

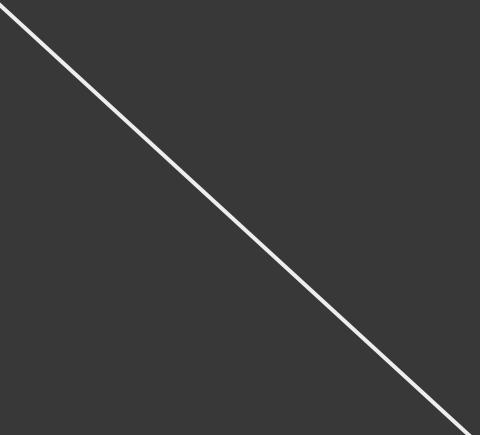


# NLP TOPIC & SENTIMENT ANALYSIS ON THE HK NATIONAL SECURITY LAW

---

16 September 2021 | by Rand Sobczak Jr.

# TABLE OF CONTENTS



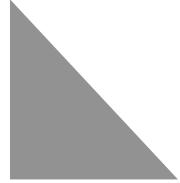
- 1 PROBLEM IDENTIFICATION  
GOAL
- 2 GENERATED DELIVERABLES  
DATA
- 3 EXPLORATORY DATA ANALYSIS  
UNDERSTANDING
- 4 MODEL DESCRIPTION  
DECISION
- 5 MODEL FINDINGS  
MODELING
- 6 READ BETWEEN THE LINES  
EXAMINE
- 7 NEXT STEPS  
FUTURE

A photograph showing a group of young men in dark clothing gathered around a person lying on the ground. One man is holding a long metal rod or stick. The scene suggests a violent confrontation or a protest.

01

# PROBLEM IDENTIFICATION

---

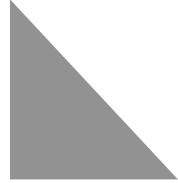


# HISTORY

---

After the end of the First Opium War, **the Qing Dynasty ceded Hong Kong (“HK”) to the United Kingdom (“UK”)**





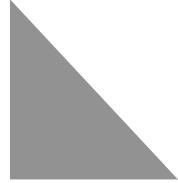
# HYBRID CULTURE

---

HK became a global financial center & its population:

- Developed a hybrid culture
  - Maintaining their Chinese ethos
- Adopting British principles; notably Common Law





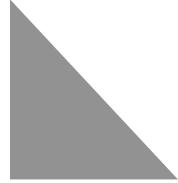
# HANDING BACK

---

The UK & China signed the Sino-British Joint Declaration whereby **the UK agreed to cede HK back to China in 1997** as “one country, two systems” (“OcTs”)

Under this agreement, **OcTs** would be **the structure between 1997 & 2047**





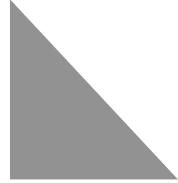
# OCCUPY MOVEMENT

---

In 2014, the Congress of China issued a decision regarding proposed reforms to the HK electoral system

This was widely seen to be highly restrictive, and tantamount to the Chinese Communist Party (“CCP”)’s pre-screening of the candidates for the Chief Executive of HK





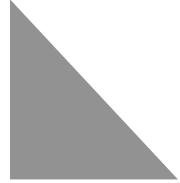
# NEW LAW

---

In 2020, the Congress of China unanimously passed the **National Security Law** (“NSL”) which **criminalizes**

- **Secession**
- **Subversion of state power**
- **Terrorism &**
- **Collusion with Foreign entities in HK**



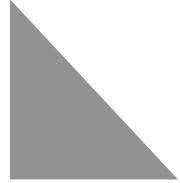


# GOAL

---

A **dichotomy in perceived legal right** to  
enact such law **is where our story begins**





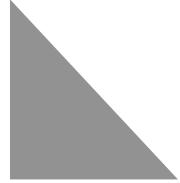
# GOAL

---

A dichotomy in perceived legal right to enact such law is where our story begins

The result was **protests** ranging between 270k to 1M in size\* **ensued**





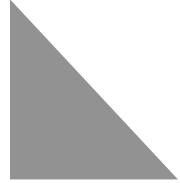
# GOAL

---

A dichotomy in perceived legal right to enact such law is where our story begins

Protests ranging between 270k to 1M in size\* ensued

The **goal of this project is NOT to establish a position on who's right**



# GOAL

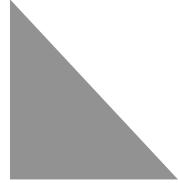
---

A dichotomy in perceived legal right to enact such law is where our story begins

Protests ranging between 270k to 1M in size\* ensued

The goal is NOT to establish a position on who's right.

**The goal is to see where certain groups stand & on what footing**



# GOAL

---

A dichotomy in perceived legal right to enact such law is where our story begins

Protests ranging between 270k to 1M in size\* ensued

The goal is NOT to establish a position on who's right.

The goal is to see where certain groups stand & on what footing **via Topic & Sentiment Analysis using Natural Language Processing** on the NSL



02

## GENERATED DELIVERABLES

---

A word cloud visualization showing the most frequently used words in the generated deliverables. The words are arranged in a cluster, with larger words indicating higher frequency. The color of the words varies, with a primary palette of blues, greens, and yellows.

The most prominent words include:

- nation
- new
- British
- last
- take
- various
- financial
- passed
- want
- central
- long
- safeguarding
- economic
- public
- political
- safeguard
- national
- right
- saying
- free
- list
- used
- enjoy
- endanger
- national
- foreign
- external
- stress
- curb
- legislative
- hope
- calling
- use
- good
- welcome
- addingable
- allow
- celebrate
- create
- existing
- offer
- help
- keep
- announced
- threatened
- necessary
- general
- real
- took
- legitimate
- first
- maintained
- focus
- safe
- lead
- enact
- principle
- proposed
- brighter
- based
- told
- illegal
- think
- current
- stable
- institutional
- radical
- chief
- maintain
- co.

# NEWS (INT'L)

---

Fifteen (15) News Articles from International sources were scraped

|    | file_name      | date       | source       | country | local | title   | article   | word_count |
|----|----------------|------------|--------------|---------|-------|---|---|------------|
| 0  | bbc_1          | 2020-06-30 | BBC          | UK      | Intl  | Hong Kong security law: What is it and is it w... | China has passed a wide-ranging new security l... | 15490      |
| 1  | bbc_2          | 2020-07-01 | BBC          | UK      | Intl  | Hong Kong's new security law: Why it scares pe... | China has introduced a new national security l... | 6021       |
| 2  | nytimes_3      | 2020-06-28 | NY Times     | US      | Intl  | What China's New National Security Law Means f... | Chinese lawmakers have approved a national sec... | 4028       |
| 3  | forbes_4       | 2020-07-01 | Forbes       | US      | Intl  | Hong Kong Makes First Arrests Under Beijing's ... | Share to FacebookShare to TwitterShare to Link... | 4408       |
| 4  | insider_5      | 2020-08-02 | Insider      | US      | Intl  | Teenage arrests, blank protest signs, and a ke... | Thursday, July 30, marked one month since Chin... | 6326       |
| 5  | abc_6          | 2020-07-01 | ABC News     | US      | Intl  | What's in Hong Kong's new national security la... | The full details of the controversial national... | 2420       |
| 6  | cnn_7          | 2020-07-03 | CNN          | US      | Intl  | Hong Kong's security law could have a chilling... | London (CNN Business)Hong Kong insists its vib... | 5683       |
| 7  | nypost_8       | 2020-07-01 | NY Post      | US      | Intl  | Hong Kong police arrest over 300 in first prot... | Hong Kong police arrested more than 300 people... | 1712       |
| 8  | nippon_9       | 2020-08-14 | Nippon       | Japan   | Intl  | Hong Kong's Security Law: How Should Japan Res... | Growing Private and Governmental Support for H... | 6367       |
| 9  | bbc_10         | 2020-06-30 | BBC          | UK      | Intl  | Hong Kong security law: Anger as China's Xi si... | The UK, EU and Nato have expressed concern and... | 11881      |
| 10 | bbc_11         | 2020-06-30 | BBC          | UK      | Intl  | Hong Kong security law: Minutes after new law...  | On Tuesday morning, the news started to break ... | 4712       |
| 11 | bbc_12         | 2020-07-08 | BBC          | UK      | Intl  | Hong Kong security law: Beijing security offic... | A new national security office has been offici... | 3238       |
| 12 | bbc_13         | 2020-07-05 | BBC          | UK      | Intl  | Hong Kong security law: Pro-democracy books pu... | Books by pro-democracy figures have been remov... | 4325       |
| 13 | bbc_14         | 2020-06-30 | BBC          | UK      | Intl  | Hong Kong security law: Life sentences for bre... | People in Hong Kong could face life in jail fo... | 6500       |
| 14 | theDiplomat_15 | 2020-07-01 | the Diplomat | US      | Intl  | Hong Kong Through Water and Fire                  | Credit: AP Photo/Kin CheungAdvertisement! has...  | 9868       |

# NEWS (CHINESE)

Fifteen (15) News Articles from Chinese sources were also scraped

|    | file_name          | date       | source                      | country   | local | title  | article  | word_count |
|----|--------------------|------------|-----------------------------|-----------|-------|--|--|------------|
| 15 | peoples_daily_1    | 2020-07-03 | People's Daily              | China     | Local | Newly passed national security law a seawall s...  | The Law of the People's Republic of China on S...  | 3486       |
| 16 | peoples_daily_2    | 2020-07-02 | People's Daily              | China     | Local | Hong Kong national security law helps ensure i...  | On June 30, the Standing Committee of the 13th...  | 8028       |
| 17 | peoples_daily_3    | 2020-07-02 | People's Daily              | China     | Local | Law and order dawns in Hong Kong as new law ta...  | -- The festive mood across Hong Kong on the 23...  | 9627       |
| 18 | peoples_daily_4    | 2020-07-02 | People's Daily              | China     | Local | National Security Law paves the way for more p...  | Hong Kong, the popular Asian city, returned to...  | 9498       |
| 19 | peoples_daily_5    | 2020-07-02 | People's Daily              | China     | Local | HKSAR national security law to put HK back on ...  | Zhang Xiaoming, deputy director of the Hong Ko...  | 6744       |
| 20 | globaltimes_6      | 2021-04-19 | Global Times                | China     | Local | Biased tone, misinformation 'major mistakes' B...  | Victor Gao (center), chair professor of Soochow... | 16593      |
| 21 | globaltimes_7      | 2020-07-06 | Global Times                | China     | Local | London's measures meaningless, 'bluff rather t...  | A Huawei store stands next to a Globe Telecom ...  | 2960       |
| 22 | globaltimes_8      | 2020-06-30 | Global Times                | China     | Local | National Security Law to protect HK democracy...   | Hong Kong citizens on Tuesday gather to suppor...  | 3711       |
| 23 | min_foreign_aff_9  | 2020-07-20 | Ministry of Foreign Affairs | China     | Local | Ambassador Liu Xiaoming Gives Exclusive Live I...  | On 19 July 2020, H.E. Ambassador Liu Xiaoming ...  | 5324       |
| 24 | min_foreign_aff_10 | 2020-07-01 | Ministry of Foreign Affairs | China     | Local | Foreign Ministry Spokesperson Zhao Lijian's Re...  | In recent years, the US government has placed ...  | 18122      |
| 25 | xinhua_11          | 2020-05-16 | Xinhua                      | China     | Local | IPCC's report comprehensive, objective, fact-b...  | Video PlayerClose\n\n\nCarrie Lam, chief exe...    | 806        |
| 26 | xinhua_12          | 2020-07-01 | Xinhua                      | China     | Local | China adopts law on safeguarding national secur... | Video PlayerClose\n\n\n Li Zhanshu, chairman...    | 4315       |
| 27 | theStandard_13     | 2020-08-26 | the Standard                | China     | Local | State media says Hong Kong has a toxic 'Apple'     | State media People's Daily today published an ...  | 3079       |
| 28 | china_daily_14     | 2020-06-20 | China Daily                 | China     | Local | Wide support for proposed national security law    | Residents sign to support the national securit...  | 4294       |
| 29 | hkgovt_15          | 2020-06-30 | Hong Kong Gov't             | Hong Kong | Local | CE welcomes passage of The Law of the People's...  | In response to the passage of The Law of the P...  | 6143       |

# NEWS CLEANING & INITIAL PREP

---

- The **cleaning & initial preparation** for both the **Titles & Articles** (📄) themselves **is the same**. They were eventually combined

# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles themselves is the same. They were eventually combined
- To begin, **they were all scraped using beautifulsoup**

Beautifulsoup



Scraped



Models

# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles themselves is the same. They were eventually combined  
To begin, they were all scraped using beautifulsoup
- Initial **formatting for the dates** were moved **to a datetime object**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   file_name   30 non-null    object  
 1   date        30 non-null    object  
 2   source      30 non-null    object  
 3   country     30 non-null    object  
 4   local        30 non-null    object  
 5   title        30 non-null    object  
 6   article      30 non-null    object  
 7   word_count   30 non-null    int64  
dtypes: int64(1), object(7)
memory usage: 2.0+ KB
```

```
news['date'] = pd.to_datetime(news['date'], format='%Y-%m-%d')
news.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype    
--- 
 0   file_name   30 non-null    object  
 1   date        30 non-null    datetime64[ns]
 2   source      30 non-null    object  
 3   country     30 non-null    object  
 4   local        30 non-null    object  
 5   title        30 non-null    object  
 6   article      30 non-null    object  
 7   word_count   30 non-null    int64  
dtypes: datetime64[ns](1), int64(1), object(6)
memory usage: 2.0+ KB
```



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles themselves is the same. They were eventually combined
- To begin, they were all scraped using beautifulsoup
- Initial formatting for the dates were moved to a datetime object
- **Non-ascii text was removed**

```
'State media says Hong Kong has a toxic 'Apple'  
"State media says Hong Kong has a toxic 'Apple'"
```



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles ( ) themselves is the same. They were eventually combined
- To begin, they were all scraped using beautifulsoup
- Initial formatting for the dates were moved to a datetime object
- Non-ascii text was removed
- The Sentiment Intensity Analyzer (“SIA”) was then **added to determine the Positive, Negative or Neutral position of the text**

|       | comp_article | neg_article | neu_article | pos_article |
|-------|--------------|-------------|-------------|-------------|
| count | 30.000000    | 30.000000   | 30.000000   | 30.000000   |
| mean  | 0.338153     | 0.087933    | 0.795733    | 0.116567    |
| std   | 0.897364     | 0.035946    | 0.034230    | 0.036774    |
| min   | -0.997900    | 0.013000    | 0.710000    | 0.064000    |
| 25%   | -0.754575    | 0.068500    | 0.772500    | 0.089000    |
| 50%   | 0.979150     | 0.093500    | 0.785500    | 0.115500    |
| 75%   | 0.996675     | 0.110000    | 0.822750    | 0.132500    |
| max   | 0.999600     | 0.145000    | 0.860000    | 0.207000    |



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles themselves is the same. They were eventually combined  
To begin, they were all scraped using beautifulsoup  
Initial formatting for the dates were moved to a datetime object  
Non-ascii text was removed  
The Sentiment Intensity Analyzer was then added to determine the Positive, Negative or Neutral position of the text
- Stop words are a **commonly used words which don't provide value to NLP**; examples are “a”, “the”, “is”, “are”, “and”; they were removed

