in large numbers & media is reporting several deaths after these people have been vaccinated. What's going on?",
 'People are literally going to die if they opt out of covid vaccination. If enough people refuse to get it when it's available the pandemic won't get much better. The desire for vaccines is polling behind the 70%-80% ideal. You shouldn't cause vaccine skepticism with bad takes.',
 "So let me get this straight. If someone dies within hours of receiving a vaccine it has nothing to do with the vaccine. But if they die within 28 days of a so called COVID 19 positive RT-PCR test (which isn't a diagnostic test) it definitely a COVID 19 death. Yeah right!",
 'About 70% of registered voters surveyed said they would take a Covid-19 vaccine, although many want to wait until it has been available for a while to see if there are major problems or side effects, a new WSJ/NBC News poll shows ',
 '_Garrett _Garrett are you sure that's right? It's reduction in incidence of infection- you are equating that to "severe hospitalisation and death". The difference is significant and some vaccines may be better or worse at the latter than that headline efficacy.',
 'Just checked the research on the vaccines. Pretty encouraging stuff! Almost no side effects or failures. Btw what's lycanthropy?',
 "GP Joe Biden's going to have to figure out a good way to explain to his voters that vaccines aren't going to be magically available and deaths aren't going magically stop as his voters seem to believe.",
 'A vaccine is still in the future. Its efficacy, side-effects, & duration of immunity may not be completely known when the vaccine is introduced. So masks, hand-washing, social distancing, etc will still be needed even after a vaccine is available-prob thru 2021 at the minimum.',
 "Dr. Melanie Swift, occupational medicine physician helping lead the COVID19 vaccination plan at Mayo Clinic, said it's important to educate Americans about a vaccine's side effects or it may deter people from getting a 2nd dose as needed. via ",
 'Special report yesterday from - care homes have had a very low incidence of virus before vaccine came. Now getting many "outbreaks". Dramatic increase in cases! Care workers being treated as "agent of state" UKC 8/2 also reported rise in deaths.',
 'The real virus \U0001f9a0 is Treason. There are more cases of Treason in this country than there are Covid deaths. *Real Talk* and the cure is Exposing it and the vaccine is removing it! That's happening right now and will continue in the coming weeks and months. 🇺🇸 is being cleansed',
 'Ok...so those who refuse to get vaccinated for Covid are not allowed into the general population that wishes to obliterate the virus by vaccination, as was done for polio, chickenpox, etc...refusnicks get to go off grid and isolate in their own community...I'm good with that.',
 'Australia right now: 😊 about to go 2 wks w/ zero local cases 😊 no Covid-19 deaths for months 😊 multiple vaccine options on horizon 😊 world looking our way & marveling Get need to question decisions, keep politicians accountable, but rest 🌎 would rather be here than there.',
 'Several of us are in an age group likely to be offered the vaccine soon. I realise there is very little evidence of side effects so far amongst healthy controls but a general guidance for ME and LongCovid patients would be gratefully received.',
 'The correct story should be "vaccine less effective at preventing mild illness and transmission. No data on whether vaccine prevents death or serious illness, but all other studies still point to vaccines being effective at doing this."',
 "People, make no mistake the time is coming when you're going to be FORCED to make a choice. Either fight for your freedom or remain passive to the Gov'ts will. I will not take a Vaccine that Bill Gates has a hand in and I'm damn sure not going to submit to being microchip.",
 'By pushing for mandatory COVID-19 vaccination, or imposing social restrictions on those who refuse, the COVID-19 vaccine is paving the way for nonconsensual medical experimentation on the general public" The Future of Vaccines Articles dot Mercola dot com Published Jan 12 2021',
 "This week's coronavirus report: As cases and hospitalizations decline, the next few weeks will be critical to ramp up vaccinations. The leading vaccines are extremely effective against severe disease, with zero Covid-19 deaths in trials. ",
 ' Thread. Email: "You recently signed the petition \'Prevent any restrictions on those who refuse a Covid-19 vaccination\': "On Tuesday 8 December there was an Urgent Question to the Department of Health and Social Care on the covid-19 vaccine roll out.. 1/n',
 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination Not taking this vaccine. We don't actually know what it will do to us in the long. Medication isn't always the answer. Why not try a healthy lifestyle to fight it off?"]

```
In [118]: # Create word cloud of top occurring words  
allwords = ' '.join([twts for twts in df_cut7['tweet']])
```

```
In [119]: plt.figure(figsize=(16,10))
wordcloud = WordCloud(stopwords=stop_words, max_words=100, width=800, height=500, random_state=32, max_font_size=90).generate(allwords)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```




```
In [124]: # 5 - Code to generate the interactive vizualization
pyLDAvis.enable_notebook()

vis = pyLDAvis.gensim.prepare(lda_model, corpus, id2word)
vis

/Users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/pyLDAvis/_prepare.py:257: FutureWarning: Sorting because non-concatenation axis is not aligned. A future version of pandas will change to not sort by default.

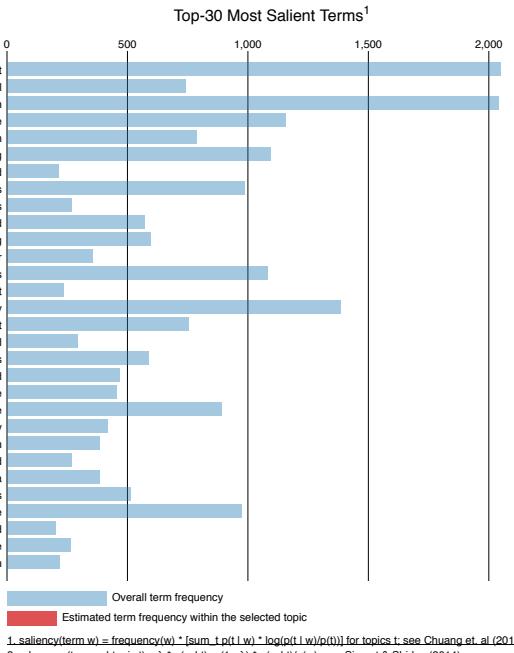
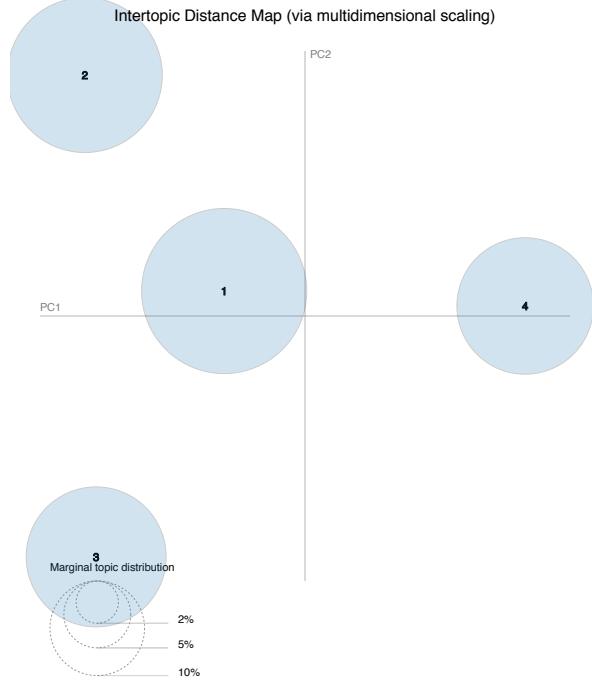
To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.
```

```
return pd.concat([default_term_info] + list(topic_dfs))
```

Out[124]: Selected Topic: 0 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 0$



Observations on the positive sentiment tweets

Here I found that with 4 topic clusters the clusters were dispersed. Overall, the topics were similar to the negative set. However, some positive words were seen... things around getting vaccinated or wanting to get vaccinated, or about when or which vaccination (1st or 2nd).

In []:

Complete Analysis with a more neutral to positive dataset

Next, I wanted to check to see if I examined tweets with more neutral keywords - "covid vaccination" would result in different findings. So this is the similar process applied to a different pull of tweets using Twint.

```
In [91]: df_start = pd.read_csv('data/cvacntn2.csv', sep='\t', lineterminator='\r')
print(df_start.shape)
df_start.head()

(28715, 36)

/Users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/IPython/core/interactiveshell.py:2714: DtypeWarning: Columns (0,21) have mixed types. Specify dtype option on import or set low_memory=False.
    interactivity=interactivity, compiler=compiler, result=result)
```

Out[91]:

		id	conversation_id	created_at	date	time	timezone	user_id	username	name	place	tweet	language	mentions	urls
0	136389364519633057	1.363894e+18		2021-02-22 08:49:19 PST	2021-02-22	08:49:19	-800.0	2.290538e+09	dominiccmc99	Dom Mcoughlin. !! MASKS SAVE LIVES!!	NaN	I can't thank enough the many scientists for...	en	[]	[]
1	1363892158399324170	1.363892e+18		2021-02-22 08:43:25 PST	2021-02-22	08:43:25	-800.0	1.989714e+07	indiatoday	IndiaToday	NaN	5 states reported a spike in COVID-19 cases ag...	en	{'screen_name': 'snehamordani', 'name': 'sneh...'}	[]
2	1363891979084570629	1.363892e+18		2021-02-22 08:42:42 PST	2021-02-22	08:42:42	-800.0	6.464306e+07	rt_com	RT	NaN	'Why would anyone get the jab after this?': AP...	en	[]	['https://on.rt.com/b2df'] ['https://pbs.twimg.co
3	1363890915467800576	1.363891e+18		2021-02-22 08:38:28 PST	2021-02-22	08:38:28	-800.0	3.814266e+07	le_parisien	Le Parisien	NaN	🔴 La vaccination contre le Covid sera possible...	fr	[]	['https://l.leparisien.fr/vZsw'] ['https://pbs.twimg.c
4	1363887481217830915	1.363887e+18		2021-02-22 08:24:49 PST	2021-02-22	08:24:49	-800.0	1.346792e+18	vaccinestatusuk	UK COVID-19 Vaccine Tracker	NaN	17,723,840 individuals in the UK received at...	en	[]	['https://pbs.twimg.com']