```
# Before Stop Word Removal
print(df_clus['title'][4][0:47])
```

Teenage arrests, blank protest signs, and a key

```
# After Stop Word Removal
print(data_words[4][0:6])
```

['teenage', 'arrest', 'blank', 'protest', 'sign', 'key']



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles ( ) themselves is the same. They were eventually combined
  - To begin, they were all scraped using beautifulsoup
  - Initial formatting for the dates were moved to a datetime object
  - Non-ascii text was removed
  - The Sentiment Intensity Analyzer was then added to determine the Positive, Negative or Neutral position of the text
    - Stop words are commonly used words which don't provide value to NLP; examples are "a", "the", "is", "are", "and"; they were removed
- Lemmatization (“Lem”) | **the process of grouping together the inflected forms of a word so they can be analyzed as a single item**, identified by the word's lemma, or dictionary form

```
# Title | Before Lemmatization
print(df_clus['title'][15][0:21])
```

Newly passed national

```
# Title | After Lemmatization
print(lemmatized[15][0:19])
```

newly pass national



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles ( ) themselves is the same. They were eventually combined  
To begin, they were all scraped using beautifulsoup  
Initial formatting for the dates were moved to a datetime object  
Non-ascii text was removed  
The Sentiment Intensity Analyzer was then added to determine the Positive, Negative or Neutral position of the text  
Stop words are commonly used words which don't provide value to NLP; examples are "a", "the", "is", "are", "and"; they were removed  
Stemming ("Stem") | The process of reducing words to its stem that affixes to suffixes & prefixes or to the roots of words known as a lemma  
Lemmatization ("Stem") | the process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's lemma, or dictionary form
- **Tfidf** | Term Frequency Inverse Document Frequency; **an algorithm to transform text into a meaningful representation of numbers** which we can use to fit machine algorithms for prediction



# NEWS CLEANING & INITIAL PREP

---

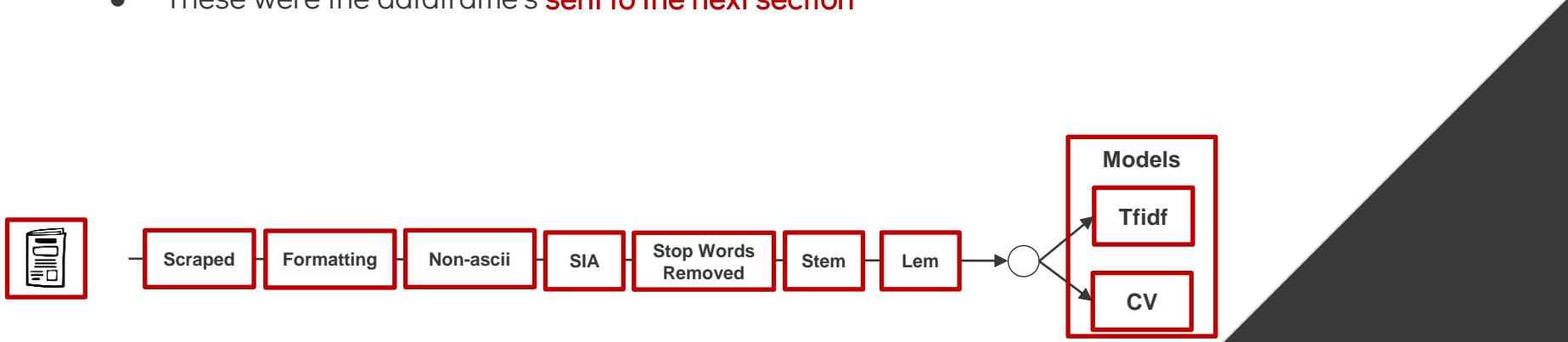
- The cleaning & initial preparation for both the Titles & Articles ( ) themselves is the same. They were eventually combined  
To begin, they were all scraped using beautifulsoup  
Initial formatting for the dates were moved to a datetime object  
Non-ascii text was removed  
The Sentiment Intensity Analyzer was then added to determine the Positive, Negative or Neutral position of the text  
Stop words are commonly used words which don't provide value to NLP; examples are "a", "the", "is", "are", "and"; they were removed  
Stemming ("Stem") | The process of reducing words to its stem that affixes to suffixes & prefixes or to the roots of words known as a lemma  
Lemmatization ("Stem") | the process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's lemma, or dictionary form
- TfIdf | Term Frequency Inverse Document Frequency; a common algorithm to transform text into a meaningful representation of numbers which we can use to fit machine algorithms for prediction
- CountVectorizer ("CV") | the **CountVectorizer is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text**



# NEWS CLEANING & INITIAL PREP

---

- The cleaning & initial preparation for both the Titles & Articles ( ) themselves is the same. They were eventually combined
- To begin, they were all scraped using beautifulsoup
- Initial formatting for the dates were moved to a datetime object
- Non-ascii text was removed
- The Sentiment Intensity Analyzer was then added to determine the Positive, Negative or Neutral position of the text
- Stop words are commonly used words which don't provide value to NLP; examples are "a", "the", "is", "are", "and"; they were removed
- Stemming ("Stem") | The process of reducing words to its stem that affixes to suffixes & prefixes or to the roots of words known as a lemma
- Lemmatization ("Stem") | the process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's lemma, or dictionary form
- TfIdf | Term Frequency Inverse Document Frequency; a common algorithm to transform text into a meaningful representation of numbers which we can use to fit machine algorithms for prediction
- CountVectorizer ("CV") | the CountVectorizer is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text
- These were the dataframe's **sent to the next section**



# TWITTER

**288k Tweets** were scraped on twelve (12) hashtags; these are all anti-Law hashtags

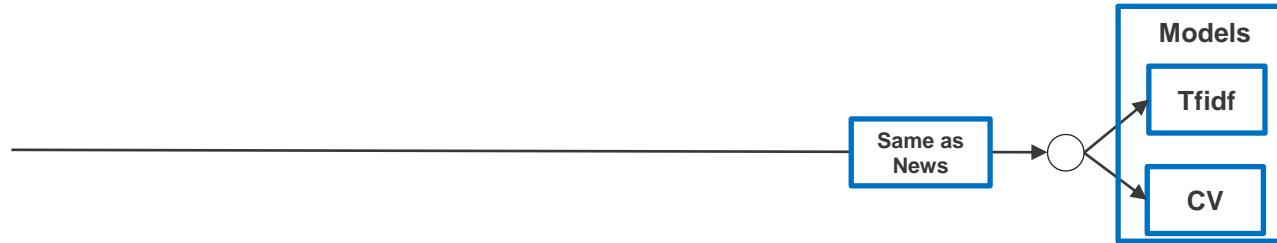
**PLEASE NOTE |** these exclusively arise from HK & Int'l Tweeters; Chinese social media's contribution is NOT present, unfortunately

|        | hash                | created_at | username        | tweet  | language | replies_count | retweets_count | likes_count | reply_to  | hashtags   |
|--------|---------------------|------------|-----------------|--|----------|---------------|----------------|-------------|---|--|
| 7      | #FightForHongKong   | 2020-08-23 | cipopopolitik   | China will never stop.<br>#HongKongProtest<br>#Figh... | en       | 0             | 0              | 0           | □   | ['hongkongprotest', 'fightforhongkong', 'prayf...  |
| 8003   | #HongKongProtest    | 2020-06-04 | paulisally      | Hong Kong Police is MURDERER!!!!!! 흐름 경찰은 살인범...       | ja       | 0             | 0              | 1           | □   | ['policebrutality', 'hongkongpolice', 'hongk...    |
| 8503   | #PrayForHongkong    | 2020-06-30 | cartwheelwombat | Please pray for Hong Kong. You can follow him...       | en       | 1             | 2              | 2           | □   | ['prayforhongkong', 'standwithhongkong', 'hong...  |
| 9025   | #WeAreNotAlone      | 2020-07-04 | whats_a_gurg    | There's not much more satisfying than seeing o...      | en       | 0             | 0              | 0           | □   | ['wearenotalone', 'questioneverything']            |
| 9307   | #nochinaextradition | 2020-05-31 | techno_tech4u   | Remove China Apps' crosses 1 million download...       | en       | 0             | 0              | 0           | □   | ['china', 'chinaapp', 'antichina', 'nochinext...   |
| 24003  | #standwithhongkong  | 2020-08-17 | minsguga1993309 | Please #StandWithHongKong!                             | en       | 0             | 0              | 1           | □   | ['standwithhongkong']                              |
| 175007 | #hkprotests         | 2020-08-24 | juniushogodie   | It's also the 1st time MTR closed its station ...      | en       | 0             | 7              | 17          | □   | ['hkprotests']                                     |
| 180023 | #freehongkong       | 2020-08-27 | pppongo         | @StateDept @SecPompeo Thank you U.S so much fo...      | en       | 0             | 0              | 0           | [[{"screen_name": "StateDept", "name": "Department..."}]] | ['standup4humanrights', 'standwithhongkong', ...]  |
| 260001 | #hkpolicebrutality  | 2020-06-16 | akaribug        | They keep telling us to trust a liar while tur...      | en       | 0             | 1              | 0           | □   | ['hkpolicebrutality', 'xinjiang', 'standwithho...  |
| 280017 | #hkpolicestate      | 2020-06-14 | ylye33          | A 17 years old girl was attacked by the #HKpol...      | en       | 0             | 0              | 3           | □   | ['hkpolice', '612protest', 'hkpolicebrutality'...] |
| 288002 | #hkpolicererrorism  | 2020-06-08 | leung18188868   | @foxthepopo They are not police. ....                  | en       | 0             | 1              | 0           | [[{"screen_name": "foxthepopo", "name": "iamthe..."}]]    | ['hkpolicebrutality', 'hkpolicererrorism', 'hk...  |
| 284002 | #HKpoliceterrorist  | 2020-07-25 | bethenewyorkers | @PENamerica Journalists are always the target ...      | en       | 0             | 2              | 2           | [[{"screen_name": "PENamerica", "name": "PEN Am..."}]]    | ['hkpolicererrorist', 'nationalsecuritylaw']       |

# TWITTER CLEANING & INITIAL PREP

---

- Twitter had the **same process** as the News **but scraped with Twint (T)** and had five (**5**) steps before

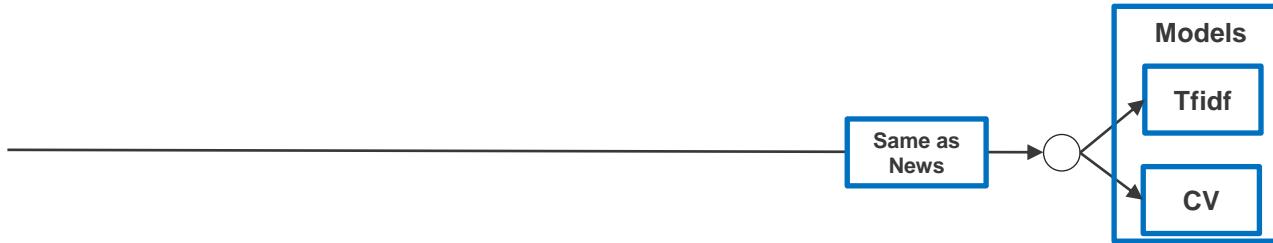


# TWITTER CLEANING & INITIAL PREP

---

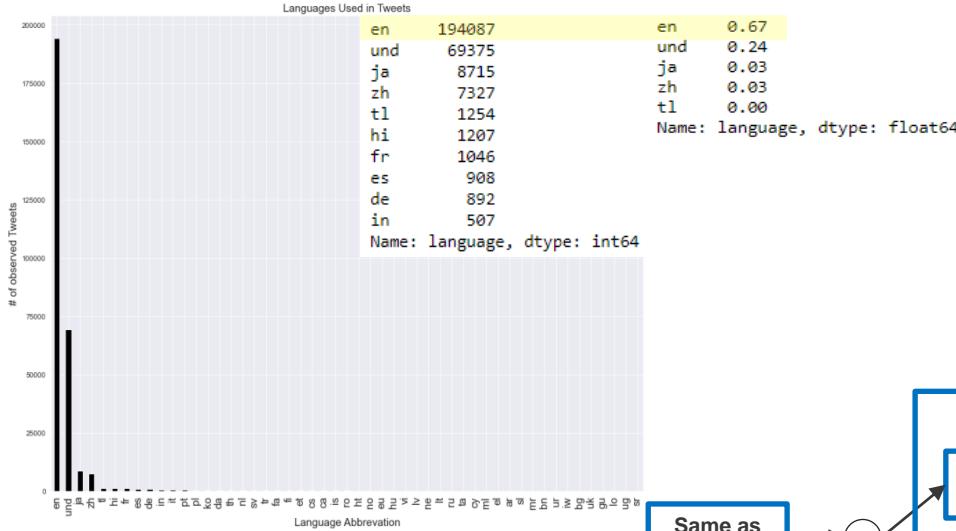
 288,416

- Twitter had the same process as the News but scraped with Twint (⌚) and had five (5) steps before
  - This **generated 288,416 Tweets** for the DataFrame



# TWITTER CLEANING & INITIAL PREP

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before
- Focused on English Tweets

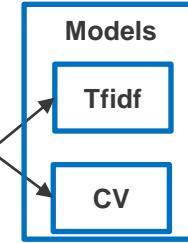


288,416

194,087



Same as  
News



# TWITTER CLEANING & INITIAL PREP

---

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before Focused on English Tweets
- Removed Duplicates

 288,416

 194,087

 163,699



# TWITTER CLEANING & INITIAL PREP



163,699

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before
- Focused on English Tweets
- Removed Duplicates
- Mentions of other users & links were removed

```
df[['hash', 'created_at', 'tweet']].head(1)
```

| hash                | created_at | tweet   |
|---------------------|------------|---|
| 0 #FightForHongKong | 2020-08-28 | @benedictrogers @NOW4humanity Thanks for speak... |

```
df['tweet'] = df['tweet'].str.replace("@[A-Za-z0-9]+", "")
```

```
df[['hash', 'created_at', 'tweet']].head(1)
```

| hash                | created_at | tweet   |
|---------------------|------------|---|
| 0 #FightForHongKong | 2020-08-28 | Thanks for speaking up for us #fridaysforfre... |