```
In [92]: # Remove unneeded columns
df_cuta = df_start.drop(columns=['conversation_id', 'created_at', 'time', 'timezone',
                                 'user_id', 'username', 'mentions',
                                 'urls', 'photos', 'replies_count', 'cashtags', 'link', 'retweet', 'quote_url', 'video',
                                 'thumbnail', 'source', 'user_rt_id', 'user_rt',
                                 'retweet_id', 'reply_to', 'retweet_date', 'translate', 'trans_src',
                                 'trans_dest'], axis=1)
df_cuta.head()
```

Out[92]:

	id	date	name	place	tweet	language	retweets_count	likes_count	hashtags	near	geo
0	136389364519633057	2021-02-22	Dom Mcoughlin. !! MASKS SAVE LIVES!!	NaN	I can't thank enough the many scientists for t...	en	3.0	15.0	[]	NaN	NaN
1	1363892158399324170	2021-02-22	IndiaToday	NaN	5 states reported a spike in COVID-19 cases ag...	en	3.0	24.0	[]	NaN	NaN
2	1363891979084570629	2021-02-22	RT	NaN	'Why would anyone get the jab after this?': AP...	en	11.0	18.0	['covid']	NaN	NaN
3	1363890915467800576	2021-02-22	Le Parisien	NaN	🔴 La vaccination contre le Covid sera possible...	fr	14.0	20.0	['covid--19']	NaN	NaN
4	1363887481217830915	2021-02-22	UK COVID-19 Vaccine Tracker	NaN	17,723,840 individuals in the UK received at...	en	3.0	16.0	['vaccine', 'vaccination']	NaN	NaN

```
In [93]: df_cuta = df_cuta.drop(columns=['place', 'near', 'geo'], axis=1)
df_cuta.shape
```

Out[93]: (28715, 8)

```
In [94]: df_cuta = df_cuta.loc[df_cuta['language'] == 'en']
df_cuta.shape
```

Out[94]: (25896, 8)

```
In [95]: df_cuta = df_cuta.loc[df_cuta['date'] > "2020-02-20"]
df_cuta.shape
```

Out[95]: (25894, 8)

```
In [96]: # Check for duplicate rows
duplicate = df_cuta[df_cuta.duplicated()]
duplicate.shape
```

Out[96]: (0, 8)

```
In [97]: print("Columns with null values")
display(df_cuta.isnull().sum())
Columns with null values
```

id	0
date	0
name	0
tweet	0
language	0
retweets_count	0
likes_count	0
hashtags	0

dtype: int64

```
In [98]: df_cutb = df_cuta
```

Begin NLP - for new dataset - keywords "covid vaccination"

```
In [128]: pd.set_option('display.max_colwidth', 80)
```

```
In [100]: # START with TextBlob processing and sentiment assignment.
# First strip out special characters in the tweets
# Strip out @, #, RT, and url from tweets
def cleantext(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = re.sub(r'#', '', text)
    text = re.sub(r'RT[\s]+', '', text)
    text = re.sub(r'https?:\/\/[\S]+', '', text)
    return text

df_cutb['tweet'] = df_cuta['tweet'].apply(cleantext)
```

```
In [101]: # Add subjectivity and polarity scores from TextBlob
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

df_cutb['subjectivity'] = df_cutb['tweet'].apply(getSubjectivity)
df_cutb['polarity'] = df_cutb['tweet'].apply(getPolarity)
df_cutb.head()
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity
0	1363893645196333057	2021-02-22	Dom McLoughlin. !!MASKS SAVE LIVES!!	I can't thank enough the many scientists for their fabulous development of what I believe are 99%...	en	3.0	15.0	[]	0.642857	0.371429
1	1363892158399324170	2021-02-22	IndiaToday	5 states reported a spike in COVID-19 cases again. On the other hand, India may have to rework t...	en	3.0	24.0	[]	0.591667	0.125000
2	1363891979084570629	2021-02-22	RT	'Why would anyone get the jab after this?': AP criticized for saying COVID vaccination WON'T bri...	en	11.0	18.0	['covid']	0.325000	0.075000
4	1363887481217830915	2021-02-22	UK COVID-19 Vaccine Tracker	17,723,840 individuals in the UK received at least one dose of a COVID vaccine as of 22 Febru...	en	3.0	16.0	['vaccine', 'vaccination']	0.350000	-0.100000
5	1363883815647801347	2021-02-22	ALT US INS (OG Immigration)	Here we go with Megyn whining about how difficult the virus is for her and that the vaccination...	en	3.0	25.0	[]	1.000000	-0.333333

```
In [102]: df_cutb['polarity'].describe()
```

```
Out[102]: count    25894.000000
mean     0.141009
std      0.240109
min     -1.000000
25%     0.000000
50%     0.100000
75%     0.266667
max      1.000000
Name: polarity, dtype: float64
```

```
In [103]: df_cutb['subjectivity'].describe()
```

```
Out[103]: count    25894.000000
mean     0.389380
std      0.270059
min     0.000000
25%     0.183333
50%     0.400000
75%     0.562500
max      1.000000
Name: subjectivity, dtype: float64
```

```
In [104]: # Group the tweets by positive, negative, or neutral based on Polarity score
```

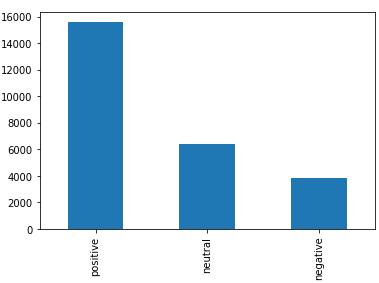
```
def getanalysis(score):
    if score <0:
        return "negative"
    elif score ==0:
        return "neutral"
    else:
        return "positive"

df_cutb['Analysis'] = df_cutb['polarity'].apply(getanalysis)
df_cutb.head()
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis
0	1363893645196333057	2021-02-22	Dom McLoughlin. !!MASKS SAVE LIVES!!	I can't thank enough the many scientists for their fabulous development of what I believe are 99%...	en	3.0	15.0	[]	0.642857	0.371429	positive
1	1363892158399324170	2021-02-22	IndiaToday	5 states reported a spike in COVID-19 cases again. On the other hand, India may have to rework t...	en	3.0	24.0	[]	0.591667	0.125000	positive
2	1363891979084570629	2021-02-22	RT	'Why would anyone get the jab after this?': AP criticized for saying COVID vaccination WON'T bri...	en	11.0	18.0	['covid']	0.325000	0.075000	positive
4	1363887481217830915	2021-02-22	UK COVID-19 Vaccine Tracker	17,723,840 individuals in the UK received at least one dose of a COVID vaccine as of 22 Febru...	en	3.0	16.0	['vaccine', 'vaccination']	0.350000	-0.100000	negative
5	1363883815647801347	2021-02-22	ALT US INS (OG Immigration)	Here we go with Megyn whining about how difficult the virus is for her and that the vaccination...	en	3.0	25.0	[]	1.000000	-0.333333	negative

```
In [105]: # Get percentage of tweets by polarity bucket (aka analysis)
print(df_cutb['Analysis'].value_counts(normalize=True)*100)
df_cutb['Analysis'].value_counts().plot(kind='bar')
plt.show()
```

```
positive    60.203136
neutral     24.812698
negative    14.984166
Name: Analysis, dtype: float64
```



```
In [106]: view_neg = df_cutb.loc[df_cutb['Analysis'] == 'negative']
```

```
In [107]: # Read over a sample of the tweets for context and deeper understanding.  
list(view.neg_tweet.sample(20))
```

[Out[107]: ['We support this risk based approach that means clinically vulnerable pregnant women & breastfeeding women will now have access to COVID-19 vaccination, until data exist to support routine vaccination in pregnancy', 'Dubious websites and social media posts are falsely accusing a member of President-elect Joe Biden's COVID-19 advisory team of wanting to mandate vaccination by tying it to food assistance. FactCHAT', 'COVID-19 is 10 times deadlier for people with Down syndrome, raising calls for early vaccination ', 'Disappointed that a New York hospital canceled my Covid vaccination scheduled for a few days from now. It doesn't have enough, it says. I'm sure the distribution will be sorted out before too long but, jeez, this is one lousy, ill-coordinated rollout.', 'please help us get the second vaccination on schedule I don't want to get covid just cause I'm a nurse, I don't want to bring it home to my disabled daughter _Wish ', 'An excellent GP tells me Covid "takes me at least three times as long as flu vaccination" Why? Frail patients are slower physically, hard of hearing, he has to mix saline + vaccine carefully and slowly, watch the temperature; then wait 15 minutes.', 'When China promised to help and provide other countries with the COVID vaccine, some people in the West called it "vaccine diplomacy." And now they are complaining the vaccine is not delivered soon enough? COVIDVaccination COVID19 ', '"It was like trying to get concert tickets - you just kept calling, calling, calling." People 65 and older are battling confusing rules and crashing websites as they try to schedule a Covid-19 vaccination. Here's what to do. ', '"JFC Dodger Stadium's COVID-19 vaccination site shut down after protesters gather at entrance ", "Poor support for self-isolation undermines the UK's Covid vaccination effort ", "It's ridiculous that I can get a COVID-19 vaccination before my father - who's in his 60s and works in a population dense area - because I'm a member of the media', 'Mentally I'm here after getting my 2nd Covid vaccination ', 'I will not take a covid vaccination. Should it be forced onto me, I will force back. You can take your covid vaccine and stick it up your ass!', 'TRUELY SCARY! 🤦 The National Cabinet has endorsed: "a national system to monitor (COVID-19) individual vaccination status" The Govt has NOT RULED OUT that Aussies may be denied goods & services if they don't take the vaccine! MY STATUS IS MY BUSINESS! ', 'Covid-19 spear phishing scams --Be cautious when opening emails and texts from unknown senders who promise information about getting a COVID-19 vaccination. Never open attachments or links from unknown senders. dontbeTHISguy See: ', 'can we talk about how stupid it was to build a soccer stadium without a roof in Minnesota??? it's been vacant for months due to covid and could have easily been used as housing, testing, vaccination center etc if there was a roof ?? whose idea was that', 'Lockdowns, masks, and other restrictions are justified by the claim that we are obligated to protect the elderly by slowing the spread of Covid. Once the elderly are protected by vaccination there is no need to slow the spread, and restrictions become tyrannical on their face.', '44% of Republicans believe Bill Gates is plotting to use a mass COVID-19 vaccination campaign as a pretext to implant microchips in billions of people and monitor their movements. Dear Independents, do u really want to empower these crazy assholes in November & keep this shit up?', 'NY Gov. Cuomo already came under fire for sending elderly covid patients into care homes prior to vaccination But remember "don't get cocky with covid". ', '.. is getting hammered in the German press over the EU COVIDVaccination debacle and her poor performance when she was German Defence Minister. This latest outburst from her is rather desperate.']]