```
df['tweet'][9]
```

```
'Joshua Wong: 'Forms of resistance need to be fluid and flexible' #FightForHongKong #HongKongProtest https://t.co/COPpnczk p4'
```

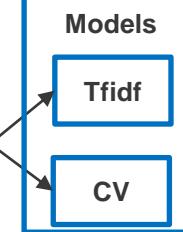
```
df['tweet'] = df['tweet'].str.replace(r'HTTP\S+', '')
```

```
df['tweet'][9]
```

```
'Joshua Wong: 'Forms of resistance need to be fluid and flexible' #FightForHongKong #HongKongProtest '
```



Same as  
News



# TWITTER CLEANING & INITIAL PREP



163,699

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before
- Focused on English Tweets
- Removed Duplicates
- Mentions of other users & links were removed
- **Emoji's were then converted to bigrams & trigrams & kept**

```
df.tweet[12]
```

```
'Heartbreaking...😢 we have to do whatever we can do to #FightForHongKong #StandWithHongKong'
```

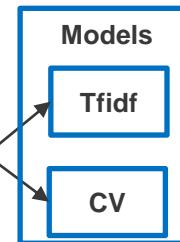
```
df['tweet'] = df.tweet.apply(lambda x: convert_emojis_to_word(x))
```

```
df.tweet[12]
```

```
'Heartbreaking... loudly_crying_face we have to do whatever we can do to #FightForHongKong #StandWithHongKong'
```



Same as  
News



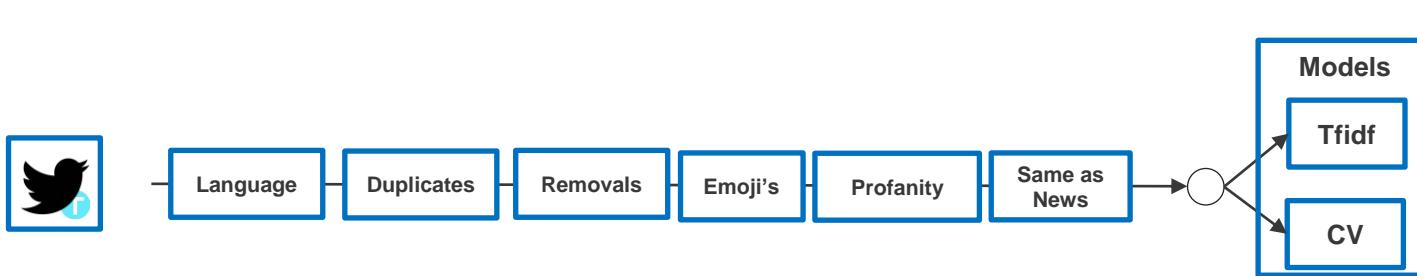
# TWITTER CLEANING & INITIAL PREP

---



163,699

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before Focused on English Tweets
  - Removed Duplicates
  - Mentions of other users & links were removed
  - Emoji's were then converted to words & kept
- All **profanity** were then **converted to the word “NEGATIVE”**; this word was chosen to be well positioned in SIA



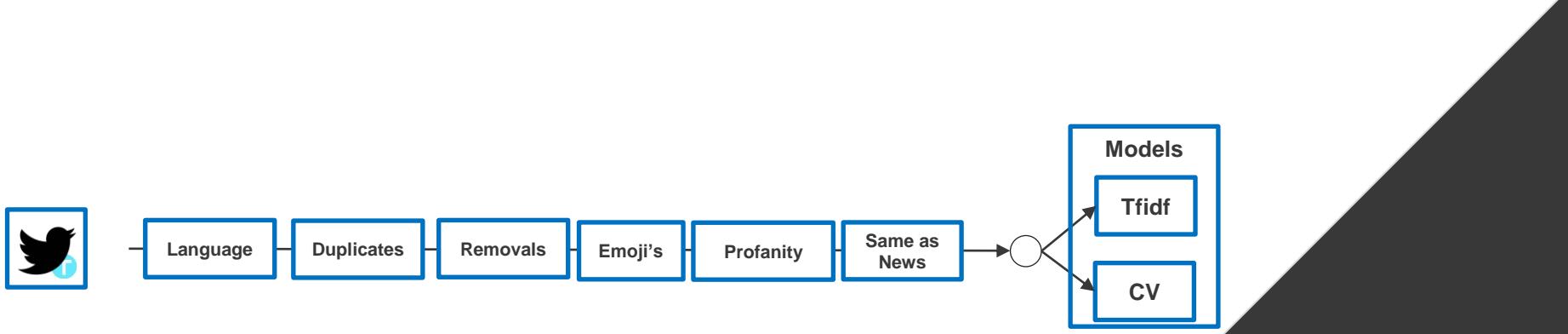
# TWITTER CLEANING & INITIAL PREP

---



163,699

- Twitter had the same process as the News but scraped with Twint and had five (5) steps before Focused on English Tweets
- Removed Duplicates
- Mentions of other users & links were removed
- Emoji's were then converted to words & kept
- All profanity were then converted to the word "NEGATIVE"; this word was chosen to be well positioned in SIA