```
In [109]: # Select just the negative sentiment tweets
df_cutc = df_cutb.loc[df_cutb['Analysis'] == 'negative']
df_cutc.head()
```

Out[109]:

	id	date	name	text	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis
4	1363887481217830915	2021-02-22	UK COVID-19 Vaccine Tracker	17,723,840 individuals in the UK received at least one dose of a COVID vaccine as of 22 February 2021.		en	3.0	16.0	['vaccine', 'vaccination']	0.350000	-0.100000	negative
5	1363883815647801347	2021-02-22	ALT US INS (OG Immigration)	Here we go with Megyn whining about how difficult the virus is for her and that the vaccination...		en	3.0	25.0	[]	1.000000	-0.333333	negative
17	1363874272486252546	2021-02-22	((Shane))	Just got my COVID vaccination. If I should happen turn into a rage zombie just remember grr arg...		en	0.0	18.0	[]	0.800000	-0.700000	negative
18	1363873899470082058	2021-02-22	*Walter Bloomberg	*UK Data Show 57% Fewer Observed Cases Than Expected Among Over 80s 28 Days After Single Dose *		en	7.0	31.0	[]	0.307143	-0.085714	negative
50	1363843487104344067	2021-02-22	Pirates	Our partnership with keeps growing stronger. On Saturday we teamed-up with AHN for our second m...		en	6.0	98.0	[]	0.200000	-0.100000	negative

In [125]: df_cutc.shape

Out[125]: (3880, 12)

Begin LDA process

```
In [111]: # 1 - TEXT Pre-processing
df_cutc['text_processed'] = \
df_cutc['text'].map(lambda x: re.sub('([,.!?])', '', x))
# Convert the titles to lowercase
df_cutc['text_processed'] = \
df_cutc['text_processed'].map(lambda x: x.lower())
# Print out the first rows of papers
df_cutc['text_processed'].head()
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
This is separate from the ipykernel package so we can avoid doing imports until
/Users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/ipykernel_launcher.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
Out[111]: 4    17723840 individuals in the uk received at least one dose of a covid vaccine as of 22 februar...
5    here we go with meghan whining about how difficult the virus is for her and that the vaccination...
17   just got my covid vaccination if i should happen turn into a rage zombie just remember grrr arggg...
18   *uk data show 57% fewer observed cases than expected among over 80s 28 days after single dose ...
50   our partnership with keeps growing stronger on saturday we teamed-up with ahn for our second ma...
Name: text_processed, dtype: object
```

```
In [112]: # 2 - Prepare the data for LDA analysis
stop_words = stopwords.words('english')
stop_words.extend(['Covid', 'covid', 'vaccine', 'vaccination', 'rt', 'amp', 'http', 'site', 'today', 'Vaccine', 'Vaccination', 'Covid vaccination'])

def sent_to_words(sentences):
    for sentence in sentences:
        # deacc=True removes punctuation
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))

def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc))
             if word not in stop_words] for doc in texts]

data = df_cute.text_processed.values.tolist()
data_words = list(sent_to_words(data))
# remove stop words
data_words = remove_stopwords(data_words)
print(data_words[:1][0][:30])

['individuals', 'uk', 'received', 'least', 'one', 'dose', 'february', 'adult', 'population', 'priority', 'groups']
```

```
In [113]: # 3 - Create dictionary and corpus
import gensim.corpora as corpora
# Create Dictionary
id2word = corpora.Dictionary(data_words)
# Create Corpus
texts = data_words
# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]
# View
print(corpus[:1][0][:30])
```

[(0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1), (9, 1), (10, 1)]

```
In [114]: # 4 - Train the model
from pprint import pprint
# number of topics
num_topics = 4
# Build LDA model
lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                         id2word=id2word,
                                         num_topics=num_topics)

# Print the Keyword in the 4 topics
pprint(lda_model.print_topics())
doc_lda = lda_model[corpus]
```

[
 (0, '0.007*"people" + 0.005*"us" + 0.004*"vaccinated" + 0.004*"would" + '
 '0.004*"health" + 0.004*"vaccines" + 0.003*"one" + 0.003*"going" + '
 '0.003*"get" + 0.003*"time"),
 (1, '0.010*"people" + 0.009*"get" + 0.006*"health" + 0.004*"one" + 0.004*"us" + '
 '0.004*"uk" + 0.004*"take" + 0.003*"care" + 0.003*"black" + 0.003*"week"),
 (2, '0.009*"people" + 0.006*"get" + 0.006*"need" + 0.005*"health" + '
 '0.005*"vaccinated" + 0.004*"vaccines" + 0.004*"us" + 0.003*"due" + '
 '0.003*"still" + 0.003*"staff"),
 (3, '0.006*"get" + 0.005*"one" + 0.005*"people" + 0.004*"mass" + 0.003*"country" +
 '0.003*"new" + 0.003*"year" + 0.003*"day" + 0.003*"us" + 0.003*"news")]

```
In [115]: # 5 - Code to generate the interactive vizualization
pyLDAvis.enable_notebook()

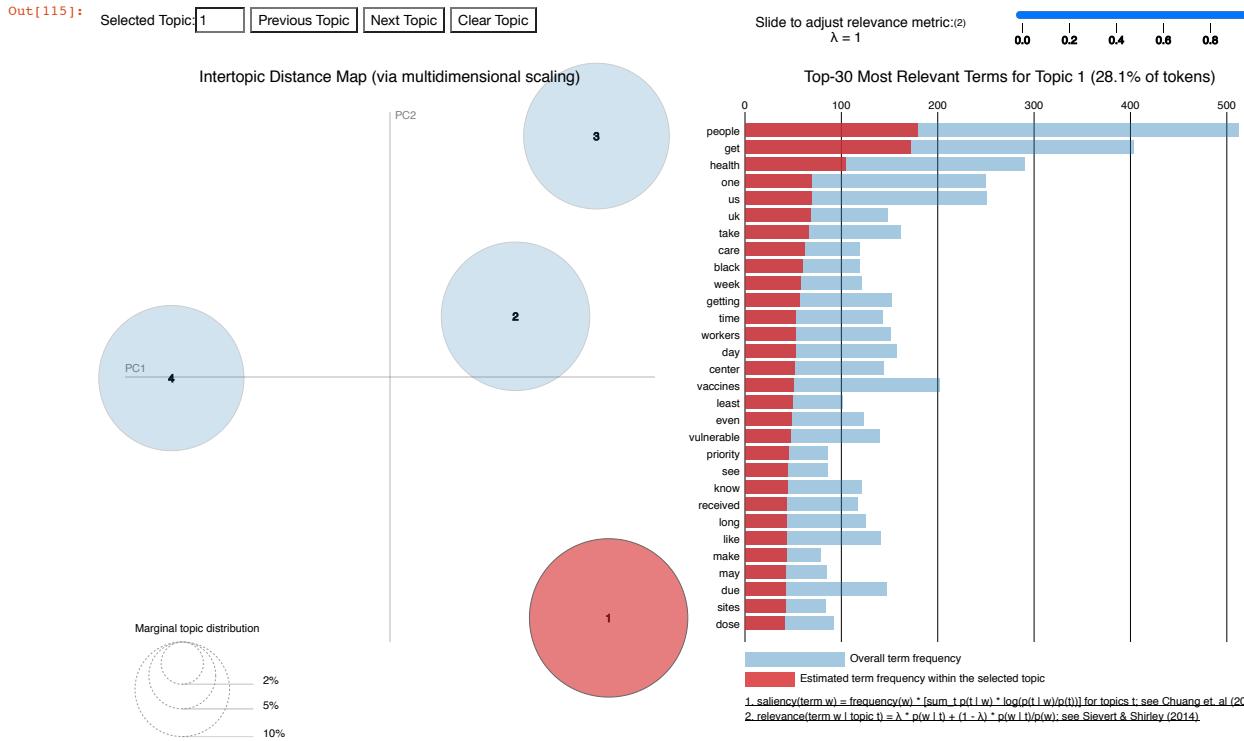
vis = pyLDAvis.gensim.prepare(lda_model, corpus, id2word)
vis

/users/markp/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/pyLDAvis/_prepare.py:257: FutureWarning: Sorting because non-concatenation axis is not aligned. A future version
of pandas will change to not sort by default.

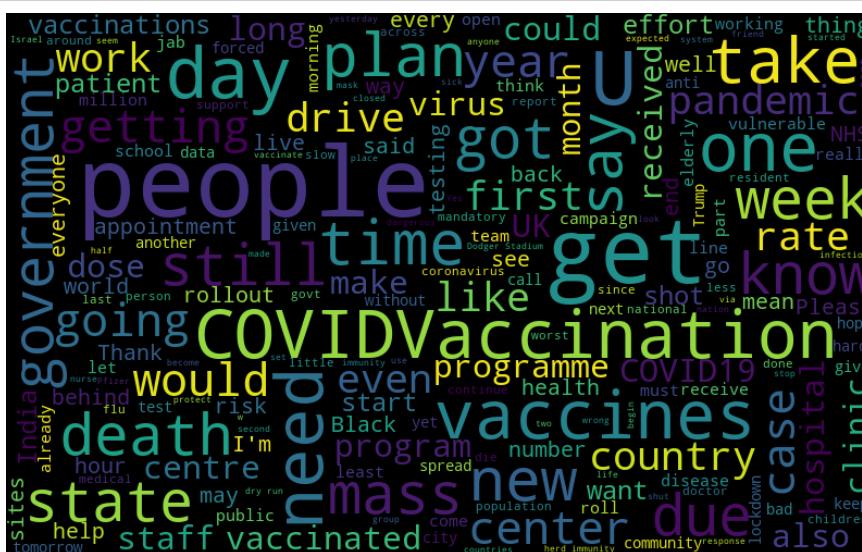
To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.

return pd.concat([default_term_info] + list(topic_dfs))
```



```
In [116]: plt.figure(figsize=(16,10))
wordcloud = WordCloud(stopwords=stop_words, max_words=200, width=800, height=500, random_state=32, max_font_size=90).generate(allwords)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



Observations on negative tweets from the more neutral subject of "covid vaccination"

Did not notice any substantial differences from what I saw earlier.

In []:

Try approach C - with different code for doing some EDA on top words, bi-grams and tri-grams

For this I will use the combo dataset of negative topics, and utilize the negative sentiment set. So this is about 25% of that dataset (~9K).

```
In [129]: df_cut5.shape
```

Out[129]: (9271, 12)

```
In [137]: pd.set_option('display.max_colwidth', 50)
```

In [138]: df_cut5.head()

Out[138]:

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% survival...	en	2.0	26.0	[]	0.684091	-0.052273	negative	remember folks this virus had a 99.7% survival...
6	\n1363887707076837376	2021-02-22	Kerry Hooper	Sick in a bag moment IDS wanting praise &...	en	7.0	14.0	['borisjohnson', 'road2recovery', 'roadmap', ...]	0.407937	-0.404762	negative	sick in a bag moment ids wanting praise &...
7	\n1363886935052324866	2021-02-22	Lawrence Gilder	DailyCovidUpdate 22nd February 2021 - Cases...	en	70.0	115.0	['dailycovidupdate']	0.266667	-0.100000	negative	dailycovidupdate 22nd february 2021 - cases...
20	\n1363876534579298316	2021-02-22	The Independent	Pfizer vaccine is cutting infections, hospital...	en	9.0	41.0	[]	0.900000	-0.600000	negative	pfizer vaccine is cutting infections hospital...
22	\n1363876368044486657	2021-02-22	ITV News	Boris Johnson says Public Health England has f...	en	14.0	32.0	[]	0.233333	-0.150000	negative	boris johnson says public health england has f...

```
In [130]: # Create a pipeline to clean the text
```

```
stopwords = nltk.corpus.stopwords.words('english')
ps = nltk.PorterStemmer()
wn = nltk.WordNetLemmatizer()
string.punctuation
```

```
Out[130]: '!#!#$%&\'()**+, -./:; <=>?@[\\"\\]^`{|}|`|'
```

```
In [139]: def clean_tweet(tweet):
    tweetn = "".join([word for word in tweetn if word not in string.punctuation])
    tokens = re.split("\W+", tweetn)
    tweetn = [wn.lemmatize(word) for word in tokens if word not in stopwords]
    return tweetn

df_cut5['tweet_clean'] = df_cut5['tweet'].apply(lambda x: clean_tweet(x.lower()))
```

If you're using .loc[row_indexer,col_indexer] - value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/cross_platform.html

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/tutorials/index.html>

```
a.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
import sys
```

```
In [140]: df_cuts.head()
```

Out[140]:

ID	Date	Name	Text	Language	tokens_count	words_count	Message	Confidence	Polarity	Analysis	Process	
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% survival...	en	2.0	26.0	[]	0.684091	-0.052273	negative	remember folks this virus had a 99.7% survival...
6	\n1363887707076837376	2021-02-22	kerry Hooper []	Sick in a bag moment IDS wanting praise &...	en	7.0	14.0	['borisjohnson', 'road2recovery', 'roadmap', ...]	0.407937	-0.404762	negative	sick in a bag moment ids wanting praise &...
7	\n1363886935052324866	2021-02-22	Lawrence Gilder	DailyCovidUpdate 22nd February 2021 - Cases...	en	70.0	115.0	['dailycovidupdate']	0.266667	-0.100000	negative	dailycovidupdate 22nd february 2021 - cases...
20	\n1363876534579298316	2021-02-22	The Independent	Pfizer vaccine is cutting infections, hospital...	en	9.0	41.0	[]	0.900000	-0.600000	negative	pfizer vaccine is cutting infections hospital...
22	\n1363876368044486657	2021-02-22	ITV News	Boris Johnson says Public Health England has f...	en	14.0	32.0	[]	0.233333	-0.150000	negative	boris johnson says public health england has f...

```
In [141]: #most frequent and least frequent words
```

```

negative_list = [] # list containing all words of all texts
for x in df_cut5['tweet_clean']: # loop over lists in df
    negative_list += x # append elements of lists to full list

neg_val_counts = pd.Series(negative_list).value_counts() # make
neg_val_counts

```

```
Out[141]: vaccine      10569
           death        5889
           effect       3098
           side         2973
           ...
hexavalent      1
hepatic         1
blinding        1
technofanart    1
pasta           1
Length: 18029, dtype: int64
```

```
In [170]: pd.set_option('display.max_rows', 100)
```

```
In [171]: neg_val_counts[:100]
```

```
Out[171]: vaccine      10569  
          5889  
          death       4794  
          effect      3098  
          side        2973  
          covid       2649  
          people      2056  
          freedom     1485  
          covid19     1471  
          amp         1400  
          get         1386  
          u           958  
          take        922  
          vaccination 860  
          one         811  
          day          684  
          case         672  
          virus        662  
          like         662  
          refuse       636  
          choice       611  
          would       609  
          know         603  
          year         588  
          need         587  
          long         577  
          rate         576  
          dont         541  
          want         529  
          even         517  
          say          508  
          getting      505  
          health       504  
          life         502  
          still        482  
          mask         473  
          make         463  
          time         444  
          serious      444  
          going        438  
          risk          438  
          bad           420  
          government   414  
          term          404  
          month         402  
          think         397  
          2             395  
          trial         389  
          uk            384  
          got           382  
          1             382  
          country       369  
          new           365  
          care          365  
          week          359  
          dose          358  
          go             355  
          world         355  
          see            351  
          could          350  
          million        349  
          work          347  
          today          345  
          also          343  
          cause          339  
          every          339  
          home          336  
          mean          334  
          may            330  
          first          327  
          way            326  
          lockdown       325  
          trump          324  
          sick           321  
          pfizer          321  
          back           320  
          flu            317  
          give           317  
          due            301  
          im             298  
          vaccinated    297  
          yet            296  
          number         296  
          pandemic       294  
          let             284  
          right          283  
          stop           280  
          last            278  
          shot            278  
          prevent         276  
          dangerous       273  
          taking          273  
          coronavirus    269  
          3               268  
          worse          267  
          le              261  
          worst          260  
          said           254  
          american        252  
          rollout         252  
          dtype: int64
```

```
In [153]: # MORE ISSUES with trying to adjust the display and the run off the page problem.  
# pd.set_option('display.max_columns', None)  
# pd.set_option('display.expand_frame_repr', False)  
# pd.set_option('max_colwidth', -1)
```

```
pd.set_option("display.max_rows", 50, "display.max_columns", 60)
```

```
In [155]: pd.set_option('display.max_colwidth', 60)
```

```
In [156]: df_cut5.head()
```

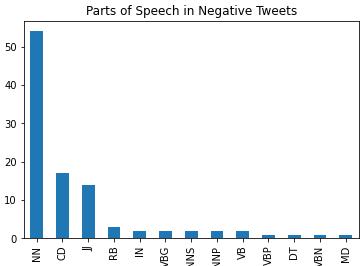
```
Out[156]:
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed	tweet_clean
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% survival rate bef...	en	2.0	26.0	[]	0.684091	-0.052273	negative	remember folks this virus had a 997% survival rate befo...	[, remember, folk, virus, 997, survival, rate, vaccine, ...]
6	\n1363887707076837376	2021-02-22	kerry Hooper □	Sick in a bag moment IDS wanting praise & applauds...	en	7.0	14.0	['borisjohnson', 'road2recovery', 'roadmap', 'toryfail']	0.407937	-0.404762	negative	sick in a bag moment ids wanting praise & applauds...	[sick, bag, moment, id, wanting, praise, amp, applauds, ...]
7	\n1363886935052324866	2021-02-22	Lawrence Gilder	DailyCovidUpdate 22nd February 2021 - Cases: 4,126,15...	en	70.0	115.0	['dailycovidupdate']	0.266667	-0.100000	negative	dailycovidupdate 22nd february 2021 - cases: 4126150 ...	[dailycovidupdate, 22nd, february, 2021, case, 4126150, ...]
20	\n1363876534579298316	2021-02-22	The Independent	Pfizer vaccine is cutting infections, hospitalisations a...	en	9.0	41.0	[]	0.900000	-0.600000	negative	pfizer vaccine is cutting infections hospitalisations an...	[pfizer, vaccine, cutting, infection, hospitalisation, d...]
22	\n136387636804486657	2021-02-22	ITV News	Boris Johnson says Public Health England has found that ...	en	14.0	32.0	[]	0.233333	-0.150000	negative	boris johnson says public health england has found that ...	[boris, johnson, say, public, health, england, found, on...]