- Then **sent to the next section** as well



# GENERATED DELIVERABLES

---



## SOURCE CODE

Each are found on my  
GitHub account  
referenced at the end



## RESEARCH REPORT

Each are found on my  
GitHub account  
referenced at the end



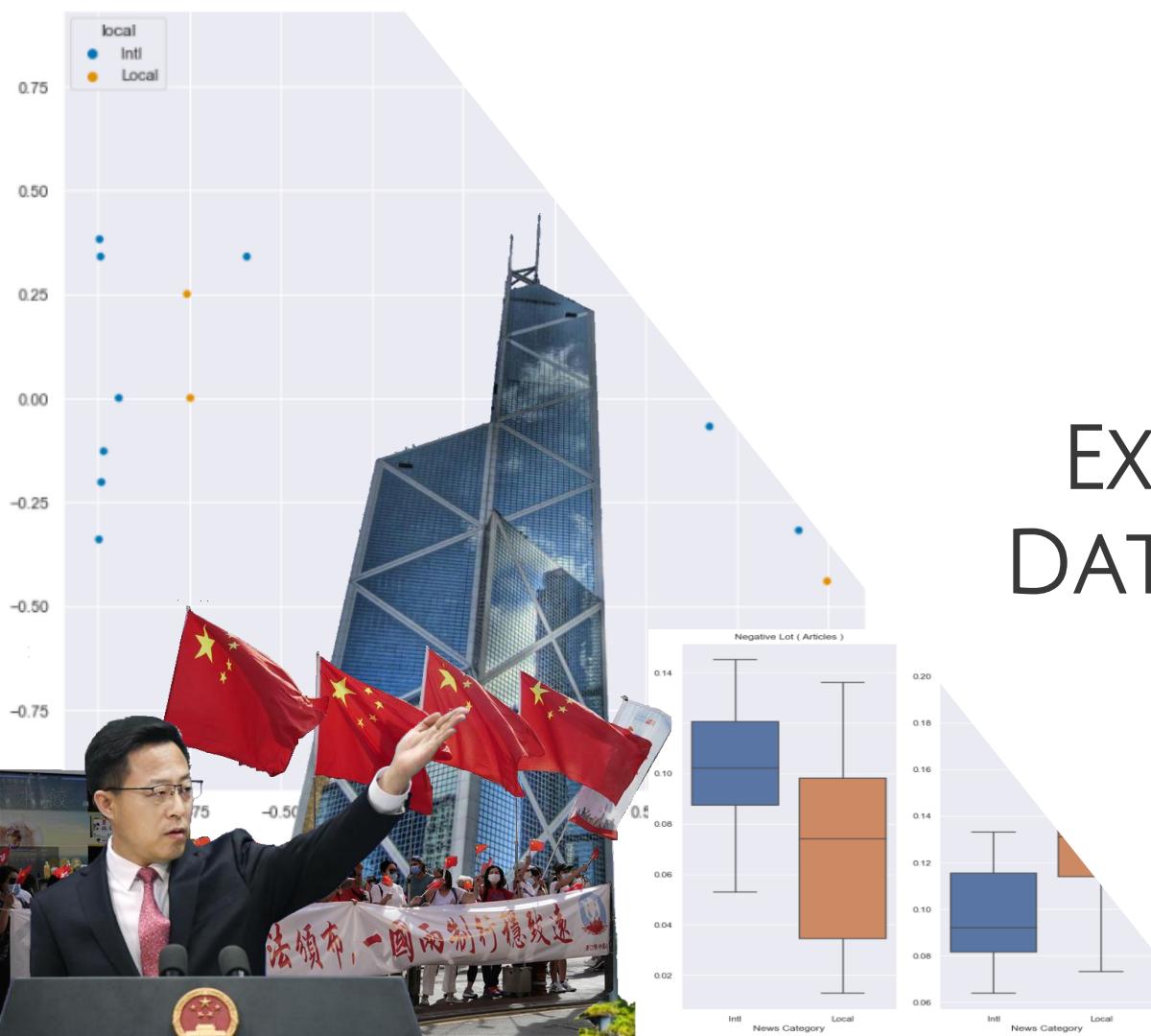
## PRESENTATION REPORT

This one...

# 03

## EXPLORATORY DATA ANALYSIS

---



# EXPLORATORY DATA ANALYSIS

---

- The **Exploratory data analysis** (“EDA”) was undertaken on both the News Articles & Tweets

# EXPLORATORY DATA ANALYSIS

---

- The Exploratory data analysis (“EDA”) was undertaken on both the News Articles & Tweets
- The **News & Tweets** were **done separately**

# EXPLORATORY DATA ANALYSIS

---

- The Exploratory data analysis (“EDA”) was undertaken on both the News Articles & Tweets
- They were done separately
- **The News will be reviewed first**, followed by the Tweets

# NEWS EDA

---

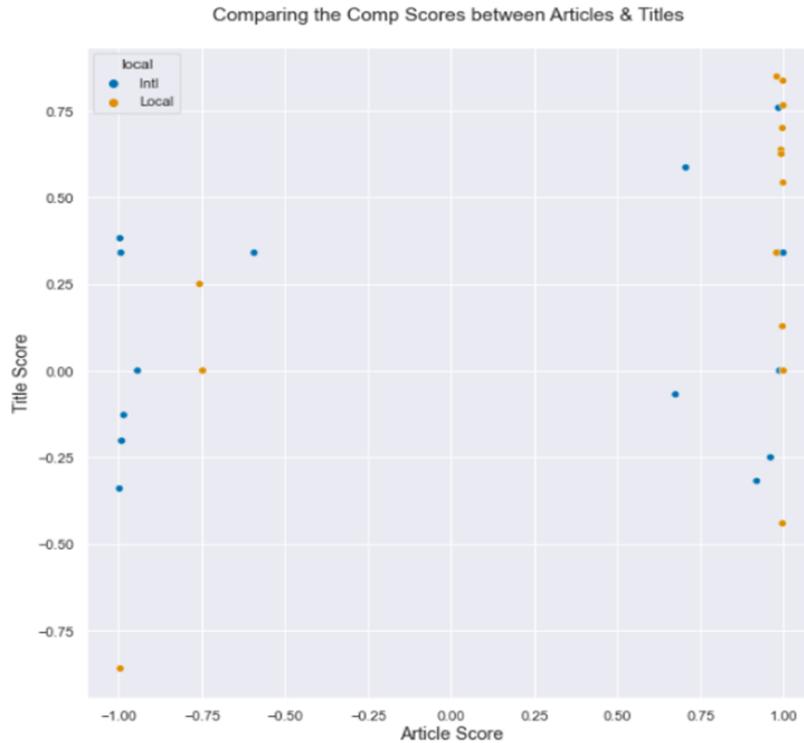
With the help of Sentiment Analysis using VADER, the **initial observation** of the News Articles was **the lack of a middle ground\*** on the Articles

Comp Score



# NEWS EDA

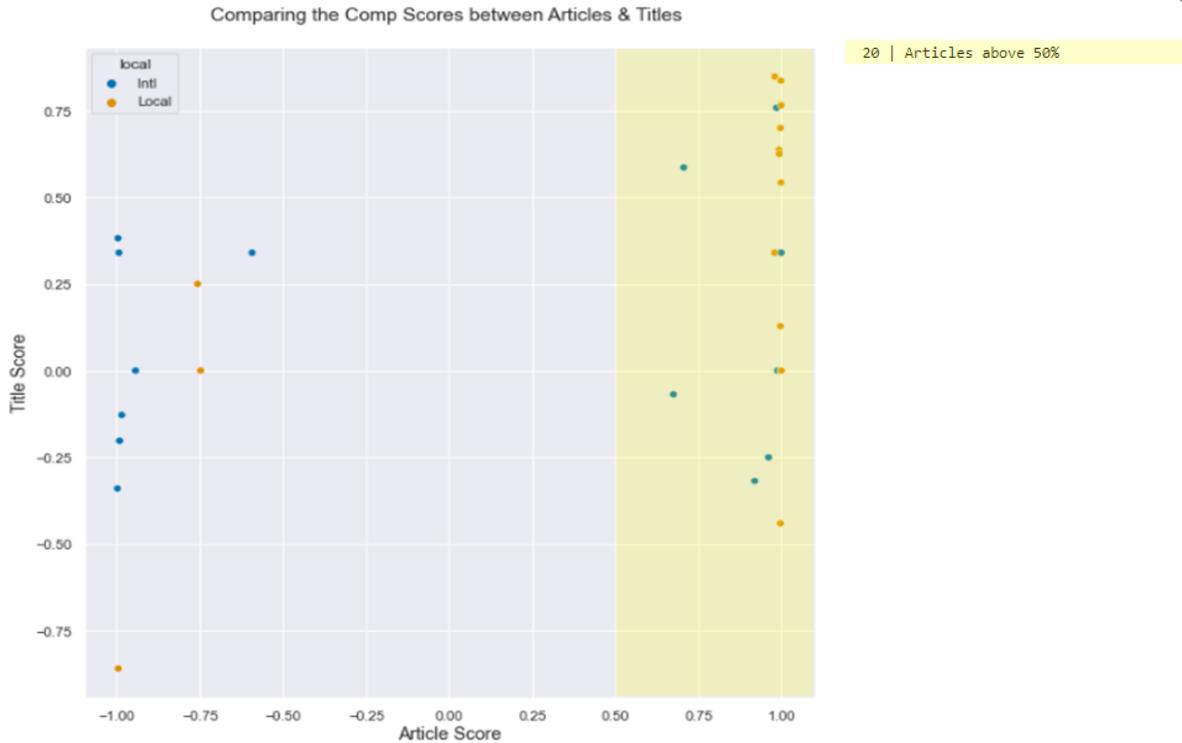
The initial observation was that the Sensitivity Scores for Titles don't "always" align with the Articles



# NEWS EDA

Digging deeper |

20 Positive Articles

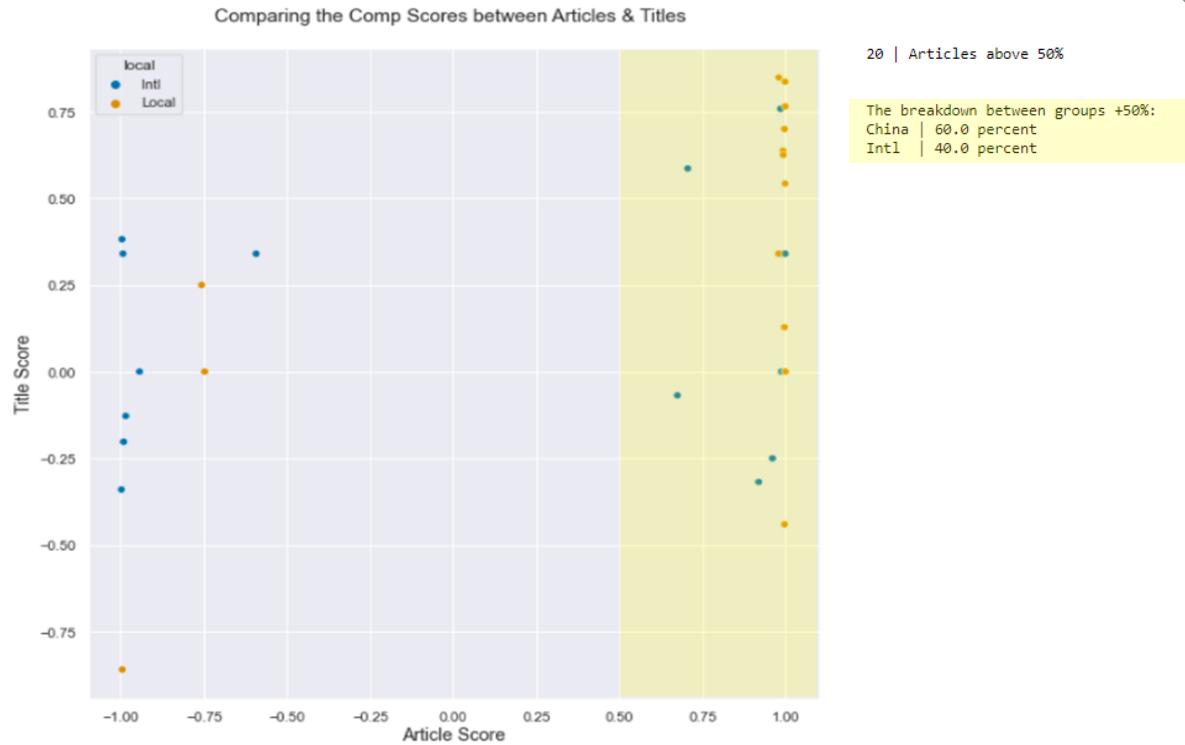


\* All the Sentiment Scores present Positivity as 1 & conversely Negativity as -1

# NEWS EDA

Digging deeper |

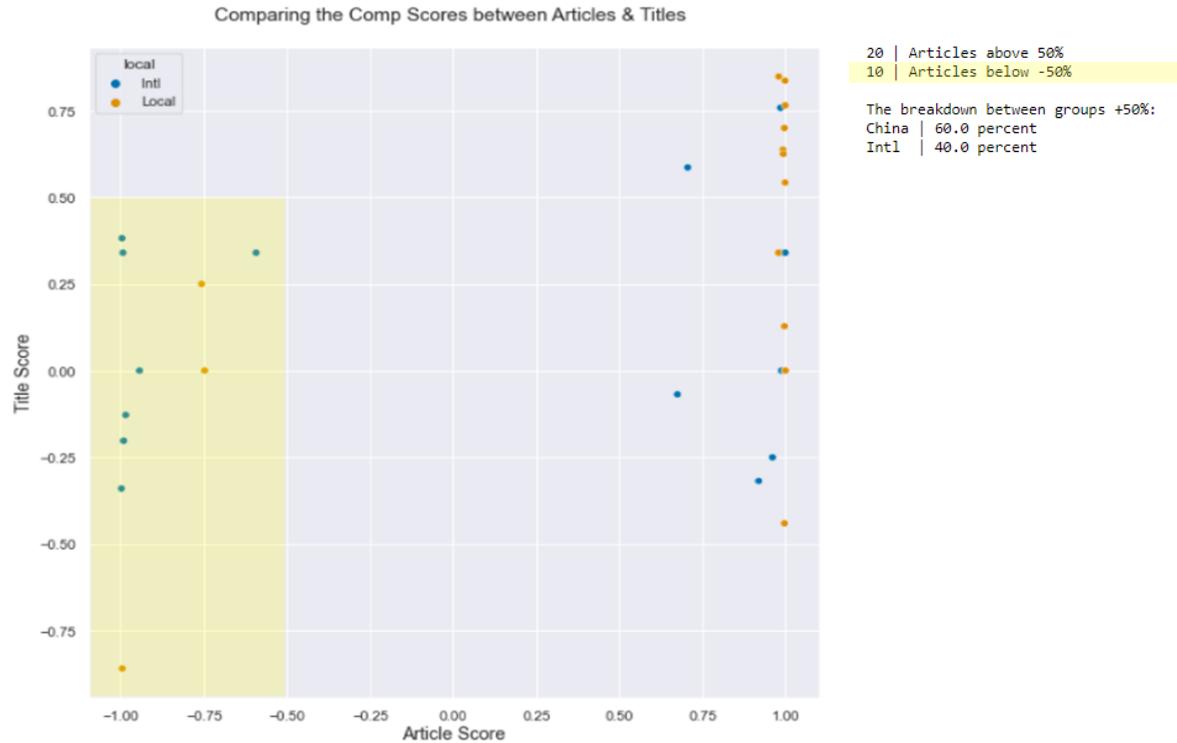
20 Positive Articles | China controls this category but not domineering



# NEWS EDA

Digging deeper |

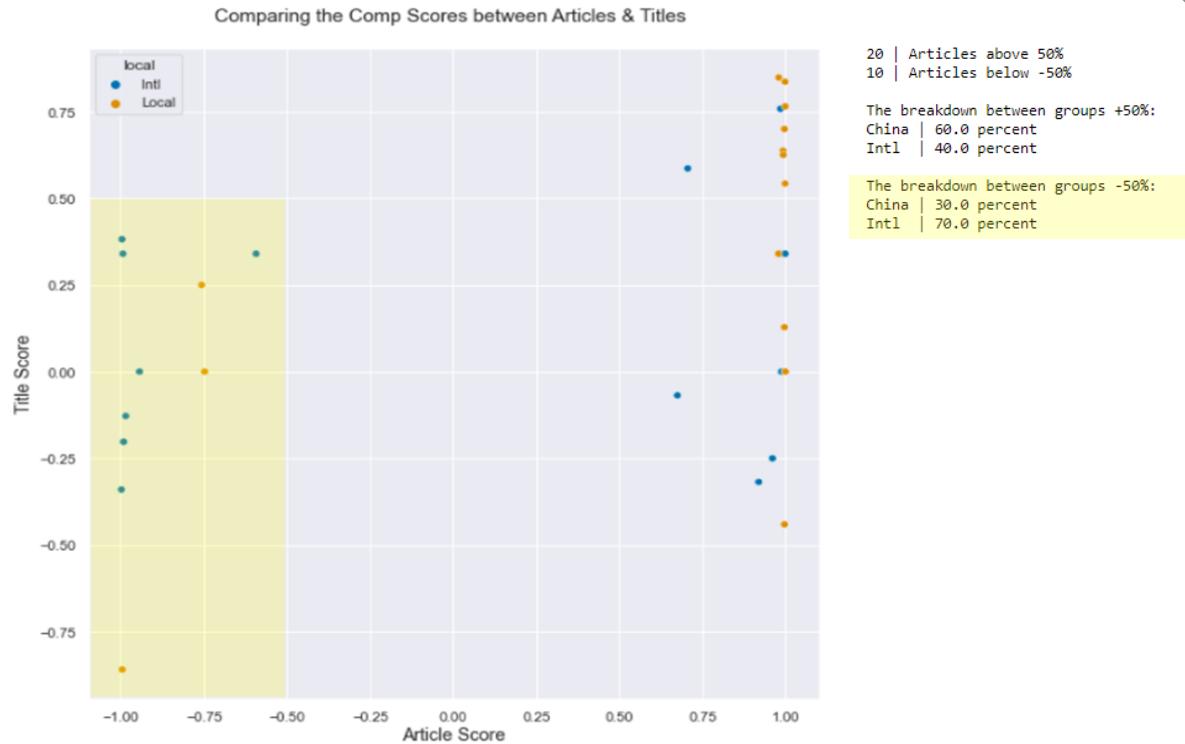
Conversely, **10 Negative Articles**



# NEWS EDA

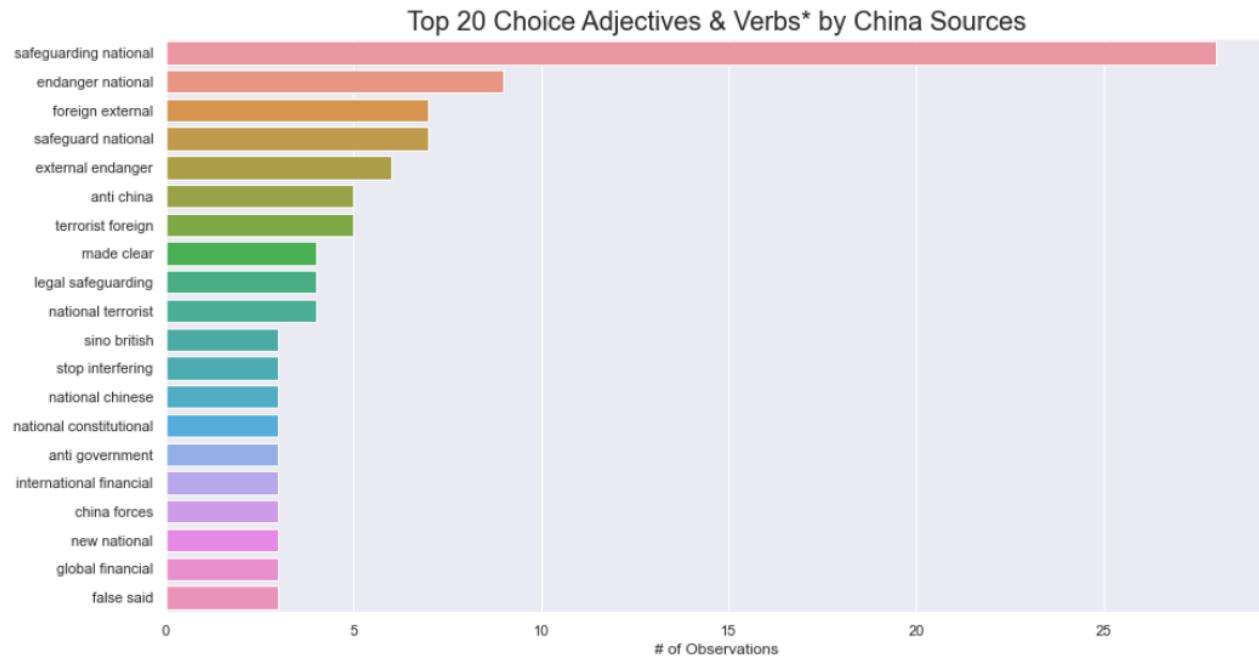
Digging deeper |

10 Negative Articles | International News controls this category but it's also not domineering



# NEWS EDA

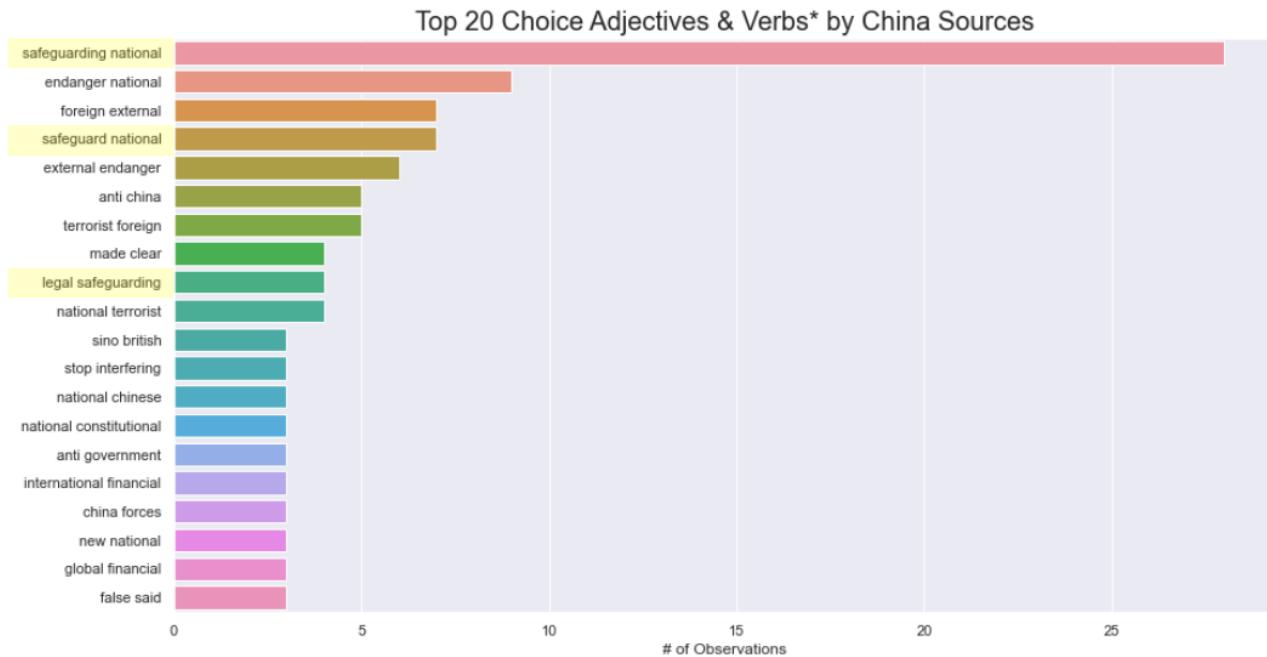