```
In [162]: # Plot the parts of speech
```

```
blob = TextBlob(str(df_cut5['tweet_clean']))  
pos_df = pd.DataFrame(blob.tags, columns = ['word' , 'pos'])  
pos_df = pos_df.pos.value_counts()[:20]  
pos_df.plot(kind='bar', title="Parts of Speech in Negative Tweets")
```

```
Out[162]: <AxesSubplot:title={'center': 'Parts of Speech in Negative Tweets'}>
```



```
In [172]: (pd.Series(nltk.ngrams(negative_list, 2)).value_counts())[100]
```

```
Out[172]: {(side, effect)      : 2908,
( , )                  : 921,
(covid, vaccine)       : 856,
(covid19, vaccine)     : 571,
(long, term)           : 336,
(covid, death)          : 321,
(death, rate)           : 321,
(vaccine, )              : 317,
(get, vaccine)          : 316,
(effect, vaccine)       : 302,
(vaccine, side)         : 300,
(take, vaccine)         : 238,
( , vaccine)             : 218,
(term, side)             : 199,
(vaccine, rollout)       : 197,
(covid19, vaccination)   : 193,
(death, toll)             : 191,
(serious, side)          : 162,
(pfizer, vaccine)        : 160,
(death, )                 : 142,
(effect, )                 : 136,
(death, vaccine)          : 133,
(wear, mask)              : 130,
(refuse, covid19)         : 124,
(vaccine, death)          : 123,
(getting, vaccine)        : 123,
(coronavirus, vaccine)    : 121,
(care, home)              : 110,
(vaccine, trial)          : 107,
(restriction, refuse)      : 104,
(covid19, death)           : 100,
(covid, vaccination)       : 97,
(prevent, restriction)     : 96,
(amp, death)               : 91,
(taking, vaccine)          : 91,
(petition, prevent)        : 90,
(effect, covid)            : 86,
(flu, vaccine)              : 86,
(refuse, vaccination)       : 86,
(2nd, dose)                : 84,
(nursing, home)             : 84,
(herd, immunity)            : 82,
(vaccine, roll)             : 81,
(death, covid)              : 80,
(get, covid)                : 77,
(vaccine, vaccine)          : 76,
(vaccine, cause)            : 75,
(vaccine, amp)               : 74,
(case, death)                : 74,
(cause, death)               : 73,
(public, health)             : 72,
(got, vaccine)               : 72,
(freedom, )                  : 71,
(vaccine, people)            : 71,
(immune, system)              : 70,
(bill, gate)                  : 69,
(freedom, choice)             : 69,
(possible, side)              : 68,
(vaccination, )                : 68,
(medical, freedom)            : 67,
(death, amp)                  : 67,
(need, vaccine)                : 65,
(mrna, vaccine)                : 65,
(worst, death)                  : 65,
(second, dose)                : 65,
( , people)                  : 65,
(vaccine, passport)            : 64,
(number, death)                : 64,
(illness, death)                : 64,
(vaccine, distribution)        : 63,
(bad, side)                  : 63,
(vaccine, work)                : 62,
(vaccine, may)                  : 62,
(death, due)                  : 61,
( , covid)                    : 60,
(vaccine, today)                : 60,
(freedom, speech)                : 60,
(1st, dose)                  : 59,
(people, refuse)                : 59,
(unnecessary, death)              : 58,
(getting, covid)                : 58,
(sore, arm)                  : 58,
(choice, )                    : 57,
(serious, illness)              : 57,
(moderna, vaccine)                : 56,
(dont, want)                  : 56,
(save, life)                  : 54,
(refuse, take)                  : 54,
(average, death)                : 54,
(wearing, mask)                : 54,
(want, vaccine)                : 53,
(survival, rate)                : 53,
(vaccine, freedom)                : 53,
(rushed, vaccine)                : 53,
(hospitalization, death)        : 52,
(dont, know)                  : 52,
(via, )                      : 51,
(longterm, side)                : 51,
(vaccine, get)                  : 50,
(get, vaccinated)                : 50
dtype: int64
```

```
In [173]: (pd.Series(nltk.ngrams(negative_list, 3)).value_counts())[:100]
```

```
Out[173]: 
(vaccine, side, effect)      297
(side, effect, vaccine)      275
(term, side, effect)         195
(long, term, side)           187
(serious, side, effect)      161
(side, effect, )              129
(refuse, covid19, vaccination) 116
(restriction, refuse, covid19) 100
(prevent, restriction, refuse) 96
(petition, prevent, restriction) 90
(side, effect, covid)        69
(possible, side, effect)      68
(vaccine, , )                 66
(bad, side, effect)           62
(covid, vaccine, side)         53
(longterm, side, effect)       51
(potential, side, effect)      50
(effect, covid, vaccine)       49
(, , vaccine)                  48
(death, rate, world)           46
(know, side, effect)           45
(vaccine, 1st, dose)            45
(vaccine, 2nd, dose)            44
(side, effect, covid19)         43
(negative, side, effect)        43
(long, term, effect)           41
(covid19, vaccination, )        41
(severe, side, effect)          40
(effect, covid19, vaccine)       39
(got, covid, vaccine)           39
(know, long, term)               37
(side, effect, like)             37
(serious, illness, death)        35
(unknown, side, effect)          35
(worst, death, rate)              33
(get, covid, vaccine)             33
(adverse, side, effect)           32
(covid19, vaccine, )              31
(side, effect, may)                30
(side, effect, could)             30
(getting, covid, vaccine)         30
(dangerous, side, effect)          29
(side, effect, people)             29
(covid19, vaccine, side)           28
(side, effect, including)          28
(covid, vaccine, )                 28
(, covid, vaccine)                  27
(sign, reshare, widely)            27
(common, side, effect)             27
(minor, side, effect)                26
(death, , )                         25
(terrible, side, effect)             25
(highest, death, rate)                25
(, side, effect)                   24
(take, covid, vaccine)                 24
(horrible, side, effect)                24
(pfizer, covid, vaccine)                 23
(covid, death, rate)                  23
(see, side, effect)                   23
(dose, covid, vaccine)                  23
(johnson, amp, johnson)                 23
(february, 2021, case)                  22
(hank, aaron, death)                   22
(covid, vaccine, today)                  22
(nvic, vaccinefreedom, informedconsent) 22
(effect, , )                           22
(refuse, covid, vaccination)             21
(side, effect, far)                   21
(january, 2021, case)                  21
(vaccinefreedom, informedconsent, )     21
(vaccine, serious, side)                  21
(health, care, worker)                   21
(worst, death, toll)                     21
(, covid19, vaccine)                   20
(covid19, vaccination, sign)                20
(zero, side, effect)                   20
(one, side, effect)                   20
(side, effect, long)                   20
(care, home, death)                     20
(average, age, death)                   20
(swine, flu, vaccine)                   19
(side, effect, pfizer)                   19
(vaccination, , )                      19
(side, effect, get)                     19
(get, vaccine, first)                   19
(phase, 3, trial)                      18
(vaccine, prevent, death)                  18
(within, 28, day)                      18
(case, amp, death)                      18
(amp, side, effect)                      18
(unknown, long, term)                   18
(vaccine, trial, participant)             17
(covid19, death, toll)                   17
(worse, side, effect)                   17
(effect, side, effect)                   17
(unpleasant, side, effect)                  17
(pfizer, covid19, vaccine)                  17
(side, effect, bad)                      17
(side, effect, second)                   17
(side, effect, include)                   16
dtype: int64
```

Observations on Top Words, Bi-grams, Tri-grams - for the negative sentiment tweets

This method worked well for quickly generating lists of top words, bi-grams, and tri-grams. The bi-grams are a big help in improving understanding of single words; tri-grams did not add much more understanding - at least with this data set.

Some of the top words and combinations included:

- * serious, side, effect
- * long, term, side
- * petition, prevent, restrictions
- * freedom, choice
- * freedom, medical
- * rushed, vaccine
- * bill, gate

It is still important to read over a sample of the tweets for deeper understanding and context. This was an important part of this analysis as well. Some of the key themes that emerged from the negative sentiment tweets were:

- * vaccine itself: ineffective, rushed, not thoroughly tested, side effects / long term / deaths
- * process: forced, lack of choice = lack of freedom, status tracked, denial of services if not vaccinated
- * actions: sign petition to prevent restrictions, protest close vaccination site
- * blame or source: microchip to track you; medical experiment; Bill Gates, WHO, Democrats, government, big pharma, the church

In [1]:

Like this approach... lets go through the neutral and positives too

In [174]: df_cut3.head()

Out[174]:

	id	date	name	text	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed
0	\n1363895782475243527	2021-02-22	Ms M	It's a death sentence for me. If mine go back on 8th Ma...	en	2.0	12.0	[]	0.250000	0.000000	neutral	it's a death sentence for me if mine go back on 8th mar...
1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certificate /passpo...	en	0.0	13.0	[]	0.712500	0.362500	positive	it will start off as just a vaccine certificate /passpo...
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vaccines &...	en	0.0	14.0	[]	0.583333	0.533333	positive	happy that the uk is doing well with their vaccines &...
3	\n1363889576583065600	2021-02-22	Adam Knight	Remember folks this virus had a 99.7% survival rate bef...	en	2.0	26.0	[]	0.684091	-0.052273	negative	remember folks this virus had a 99.7% survival rate befo...
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths coming from? ...	en	18.0	83.0	[]	0.404242	0.020606	positive	4/ so where are these new cases and deaths coming from t...

```
In [176]: df3a_neutral = df_cut3.loc[df_cut3['Analysis'] == 'neutral']
print(df3a_neutral.shape)
df3a_neutral.head()

(9589, 12)
```

Out[176]:

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed
0	\n1363895782475243527	2021-02-22	Ms M	It's a death sentence for me. If mine go back on 8th Mar...	en	2.0	12.0	⌚	0.25	0.0	neutral	it's a death sentence for me if mine go back on 8th mar...
29	\n1363873071069806592	2021-02-22	🔥 This Is Fine🔥 Too 🔥	People screeching "no vaccines for anti-maskers" is peak...	en	5.0	17.0	⌚	0.00	0.0	neutral	people screeching "no vaccines for anti-maskers" is peak...
33	\n1363867929377390596	2021-02-22	The Hill	Israel: Pfizer vaccine prevents 98.9 percent of COVID-19...	en	68.0	286.0	⌚	0.00	0.0	neutral	israel: pfizer vaccine prevents 989 percent of covid-19 ...
43	\n1363855246506811394	2021-02-22	Graham Hutchinson	What is irresponsible is introducing an untested vaccine...	en	41.0	127.0	⌚	0.00	0.0	neutral	what is irresponsible is introducing an untested vaccine...
48	\n1363848291570581508	2021-02-22	The Hill	Israel: Pfizer vaccine prevents 98.9 percent of COVID-19...	en	46.0	142.0	⌚	0.00	0.0	neutral	israel: pfizer vaccine prevents 989 percent of covid-19 ...