A deeper look with Scikit-learn's CountVectorizer conveys that notable mentions of **safeguarding** & **externalities** show up **in Chinese News**



\*These are words arrived through a Bi-gram

# NEWS EDA

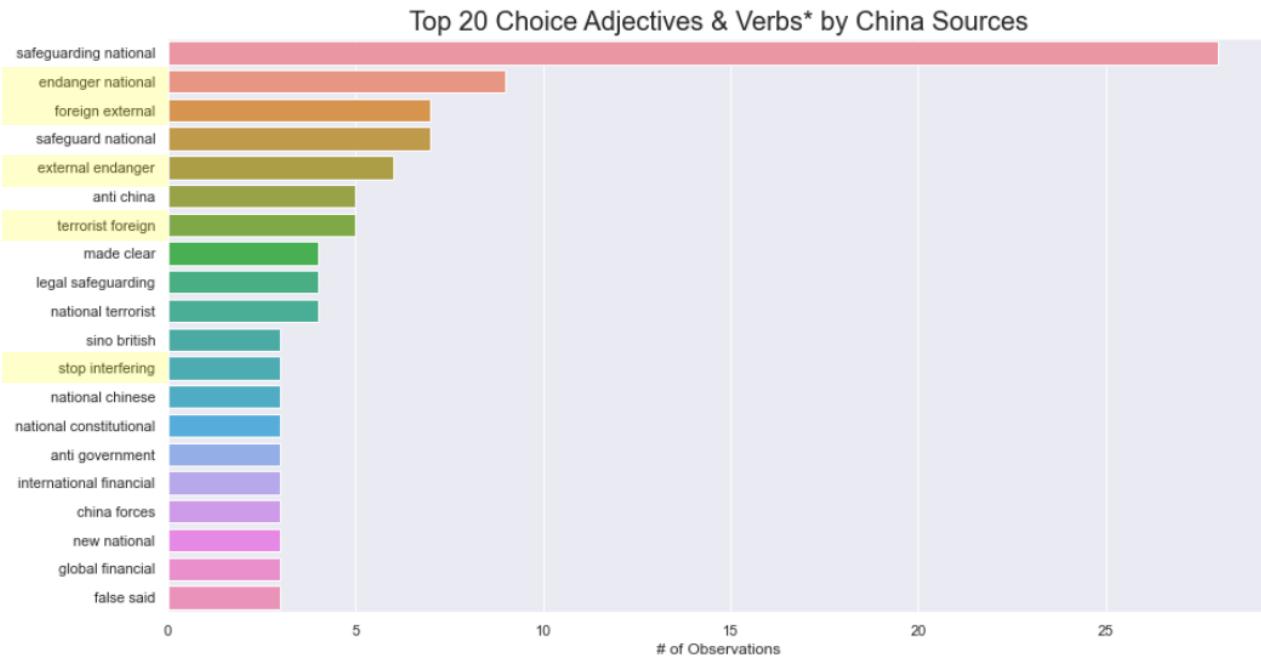
A deeper look with Scikit-learn's CountVectorizer conveys that notable mentions of **safeguarding** & **externalities** show up **in Chinese News**



\*These are words arrived through a Bi-gram

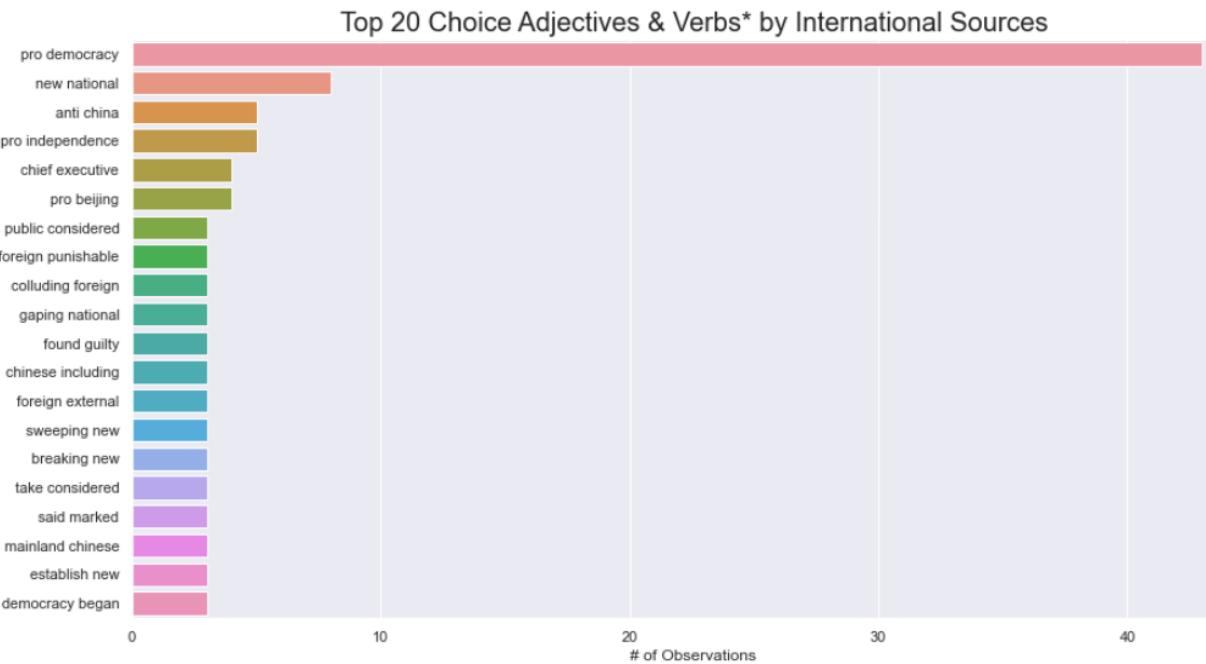
# NEWS EDA

A deeper look with Scikit-learn's CountVectorizer conveys that notable mentions of **safeguarding** & **externalities** show up **in Chinese News**



# NEWS EDA

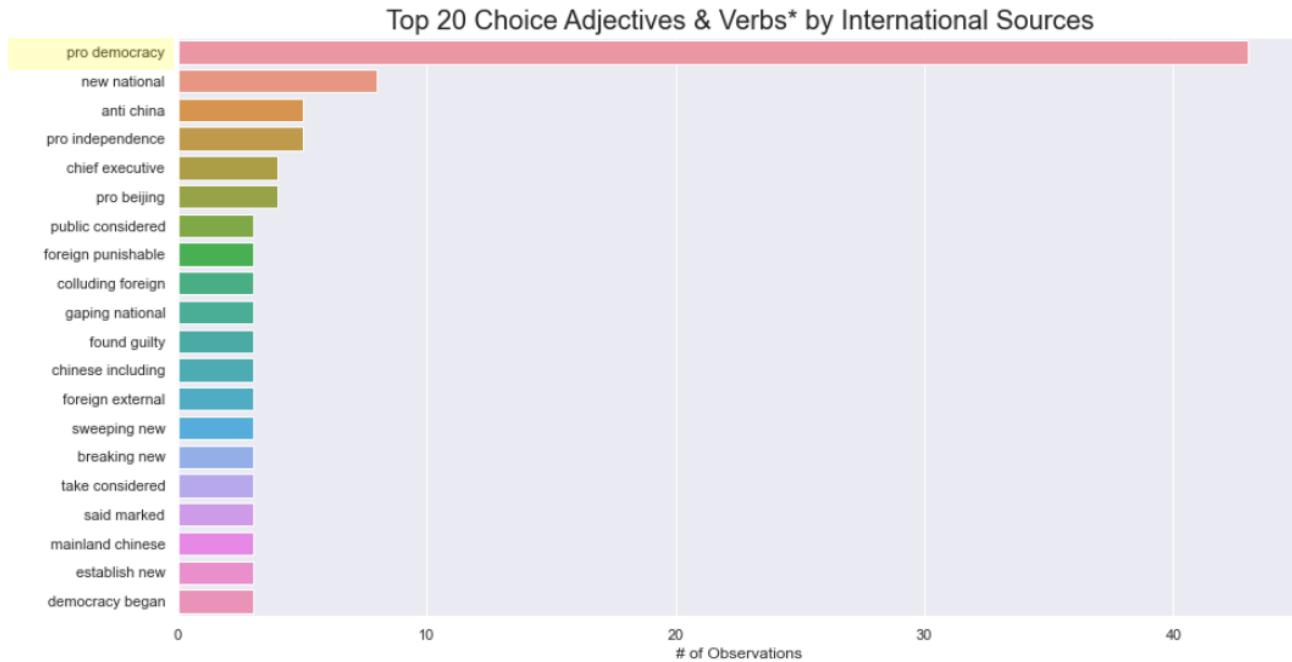
Pro democracy commands a notable majority in International News



\*These are words arrived through a Bi-gram

# NEWS EDA

Pro democracy commands a notable majority in International News



\*These are words arrived through a Bi-gram



## CHINA | POSITIVE

China & the Law's intent is to protect & support the freedom & prosperity of HK as a unified motherland



## INT'L | POSITIVE

No clear indication for a positive position; to be confirmed in later stages

# NEWS EDA TAKE-AWAYS



## CHINA | NEGATIVE

The West may be misinformed or not see the longer term as well as endangering HK by interfering



## INT'L | NEGATIVE

Concern over the freedom of the HK people, particularly the protestors

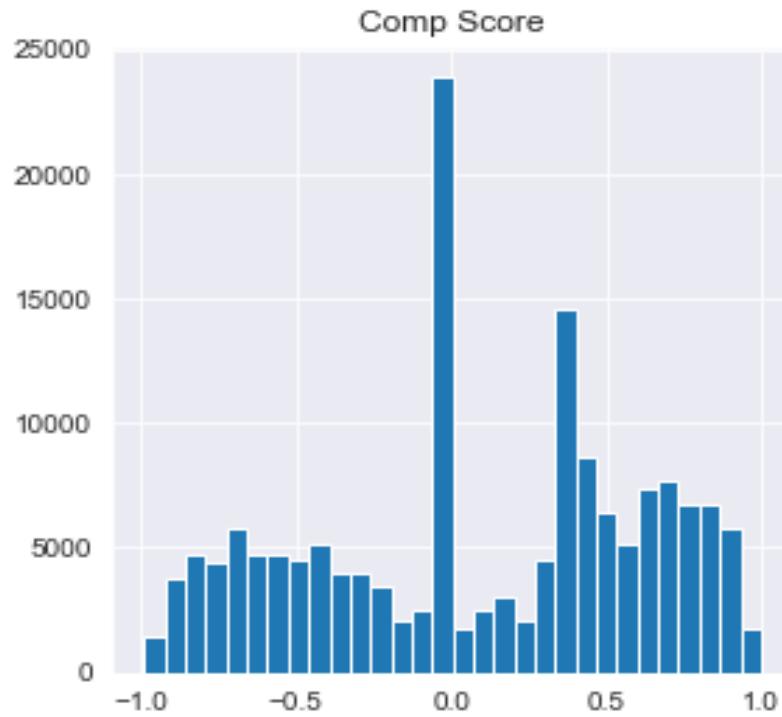
# EXPLORATORY DATA ANALYSIS

---

- The Exploratory data analysis (“EDA”) was undertaken on both the News Articles & Tweets
- They were done separately
- The News will be reviewed first, followed by the Tweets
- Moving over to the Tweets
  - To reiterate, **the Tweets are expected to be unanimously against the Law**

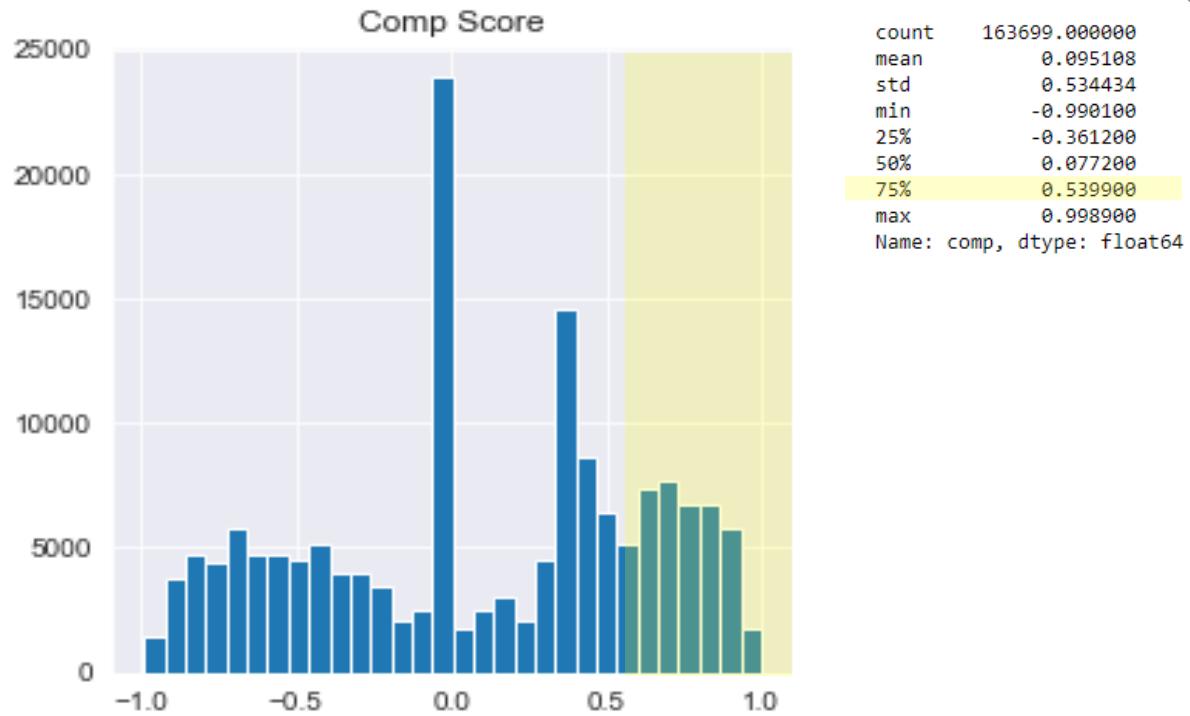
# TWITTER EDA

Twitter **Tweets** present a **bimodal** nature on the Comp Score



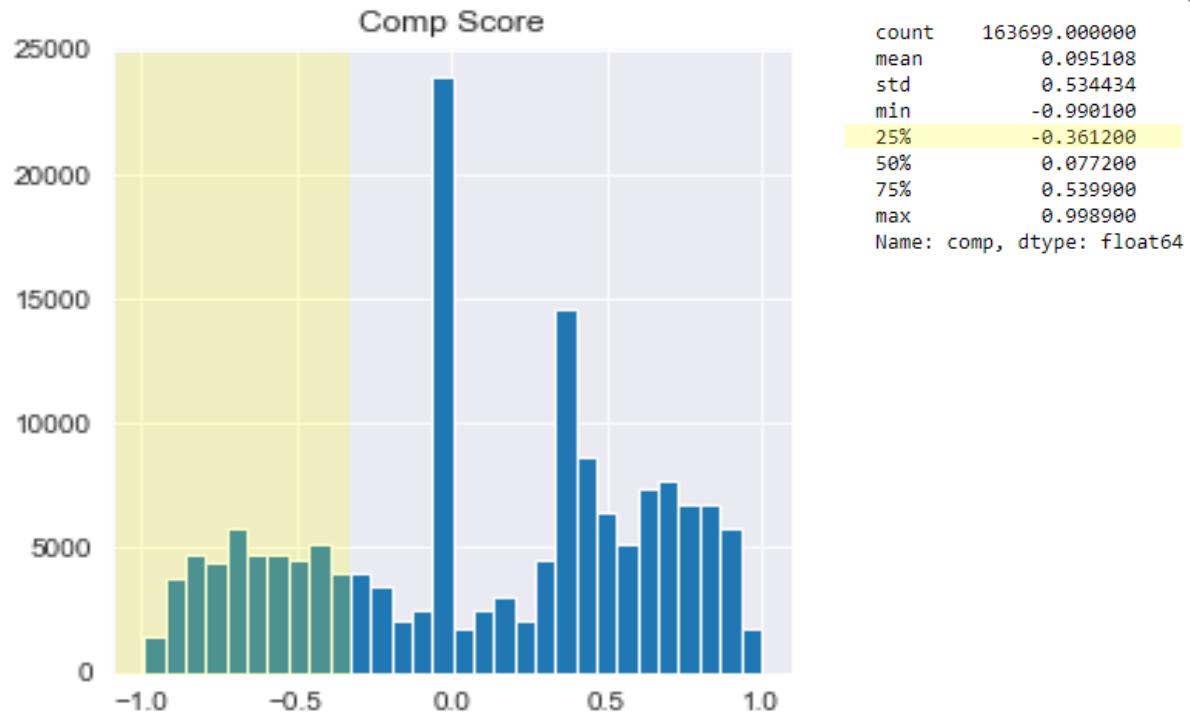