```
In [177]: df3a_positive = df_cut3.loc[df_cut3['Analysis'] == 'positive']
print(df3a_positive.shape)
df3a_positive.head()

(19123, 12)
```

Out[177]:

1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certificate /passpo...	en	0.0	13.0	□	0.712500	0.362500	positive	It will start off as just a vaccine certificate /passpo...
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vaccines &am...	en	0.0	14.0	□	0.583333	0.533333	positive	happy that the uk is doing well with their vaccines &am...
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths coming from? ...	en	18.0	83.0	□	0.404242	0.020606	positive	4/ so where are these new cases and deaths coming from t...
5	\n1363887730527260674	2021-02-22	Alex Berenson	3/ And the Israeli government claims nearly perfect effe...	en	18.0	84.0	□	0.538333	0.153333	positive	3/ and the israeli government claims nearly perfect effe...
8	\n1363886675844362248	2021-02-22	Alex Berenson	1/ I have to keep saying this: the Israeli's government'...	en	172.0	350.0	□	0.430303	0.075758	positive	1/ i have to keep saying this: the israeli's government'...

```
In [178]: # Create a pipeline to clean the text
```

```
stopwords = nltk.corpus.stopwords.words('english')
ps = nltk.PorterStemmer()
wn = nltk.WordNetLemmatizer()
string.punctuation
```

```
Out[178]: '1."#\$\$&\\"()**+ - /::<=>20[\\"1^`~{]`-
```

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
In [180]: df3a_neutral.head()
```

```
Out[180]:
```

	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed	tweet_clean
0	\n1363895782475243527	2021-02-22	Ms M	It's a death sentence for me. If mine go back on 8th Ma...	en	2.0	12.0	[]	0.25	0.0	neutral	it's a death sentence for me if mine go back on 8th mar...	[, death, sentence, mine, go, back, 8th, march, exposed...]
29	\n1363873071069806592	2021-02-22	👉 This Is Fine 👉 Too 👕	People screeching "no vaccines for anti-maskers" is peak...	en	5.0	17.0	[]	0.00	0.0	neutral	people screeching "no vaccines for anti-maskers" is peak...	[people, screeching, vaccine, antimaskers, peak, brainde...]
33	\n1363867929377390596	2021-02-22	The Hill	Israel: Pfizer vaccine prevents 98.9 percent of COVID-19...	en	68.0	286.0	[]	0.00	0.0	neutral	israel: pfizer vaccine prevents 989 percent of covid-19 ...	[israel, pfizer, vaccine, prevents, 989, percent, covid1...]
43	\n1363855246506811394	2021-02-22	Graham Hutchinson	What is irresponsible is introducing an untested vaccine...	en	41.0	127.0	[]	0.00	0.0	neutral	what is irresponsible is introducing an untested vaccine...	[irresponsible, introducing, untested, vaccine, stopping...]
48	\n1363848291570581508	2021-02-22	The Hill	Israel: Pfizer vaccine prevents 98.9 percent of COVID-19...	en	46.0	142.0	[]	0.00	0.0	neutral	israel: pfizer vaccine prevents 989 percent of covid-19 ...	[israel, pfizer, vaccine, prevents, 989, percent, covid1...]

```
In [193]: # Take a look at a sample of neutral tweets
```

```
list(df3a_neutral['text'].sample(25))
```

```
Out[193]: ['Vaccine deaths will be classified as covid-19 deaths! []', 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ', 'Statistical series that seems to show the benefit of vaccine protection is already starting to kick-in for the over-80s, with that age group making up a reducing share of Covid deaths as the month has gone on.', '£8.6bn financial support, 3 vaccines, Scotland highest deaths per capita in the world yeah lets blame WM for Nicolas failings. ', 'Low side effects?! Shouldn't every vaccine have low side effects? Actually, every vaccine should have NO side effects!', 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination obingo', 'Covid related deaths jump by 46% which coincides with the massive vaccine rollout Anyone else the slightest bit suspicious \U0001f928', '70% of COVID-19 deaths have been in Alaskans over 65 who make up only 10% of our population. So reach out to Alaskans age 65+ to help them schedule a vaccine appt. by calling 907-646-3322 or go to We can do this-together. akleg', 'Healthcare Worker Dies 4 Days After Second Pfizer Vaccine, Death Under Investigation ', 'Bet the freedom to bear arms and say no to a vaccine.', 'randy_mcgrath The problem of getting to 70% is a time problem Anti vaxxers will make up the difference by getting covid. If 60% vaccinate, 40% dont, after a third of those that refuse a vaccination get covid, we are at 70. Protecting children from antivaxxers becomes the issue', 'Britain's vaccine drive is ALREADY stopping Covid deaths, study shows ', 'Just got my second shot of The Vaccine. Will report on side effects. So grateful to have been given the chance to be front loaded for the juice. COVID19', 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ', 'Another message technique was the "personal freedom" message, which basically said if you want to get your freedoms back, and you want a return to how things were, vaccines are the only way back', 'I will never understand this argument. If you have the vaccine how am i a danger to you?? And how can you say things like "get your freedom back". I wasnt told it had been taken. Foul foul people', 'Sooner or Later, Americans Will Have to Choose Between Freedom and a Micro-Chipped Vaccine ', 'It'll be coerced vaccine. No vaccine no flights - look at quantas. Get your freedom pass stamped for flight...concerts...or alternatively get a stick up your nose twice a week to allow you to do these things with your pass. Shove it Boris.', 'Translation? No end in sight for Lockdowns No vaccine = no freedom Why would anyone vote for this feckless pawn of the globalists? ', 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ', 'Petition: Prevent any restrictions on those who refuse a Covid-19 vaccination ', 'The NY Democrats stole your freedom to refuse vaccines June 13, 2019. They will add Mandatory Covid vaccine to the list in '21. Only a Republican Senate majority will stop them. Vote for your kids. VoteRed VoteWithaVengeance MakeThemPay', 'There's literally no reason to have NPIs if you have a vaccine and don't have excess deaths.', '2 events today demonstrate the difference between KSA's vision for for the ME vs. Iran's vision of death & destruction. Khamenei prohibits use of vaccine s for his people, risking their health. King Salman takes vaccine after implementing a tech- enabled nat'l vaccination campaign. ", 'Getting my vaccine tmrw, one step closer to freedom from the shackles of COVID']
```

```
In [181]: #most frequent and least frequent words
```

```
neutral_list = [] # list containing all words of all texts
for x in df3a_neutral['text_clean']: # loop over lists in df
    neutral_list += x # append elements of lists to full list

neutral_val_counts = pd.Series(neutral_list).value_counts() # make temporary Series to count
neutral_val_counts
```

```
Out[181]:
```

```
7989
vaccine      6447
covid19      4853
vaccination   4400
refuse       4240
...
smuggled      1
shongwe      1
covid21      1
economically  1
vaccinatedrine 1
Length: 12594, dtype: int64
```

```
In [182]: neutral_val_counts[:100]
```

```
Out[182]:
```

vaccine	6447
covid19	4853
vaccination	4400
refuse	4240
restriction	3603
prevent	3545
petition	3455
death	2624
covid	1740
effect	1638
side	1607
freedom	1200
people	922
get	727
amp	514
take	503
u	489
choice	443
say	408
health	375
day	342
sign	338
pfizer	322
like	313
one	302
would	291
government	278
please	264
want	261
need	257
today	256
mask	256
dont	254
getting	249
coronavirus	246
case	237
know	229
medical	217
week	216
going	206
still	204
virus	201
doctor	196
got	195
could	195
trial	193
uk	189
reaction	189
vaccinated	186
time	176
year	176
make	176
also	170
news	170
rate	169
give	169
1	167
world	166
taking	164
public	163
life	163
work	162
cause	160
see	159
shot	158
wont	157
back	157
2	155
died	155
think	152
even	150
care	148
home	148
via	148
million	148
risk	147
let	143
mandatory	142
every	142
last	141
may	140
stop	139
moderna	138
keep	137
trump	137
never	135
sideeffects	133
said	132
dose	127
lockdown	127
number	126
worker	126
without	126
severe	124
way	123
state	122
data	122
pandemic	122
go	122

dtype: int64

```
In [183]: (pd.Series(nltk.ngrams(neutral_list), 2)).value_counts()[:100]
```

```
Out[183]: {(covid19, vaccination)      : 3687
(refuse, covid19)        : 3562
(restriction, refuse)   : 3497
(prevent, restriction)  : 3461
(petition, prevent)     : 3169
(vaccination, )          : 3118
(, petition)             : 2766
(side, effect)           : 1582
(, )                     : 845
(covid, vaccine)         : 794
(covid19, vaccine)       : 620
(vaccine, )               : 360
(vaccine, side)          : 336
(vaccine, death)         : 210
(effect, )                : 197
(covid, death)            : 181
(death, )                 : 178
(effect, vaccine)        : 178
(, vaccine)               : 177
(take, vaccine)           : 166
(covid, vaccination)      : 158
(refuse, vaccination)    : 143
(get, vaccine)             : 139
(vaccination, petition)  : 137
(please, sign)             : 133
(pfizer, vaccine)          : 133
(refuse, covid)            : 128
(people, refuse)           : 117
(, prevent)                : 115
(sign, petition)           : 110
(death, vaccine)           : 108
(vaccinefreedom, informedconsent) : 106
(vaccine, rollout)         : 98
(death, toll)               : 97
(coronavirus, vaccine)     : 96
(effect, covid)              : 96
(vaccine, reaction)        : 94
(, people)                  : 94
(, please)                  : 90
(vaccine, freedom)          : 88
(petition, )                  : 87
(via, )                     : 84
(covid19, death)             : 82
(, covid)                   : 81
(covid, 19)                  : 79
(death, rate)                 : 78
(getting, vaccine)           : 78
(informedconsent, )          : 78
(nvic, vaccinefreedom)       : 75
(freedom, )                  : 74
(get, covid)                  : 73
(, refuse)                   : 70
(sign, )                     : 67
(medical, freedom)           : 66
(wear, mask)                  : 65
(today, nvic)                  : 65
(vaccine, trial)              : 64
(refuse, get)                  : 61
(cause, death)                 : 61
(vaccine, vaccine)             : 61
(last, week)                  : 61
(get, covid19)                  : 61
(freedom, choice)              : 60
(taking, vaccine)              : 60
(goal, sign)                  : 59
(vaccination, closing)         : 59
(closing, goal)                 : 59
(death, covid)                  : 57
(, side)                      : 56
(nursing, home)                 : 54
(, death)                      : 52
(subscribe, tvr)                 : 51
(tvr, today)                   : 51
(doctor, unloads)                : 51
(unloads, covid)                 : 51
(public, health)                 : 51
(vaccination, acclaimed)        : 51
(refuse, vaccine)                 : 50
(effect, covid19)                : 50
(vaccination, covid19)           : 49
(injury, death)                  : 49
(, spain)                      : 48
(news, )                        : 48
(walk, park)                    : 47
(second, dose)                  : 46
(vaccine, news)                  : 46
(acclaimed, french)                : 46
(vaccine, shot)                  : 46
(choice, )                      : 46
(28, day)                      : 46
(french, doctor)                  : 46
(body, choice)                  : 46
(vaccination, could)              : 46
(vaccine, wont)                  : 45
(health, minister)                : 45
(, got)                          : 44
(, sign)                        : 43
(got, vaccine)                  : 43
(bill, gate)                     : 43
(vaccine, cause)                  : 43
dtype: int64
```

```
In [184]: (pd.Series(nltk.ngrams(neutral_list, 3)).value_counts()[:100]
```

```
Out[184]: (refuse, covid19, vaccination)      3551
(restriction, refuse, covid19)                3467
(prevent, restriction, refuse)                3455
(petition, prevent, restriction)              3169
(covid19, vaccination, )                     3055
(, petition, prevent)                       2735
(vaccination, , petition)                   2485
(vaccine, side, effect)                    332
(side, effect, )                           190
(side, effect, vaccine)                   171
(vaccination, , )                          156
(covid19, vaccination, petition)            115
(, prevent, restriction)                  114
(side, effect, covid)                     95
(effect, covid, vaccine)                  86
(vaccination, petition, )                 82
(vaccine, , )                            77
(vaccinefreedom, informedconsent, )        77
(vaccination, , prevent)                  75
(nvic, vaccinefreedom, informedconsent)    75
(covid, vaccine, side)                    75
(refuse, covid, vaccination)               74
(, please, sign)                         66
(today, nvic, vaccinefreedom)              65
(petition, , petition)                   59
(closing, goal, sign)                    59
(covid19, vaccination, closing)            59
(vaccination, closing, goal)              59
(covid19, vaccine, )                     59
(goal, sign, )                           58
(vaccination, , please)                  56
(, side, effect)                         55
(, refuse, vaccination)                  53
(refuse, vaccination, acclaimed)          51
(subscribe, tvr, today)                  51
(unloads, covid, vaccine)                51
(doctor, unloads, covid)                 51
(covid, vaccine, )                      49
(side, effect, covid19)                  49
(, covid, vaccine)                      47
(vaccination, acclaimed, french)          46
(acclaimed, french, doctor)              46
(sign, , petition)                      46
(french, doctor, unloads)                46
(sign, petition, prevent)                45
(effect, covid19, vaccine)                45
(covid, vaccine, news)                  45
(, , refuse)                           44
(please, sign, petition)                44
(tvr, today, nvic)                      43
(possible, side, effect)                41
(covid, 19, vaccination)                41
(walk, park, )                          41
(refuse, covid, 19)                     40
(, people, refuse)                      40
(covid19, vaccination, could)            40
(refuse, voluntary, covid19)             39
(voluntary, covid19, vaccination)         39
(vaccination, could, face)              39
(news, lifesite, )                      39
(within, 28, day)                      39
(people, refuse, voluntary)              38
(vaccine, news, lifesite)                38
(could, face, restriction)              38
(covid19, vaccination, sign)             36
(face, restriction, ontario)             36
(potential, side, effect)                36
(ontario, govt, warns)                  36
(chief, medical, officer)                36
(restriction, ontario, govt)              36
(, , vaccine)                           35
(covid19, vaccine, side)                35
(conversation, vaccination, health)     34
(freedom, move, around)                  34
(get, covid, vaccine)                   34
(enlightened, conversation, vaccination) 34
(wont, walk, park)                      34
(u, enlightened, conversation)            34
(vaccination, , sign)                   34
(vaccination, health, autonomy)          34
(world, news, )                        34
(join, u, enlightened)                  34
(effect, vaccine, )                     33
(people, refuse, covid19)                33
(ccdc, warn, people)                   33
(people, side, effect)                  32
(reaction, join, u)                     31
(registry, refuse, covid)                31
(vaccine, reaction, join)                31
(death, , )                            31
(covid19, vaccination, please)           31
(data, european, partner)                30
(list, people, refuse)                  30
(spanish, keep, list)                   30
(share, data, european)                 30
(warn, people, side)                   30
(covid19, vaccination, amp)              29
(get, covid19, vaccination)              29
(covid, vaccination, via)                29
(vaccination, amp, share)                29
dtype: int64
```

```
In [185]: # Create a pipeline to clean the text
stopwords = nltk.corpus.stopwords.words('english')
ps = nltk.PorterStemmer()
wn = nltk.WordNetLemmatizer()
string.punctuation
```

```
Out[185]: '!#$%&`(*+,-./:;<=>?@[\`\\]^_`{|}-'
```