# TWITTER PREP

Using Descriptive Statistics, we determined that **Tweets with a Comp Score at or above 0.539 to be Positive**



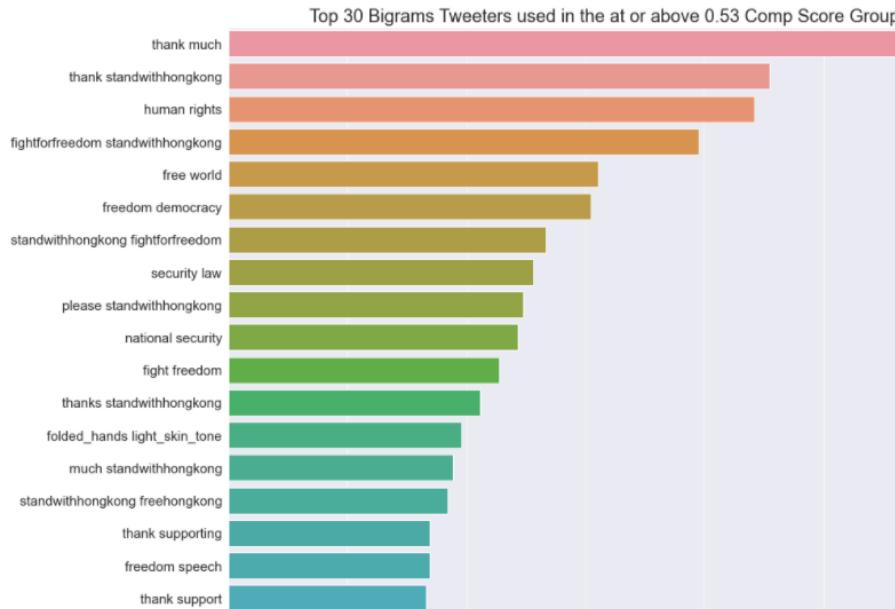
# TWITTER EDA

Using Descriptive Statistics, we determined that Tweets with a Comp Score at or above 0.539 to be Positive & Tweets **at or below -0.361 to be Negative**



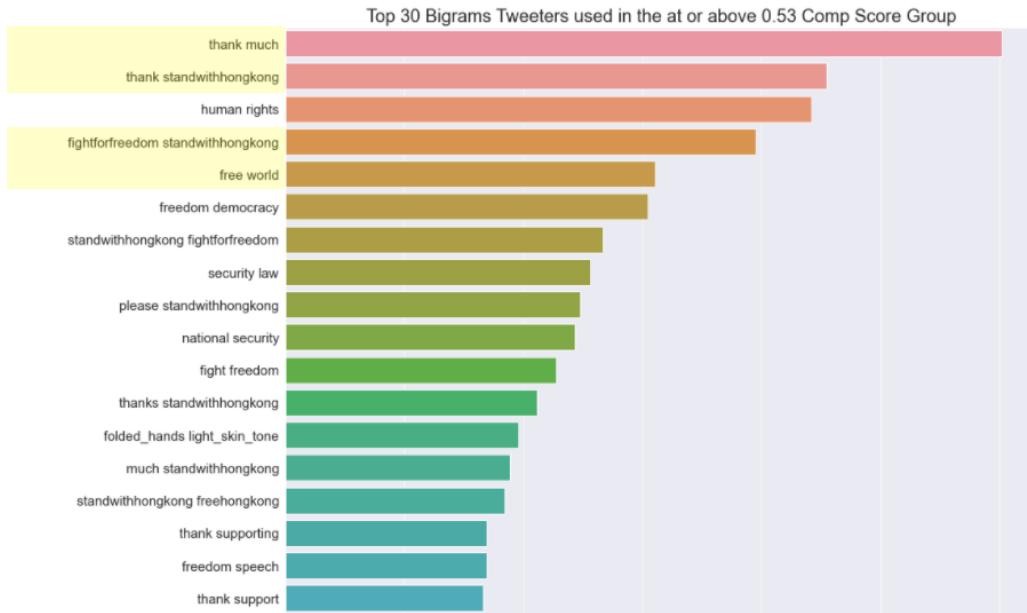
# TWITTER EDA

This is also seen in the top bigrams  
therein with thanks



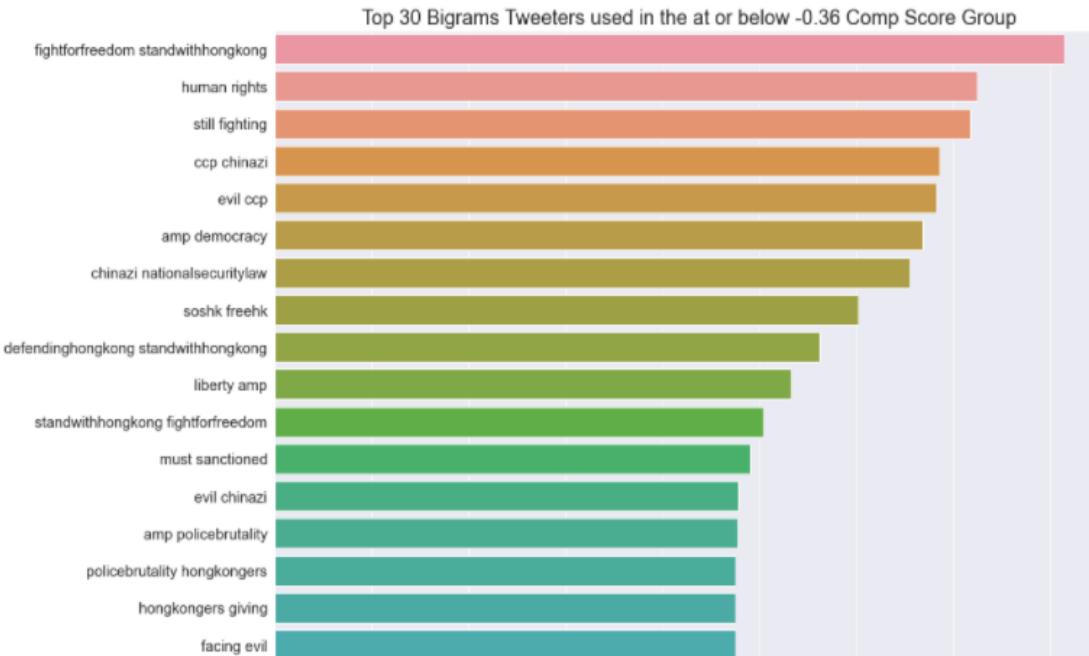
# TWITTER EDA

This is also seen in the top bigrams  
therein with thanks



# TWITTER EDA

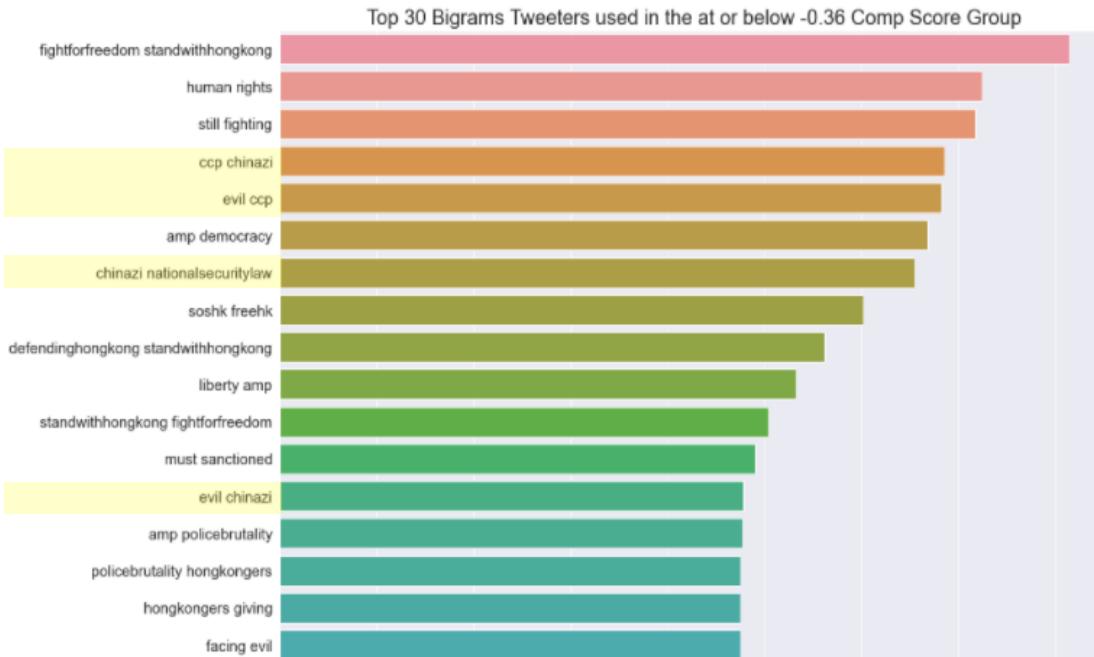
This may suggest that the optimism is low towards these positive topics



# TWITTER EDA

This may suggest that the optimism is low towards these positive topics

Especially given **negativity towards China**



# TWITTER EDA TAKE-AWAYS

---

## NEGATIVE | SUMMARY

- Evil CCP
- References to Freedom / Democracy
- Holding Hands (



## POSITIVE | SUMMARY

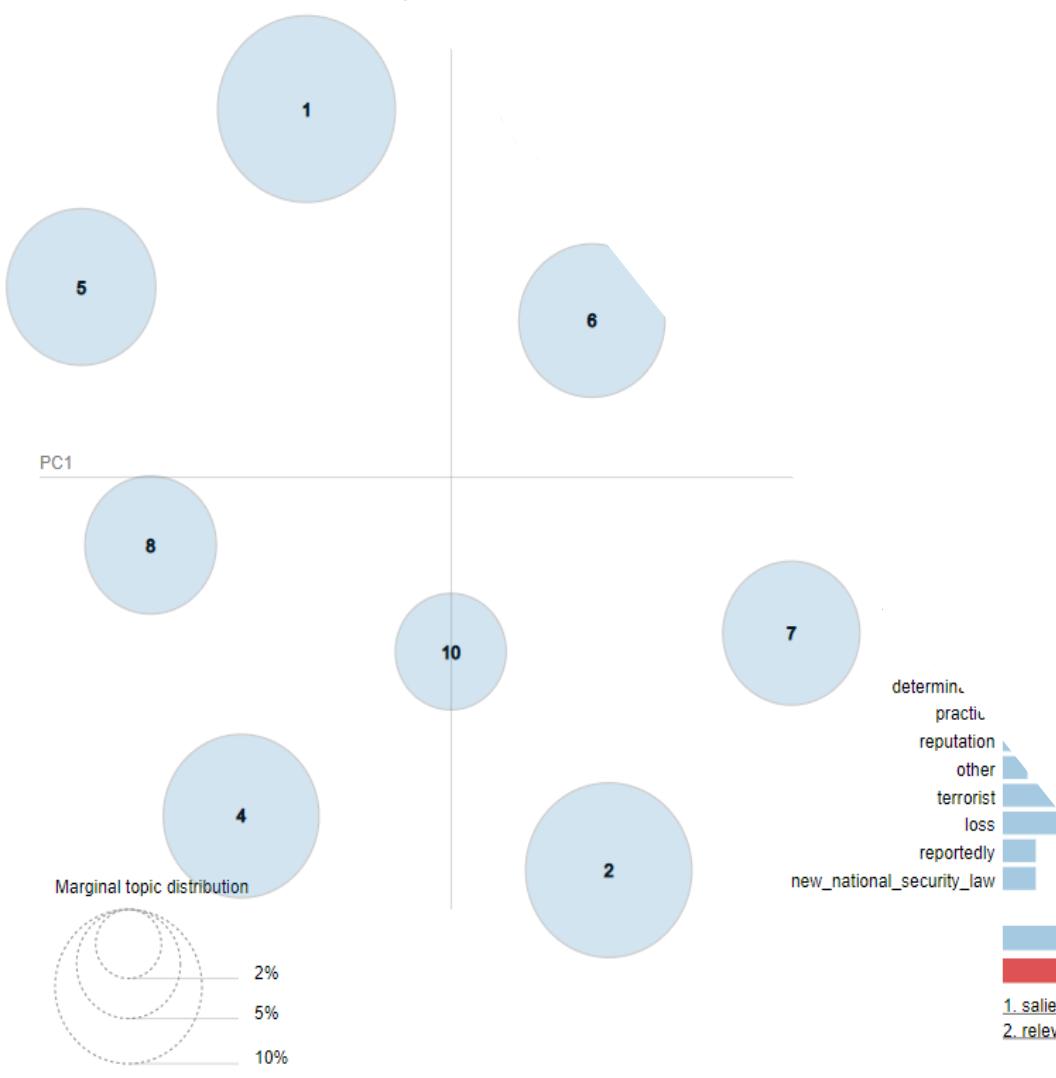
- Thanks much / support / free world
- Please standwithhongkong / support
- folded\_hands (



# 04

## MODEL DESCRIPTION

---



# MODEL APPROACH

---

- The modelling approach for the News (📰) & Twitter (🐦) is identical

# MODEL APPROACH

---

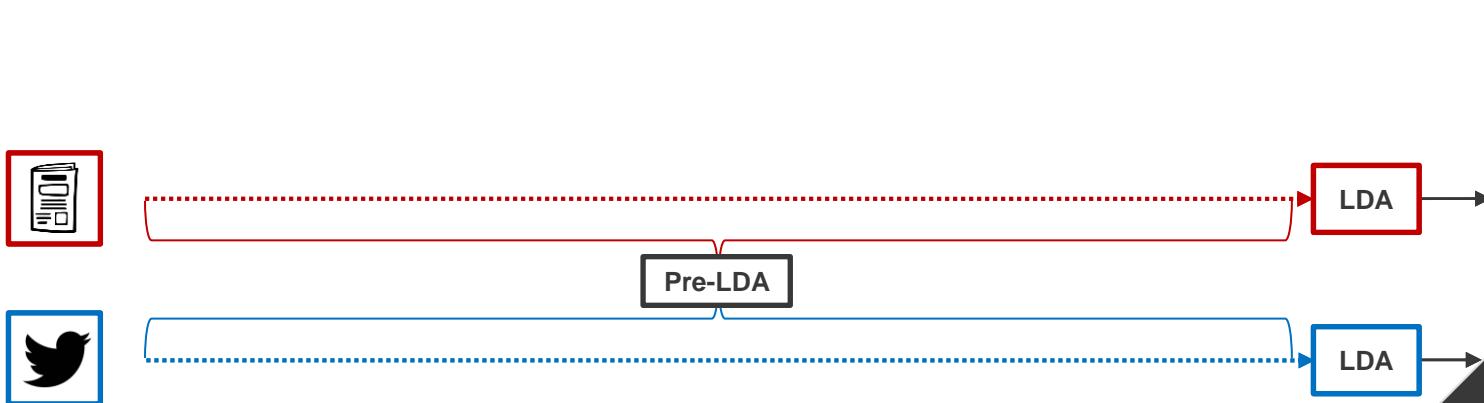
- The modelling approach for Twitter & News is identical
  - both use **Latent Dirichlet Allocation (“LDA”)** to classify text



# MODEL APPROACH

---

- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do **another “cleaning” process** which we call **Pre-LDA**



# MODEL APPROACH

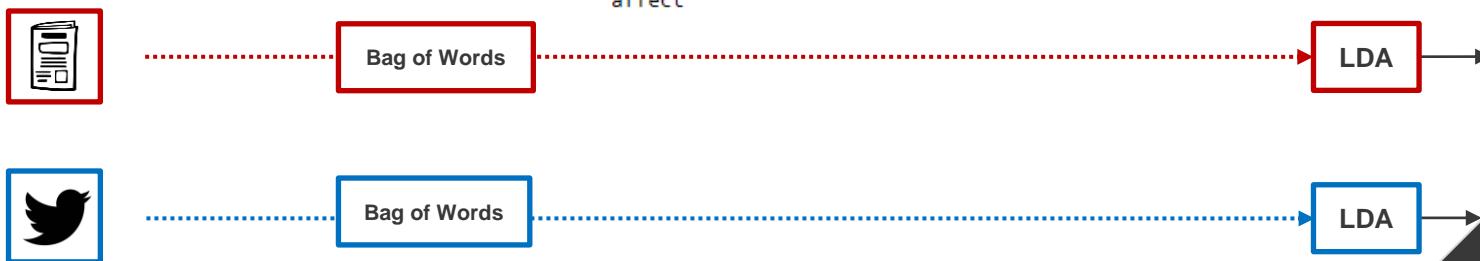
---

- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do another “cleaning” process which we call Pre-LDA
  - **Bag of Words** | A representation of text that **describes the occurrence of words**; it disregards the order or structure, the occurrence is what matters

```
print (corpus[4][0:2])  
[(1, 9), (11, 5)]
```

```
# Below is the word  
# assigned to the number  
# in the bag of words  
word = id2word[[4][:][0]]  
print(word)
```

affect



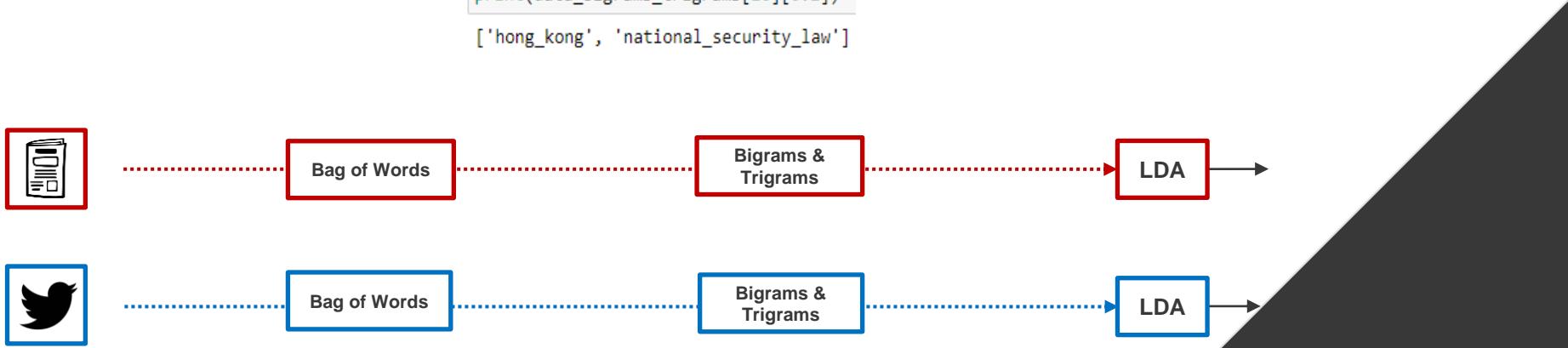
# MODEL APPROACH

---

- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do another “cleaning” process which we call Pre-LDA
  - Bag of Words | A representation of text that describes the occurrence of words; it disregards the order or structure, the occurrence is what matters
  - **Bigrams & Trigrams** | sequences of two (2) or three (3) words respectively; it's efficient to identify occurrences of groups of words appearing together

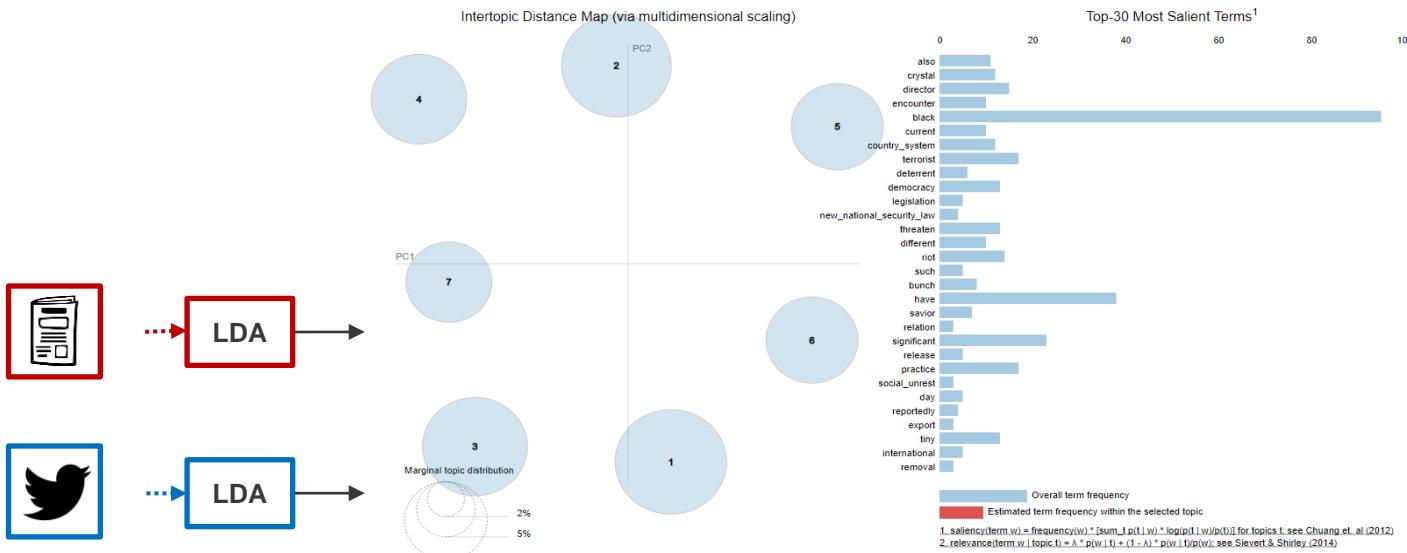
```
# below is a bigram & trigram
print(data_bigrams_trigrams[16][0:2])