In [187]:	df3a_positive.head()																																																																																				
Out[187]:	<table border="1"> <thead> <tr> <th></th><th>id</th><th>date</th><th>name</th><th>tweet</th><th>language</th><th>retweets_count</th><th>likes_count</th><th>hashtags</th><th>subjectivity</th><th>polarity</th><th>Analysis</th><th>text_processed</th><th>tweet_clean</th></tr> </thead> <tbody> <tr> <td>1</td><td>\n1363895502627078147</td><td>2021-02-22</td><td>LizW</td><td>It will start off as just a vaccine certificate /passpo...</td><td>en</td><td>0.0</td><td>13.0</td><td>[]</td><td>0.712500</td><td>0.362500</td><td>positive</td><td>it will start off as just a vaccine certificate /passpo...</td><td>[I, start, vaccine certificate, passport, eventually, us...</td></tr> <tr> <td>2</td><td>\n1363895430032031749</td><td>2021-02-22</td><td>Marion</td><td>Happy that the UK is doing well with their vaccines &amp...</td><td>en</td><td>0.0</td><td>14.0</td><td>[]</td><td>0.583333</td><td>0.533333</td><td>positive</td><td>happy that the uk is doing well with their vaccines &amp...</td><td>[happy, uk, well, vaccine, amp, great, see, hospital, ad...</td></tr> <tr> <td>4</td><td>\n1363888280597647370</td><td>2021-02-22</td><td>Alex Berenson</td><td>4/ So where are these new cases and deaths coming from? ...</td><td>en</td><td>18.0</td><td>83.0</td><td>[]</td><td>0.404242</td><td>0.020606</td><td>positive</td><td>4/ so where are these new cases and deaths coming from t...</td><td>[4, new, case, death, coming, over60, group, group, risk...</td></tr> <tr> <td>5</td><td>\n1363887730527260674</td><td>2021-02-22</td><td>Alex Berenson</td><td>3/ And the Israeli government claims nearly perfect effe...</td><td>en</td><td>18.0</td><td>84.0</td><td>[]</td><td>0.538333</td><td>0.153333</td><td>positive</td><td>3/ and the israeli government claims nearly perfect effe...</td><td>[3, israeli, government, claim, nearly, perfect, effect...</td></tr> <tr> <td>8</td><td>\n1363886675844362248</td><td>2021-02-22</td><td>Alex Berenson</td><td>1/ I have to keep saying this: the Israeli's government'...</td><td>en</td><td>172.0</td><td>350.0</td><td>[]</td><td>0.430303</td><td>0.075758</td><td>positive</td><td>1/ i have to keep saying this: the israeli's government'...</td><td>[1, keep, saying, israeli, government, claim, well, vac...</td></tr> </tbody> </table>		id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed	tweet_clean	1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certificate /passpo...	en	0.0	13.0	[]	0.712500	0.362500	positive	it will start off as just a vaccine certificate /passpo...	[I, start, vaccine certificate, passport, eventually, us...	2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vaccines &...	en	0.0	14.0	[]	0.583333	0.533333	positive	happy that the uk is doing well with their vaccines &...	[happy, uk, well, vaccine, amp, great, see, hospital, ad...	4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths coming from? ...	en	18.0	83.0	[]	0.404242	0.020606	positive	4/ so where are these new cases and deaths coming from t...	[4, new, case, death, coming, over60, group, group, risk...	5	\n1363887730527260674	2021-02-22	Alex Berenson	3/ And the Israeli government claims nearly perfect effe...	en	18.0	84.0	[]	0.538333	0.153333	positive	3/ and the israeli government claims nearly perfect effe...	[3, israeli, government, claim, nearly, perfect, effect...	8	\n1363886675844362248	2021-02-22	Alex Berenson	1/ I have to keep saying this: the Israeli's government'...	en	172.0	350.0	[]	0.430303	0.075758	positive	1/ i have to keep saying this: the israeli's government'...	[1, keep, saying, israeli, government, claim, well, vac...
	id	date	name	tweet	language	retweets_count	likes_count	hashtags	subjectivity	polarity	Analysis	text_processed	tweet_clean																																																																								
1	\n1363895502627078147	2021-02-22	LizW	It will start off as just a vaccine certificate /passpo...	en	0.0	13.0	[]	0.712500	0.362500	positive	it will start off as just a vaccine certificate /passpo...	[I, start, vaccine certificate, passport, eventually, us...																																																																								
2	\n1363895430032031749	2021-02-22	Marion	Happy that the UK is doing well with their vaccines &...	en	0.0	14.0	[]	0.583333	0.533333	positive	happy that the uk is doing well with their vaccines &...	[happy, uk, well, vaccine, amp, great, see, hospital, ad...																																																																								
4	\n1363888280597647370	2021-02-22	Alex Berenson	4/ So where are these new cases and deaths coming from? ...	en	18.0	83.0	[]	0.404242	0.020606	positive	4/ so where are these new cases and deaths coming from t...	[4, new, case, death, coming, over60, group, group, risk...																																																																								
5	\n1363887730527260674	2021-02-22	Alex Berenson	3/ And the Israeli government claims nearly perfect effe...	en	18.0	84.0	[]	0.538333	0.153333	positive	3/ and the israeli government claims nearly perfect effe...	[3, israeli, government, claim, nearly, perfect, effect...																																																																								
8	\n1363886675844362248	2021-02-22	Alex Berenson	1/ I have to keep saying this: the Israeli's government'...	en	172.0	350.0	[]	0.430303	0.075758	positive	1/ i have to keep saying this: the israeli's government'...	[1, keep, saying, israeli, government, claim, well, vac...																																																																								

```
In [188]: #most frequent and least frequent words
positive_list = [] # list containing all words of all texts
for x in df3a_positive['tweet_clean']: # loop over lists in df
    positive_list += x # append elements of lists to full list

positive_val_counts = pd.Series(positive_list).value_counts() # make temporary Series to count
positive_val_counts

Out[188]: vaccine      22109
           12390
death       9254
effect      6135
side        5922
...
northrup     1
boyd         1
palindrome   1
yourcarequotient  1
ejaculation   1
Length: 26671, dtype: int64
```

```
In [189]: positive_val_counts[:100]
```

```
Out[189]: vaccine      22109  
          12390  
          death       9254  
          effect      6135  
          side        5922  
          covid       5202  
          covid19     4012  
          people      3903  
          freedom     3045  
          amp         2972  
          get         2686  
          first       2028  
          u           1988  
          vaccination 1942  
          new         1798  
          one         1711  
          day         1681  
          take        1653  
          case        1563  
          many        1516  
          choice      1412  
          refuse      1276  
          like         1250  
          know        1234  
          say          1205  
          would       1204  
          right       1196  
          year        1158  
          need        1120  
          virus        1105  
          dose         1100  
          want        1099  
          good         1092  
          effective   1057  
          health       1032  
          risk         1018  
          even         1011  
          got          989  
          getting      980  
          safe          980  
          pfizer       976  
          today        966  
          make         959  
          life         941  
          trial        926  
          time         917  
          still        916  
          rate         892  
          uk            871  
          also         870  
          week         869  
          2             843  
          dont         822  
          see           820  
          shot          807  
          far            807  
          1             807  
          mask          796  
          vaccinated   792  
          coronavirus 780  
          work          768  
          country      753  
          think         749  
          news          742  
          could        733  
          government   730  
          million      725  
          month        716  
          long          705  
          much          700  
          going         697  
          may           687  
          data          673  
          world         654  
          im            648  
          let            645  
          go             636  
          number        634  
          really        625  
          disease       625  
          severe        619  
          prevent       614  
          way            600  
          flu            594  
          said           592  
          well           581  
          3              579  
          care           578  
          lockdown      575  
          home           560  
          back           559  
          pandemic     554  
          last           553  
          dos            549  
          better         547  
          test           546  
          great          546  
          medical        541  
          thing          537  
          second         531  
          dtype: int64
```

```
In [190]: (pd.Series(nltk.ngrams(positive_list, 2)).value_counts())[100:]
```

```
Out[190]: ((side, effect)      5798
(covid, vaccine)           1860
(, )                      1765
(covid19, vaccine)         1654
(vaccine, )                650
(covid, death)              634
(vaccine, side)             603
(get, vaccine)              570
(effect, vaccine)           444
(covid19, vaccination)      432
(, vaccine)                425
(death, rate)                402
(first, dose)                392
(pfizer, vaccine)             391
(long, term)                  386
(effect, )                   370
(take, vaccine)               370
(death, )                     362
(vaccine, rollout)            346
(death, toll)                  312
(coronavirus, vaccine)        300
(new, case)                   288
(refuse, covid19)              276
(covid19, death)                 272
(death, vaccine)                 269
(vaccine, trial)                 251
(hospitalization, death)       248
(vaccine, death)                 241
(getting, vaccine)                236
(term, side)                   228
(amp, death)                     215
(got, first)                     208
(serious, side)                   207
(covid, vaccination)                206
(wear, mask)                     206
(immune, system)                   203
(second, dose)                   202
(many, people)                   200
(refuse, vaccination)                200
(vaccine, safe)                   199
(moderna, vaccine)                  195
(restriction, refuse)                194
(freedom, choice)                   187
(prevent, restriction)                185
(dose, vaccine)                   182
(flu, vaccine)                     180
(sore, arm)                      180
(effective, preventing)                176
(medical, freedom)                  175
(28, day)                        168
(severe, disease)                   168
(vaccine, today)                   167
(vaccine, amp)                     166
(vaccine, first)                   164
(get, covid)                      164
(vaccine, dos)                     163
(vaccination, )                   159
(care, home)                      159
(first, vaccine)                   158
(case, death)                      158
(received, first)                   156
(good, news)                      156
(taking, vaccine)                   155
(vaccine, administered)                153
(mrna, vaccine)                   152
(death, covid)                      151
(nursing, home)                     150
(clinical, trial)                   148
(public, health)                   147
(get, vaccinated)                   146
(vaccinefreedom, informedconsent)    145
(illness, death)                     144
(herd, immunity)                     144
(freedom, )                          143
(vaccine, work)                      143
(many, death)                        137
(new, vaccine)                      136
(vaccine, available)                  134
(bill, gate)                         134
(effect, covid)                      132
(cause, death)                        131
(vaccine, reaction)                  131
(, got)                            130
(100, effective)                     129
(within, 28)                         124
(choice, )                           124
(informedconsent, )                  123
(vaccine, may)                        123
(experimental, vaccine)                121
(nvic, vaccinefreedom)                 121
(immune, response)                     121
(vaccine, developed)                  119
(effect, far)                         119
(2nd, dose)                           118
(year, old)                           118
(death, covid19)                      118
(vaccine, people)                     118
(need, vaccine)                        117
(risk, death)                          117
(covid, 19)                           117
dtype: int64
```

```
In [191]: (pd.Series(nltk.ngrams(positive_list, 3)).value_counts())[100]
```

```
Out[191]: 
(vaccine, side, effect)      594
(side, effect, vaccine)      412
(side, effect, )              354
(refuse, covid19, vaccination) 251
(term, side, effect)         225
(long, term, side)           208
(serious, side, effect)      206
(restriction, refuse, covid19) 189
(prevent, restriction, refuse) 183
(vaccine, , )                124
(within, 28, day)             123
(vaccinefreedom, informedconsent, ) 122
(nvic, vaccinefreedom, informedconsent) 121
(side, effect, far)           118
(side, effect, covid)          112
(mild, side, effect)          101
(petition, prevent, restriction) 97
(first, dose, vaccine)        97
(today, nvic, vaccinefreedom)   93
(possible, side, effect)       93
(longterm, side, effect)       92
(received, first, dose)        92
(, , vaccine)                 91
(get, covid, vaccine)          90
(dose, covid, vaccine)         89
(got, first, dose)             89
(potential, side, effect)      88
(effect, covid, vaccine)       88
(severe, side, effect)          84
(covid, vaccine, side)          80
(johnson, amp, johnson)        78
(covid19, vaccination, )        76
(major, side, effect)           71
(28, day, positive)            70
(first, dose, pfizer)           68
(dose, pfizer, vaccine)         67
(adverse, side, effect)         67
(death, within, 28)              65
(tvr, subscribe, today)          64
(, side, effect)                64
(first, covid, vaccine)          64
(read, full, article)            64
(side, effect, covid19)          63
(covid19, vaccine, side)          63
(pfizer, covid19, vaccine)        62
(know, side, effect)              62
(got, covid, vaccine)             60
(side, effect, first)             60
(article, tvr, subscribe)        59
(full, article, tvr)              59
(zero, side, effect)              59
(death, , )                      58
(covid19, vaccine, )              57
(getting, covid, vaccine)          56
(refuse, covid, vaccination)      56
(covid, vaccine, today)            56
(far, side, effect)                55
(subscribe, today, nvic)          55
(long, term, effect)              55
(side, effect, may)                55
(first, dose, covid)                54
(day, positive, test)               54
(severe, disease, death)            54
(side, effect, yet)                54
(side, effect, like)                54
(effect, covid19, vaccine)          54
(100, effective, preventing)        53
(people, received, first)            53
(get, covid19, vaccine)              53
(pfizer, covid, vaccine)              52
(health, care, worker)                52
(effect, , )                      52
(one, side, effect)                51
(dose, covid19, vaccine)              51
(experience, side, effect)            51
(getting, covid19, vaccine)          51
(see, side, effect)                50
(vaccine, reaction, read)            50
(reaction, read, full)                50
(freedom, choice, )                  49
(vaccine, dos, administered)        46
(experienced, side, effect)          46
(covid, vaccine, )                  46
(vaccine, 100, effective)            46
(side, effect, whatsoever)            45
(, covid, vaccine)                  45
(side, effect, amp)                  45
(medical, freedom, choice)            44
(enlightened, conversation, vaccination) 44
(no, side, effect)                  44
(u, enlightened, conversation)        44
(conversation, vaccination, health) 44
(join, u, enlightened)                44
(vaccination, health, autonomy)      44
(safe, no, side)                      44
(safe, side, effect)                  43
(wanted, let, know)                  43
(risk, side, effect)                  43
(get, vaccine, first)                  43
(first, covid19, vaccine)              42
dtype: int64
```

Observations on the positive sentiment tweets

There was quite a bit of similarity between this set and the negative sentiment tweets, but some more positive words and word combos emerged. For example:

```
* getting, get, received, want
* safe, zero side effects, 100% effective, vaccine works
* good, better, great, good news
```

```
In [ ]:
```