['hong_kong', 'national_security_law']
```



# MODEL APPROACH

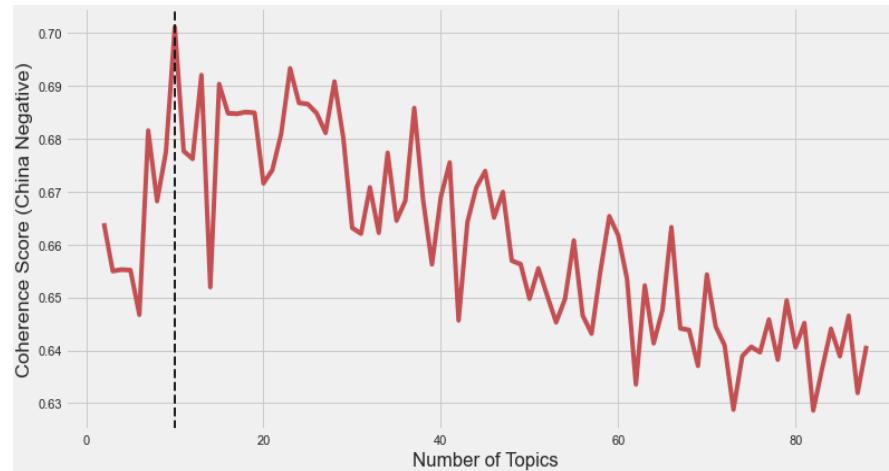
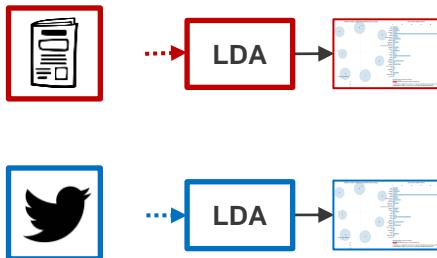
- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do another “cleaning” process which we call Pre-LDA
- Then sent to LDA & subsequently **the pyLDAvis** which is a topic model visualization tool. An initial glimpse of how many topics there are & which words are assigned to them is achieved here



# MODEL APPROACH

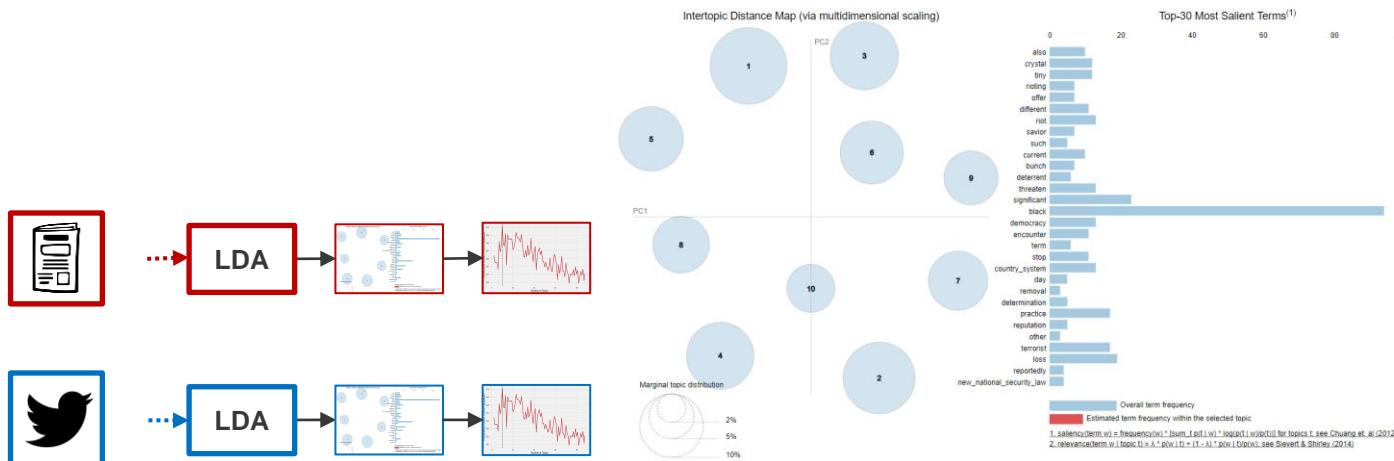
---

- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do another “cleaning” process which we call Pre-LDA
- Then sent to LDA & subsequently the pyLDAVis which is a topic model visualization tool. An initial glimpse of how many topics there are & which words are assigned to them is achieved here.
- **Determining the optimal number of topics, a Coherence Score is calculated**  
measuring the degree of semantic similarity; In other words, **a set of statements or facts that support each other are said to be coherent**



# MODEL APPROACH

- The modelling approach for Twitter & News is identical
- Before they get to LDA, they do another “cleaning” process which we call Pre-LDA
- Then sent to LDA & subsequently the pyLDAvis which is a topic model visualization tool. An initial glimpse of how many topics there are & which words are assigned to them is achieved here.
- Determining the optimal number of topics, a Coherence Score is calculated measuring the degree of semantic similarity; In other words, a set of statements or facts that support each other are said to be coherent.
- **With the Coherence Score, the LDA & pyLDAvis process is done again with the optimal number & an investigation of each those topics is reviewed**



# 05

# MODEL

# FINDINGS

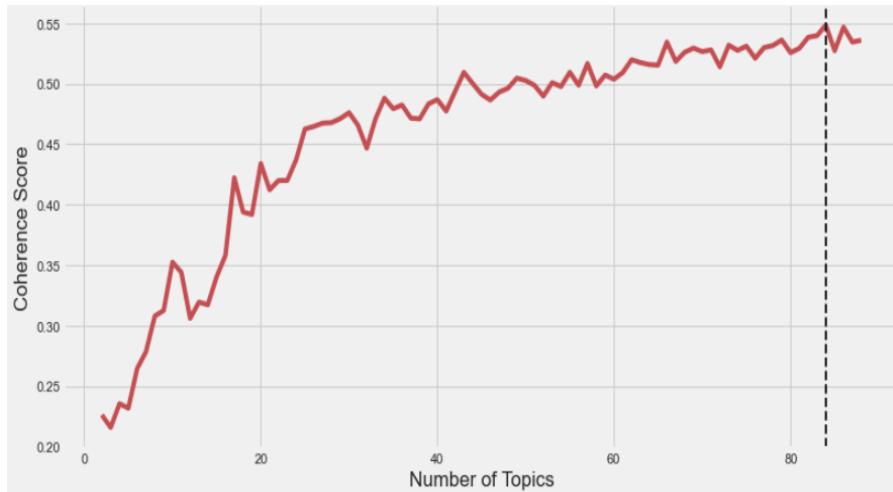
---



# NEWS | INITIAL FINDINGS

---

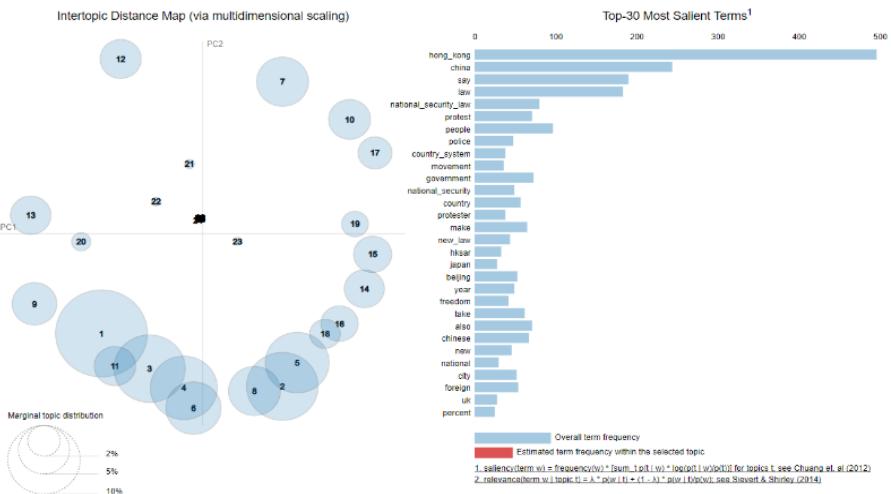
- The 30 News Articles presented 84 Topics as the ideal breakdown



# NEWS | INITIAL FINDINGS

---

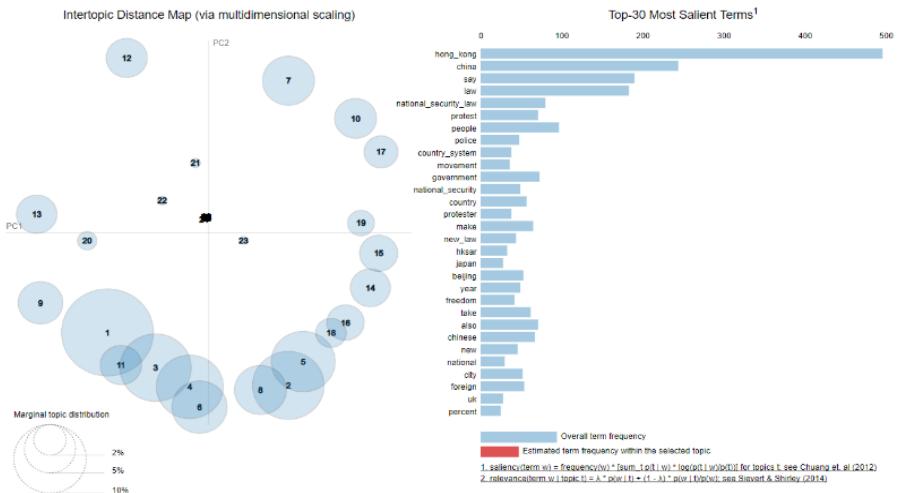
- The 30 News Articles presented 84 Topics as the ideal breakdown
  - The **sizes vary** in size
  - **Distance between the center** for many indicates some are distinct while **21+ hover in the center**
  - A **notable amount of overlap** is presented



# NEWS | INITIAL FINDINGS

---

- The 30 News Articles presented 84 Topics as the ideal breakdown
- Removing the 21 topics given their proximity to the center & size is an option, **another option was decided to better handle the LDA Modeling process on the News Articles**

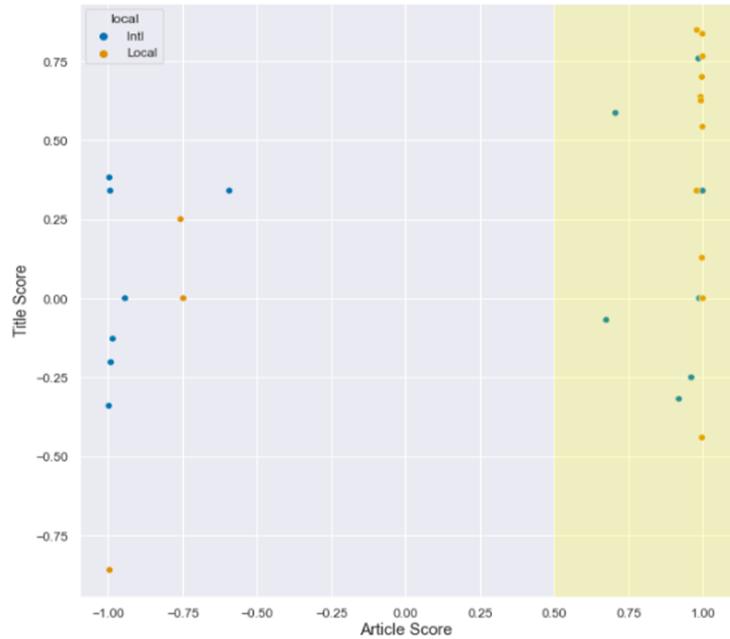


# NEWS | INITIAL FINDINGS

---

- The 30 News Articles presented 84 Topics as the ideal breakdown  
Removing the 21 topics given their proximity to the center & size is an option, another option was decided to better handle the LDA Modeling process on the News Articles
- Heading back to EDA, we have **four (4) categories:**
  - China & International **Positive** x2

Comparing the Comp Scores between Articles & Titles

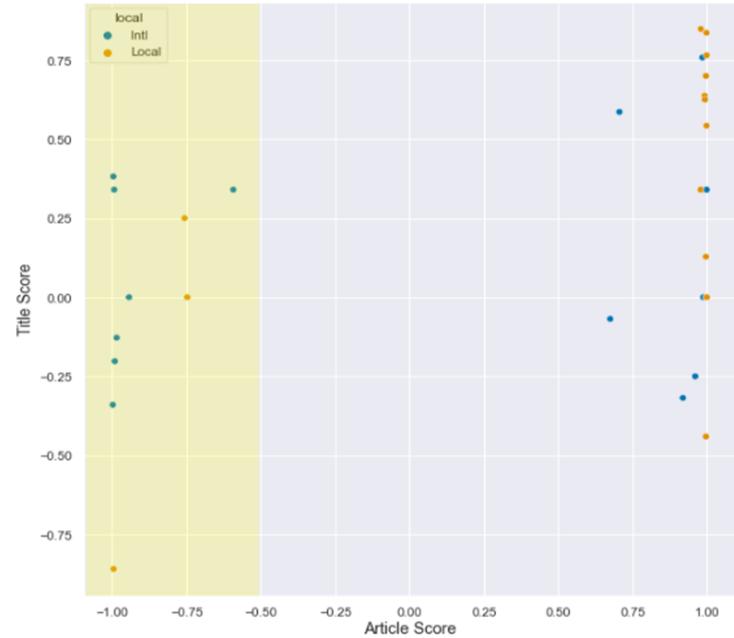


# NEWS | INITIAL FINDINGS

---

- The 30 News Articles presented 84 Topics as the ideal breakdown  
Removing the 21 topics given their proximity to the center & size is an option, another option was decided to better handle the LDA Modeling process on the News Articles
- Heading back to EDA, we have **four (4) categories:**
  - China & International Positive x2
  - China & International **Negative** x2

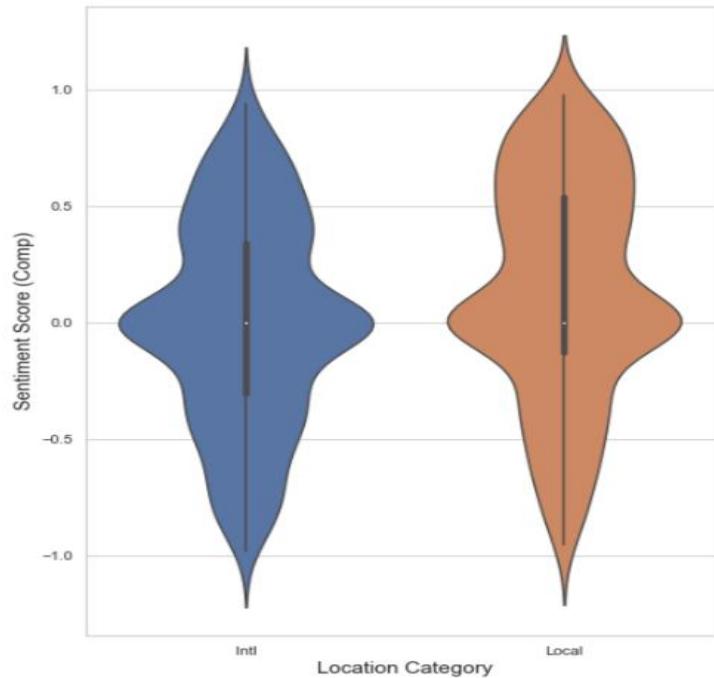
Comparing the Comp Scores between Articles & Titles



# NEWS | INITIAL FINDINGS

---

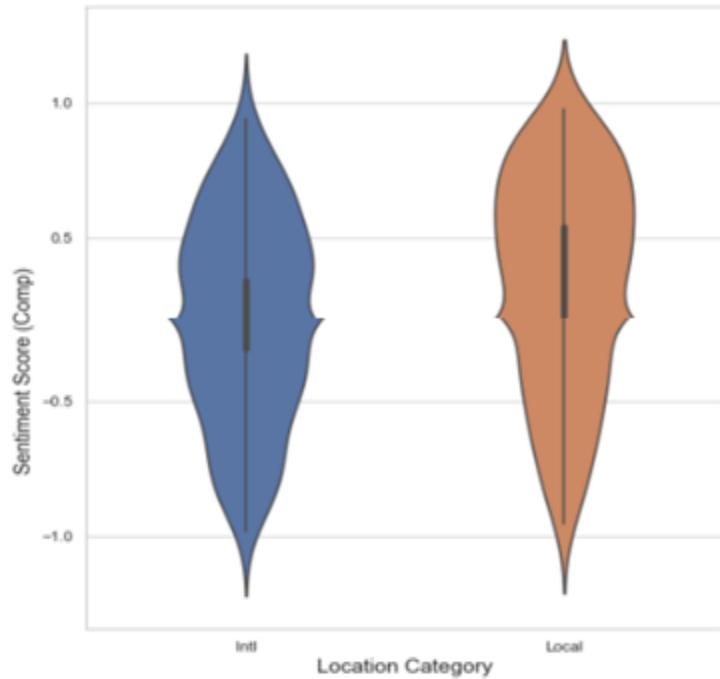
- The 30 News Articles presented 84 Topics as the ideal breakdown  
Removing the 21 topics given their proximity to the center & size is an option, another option was decided to better handle the LDA Modeling process on the News Articles
- Heading back to EDA, we have four (4) categories:
  - China & International Positive x2
  - China & International Negative x2
- The Articles were also split by their sentences & re-examined with a notable amount of Zero's; meaning Neutral



# NEWS | INITIAL FINDINGS

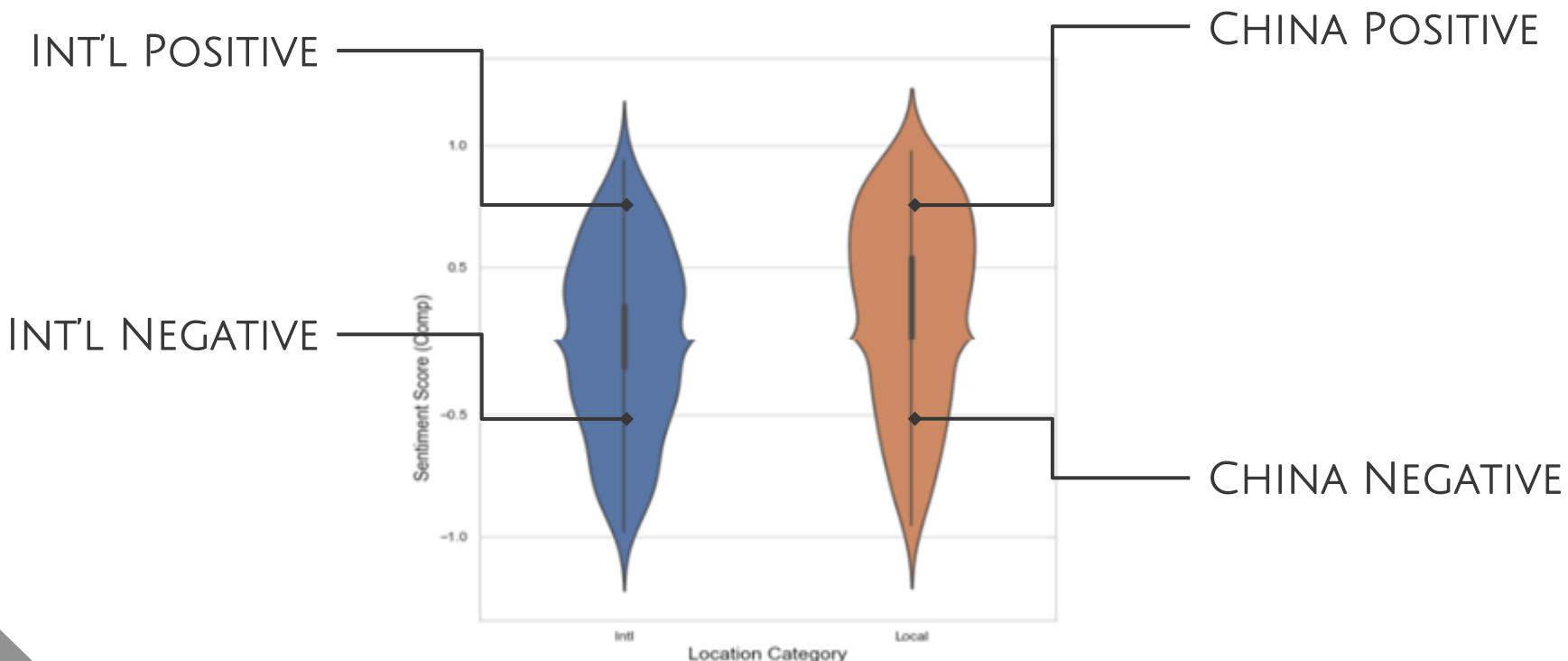
---

- The 30 News Articles presented 84 Topics as the ideal breakdown  
Removing the 21 topics given their proximity to the center & size is an option, another option was decided to better handle the LDA Modeling process on the News Articles
- Heading back to EDA, we have four (4) categories:
  - China & International Positive x2
  - China & International Negative x2
- The Articles were also split by their sentences & re-examined with a notable amount of Zero's; meaning Neutral
  - Neutrals were removed



# NEWS | INITIAL FINDINGS

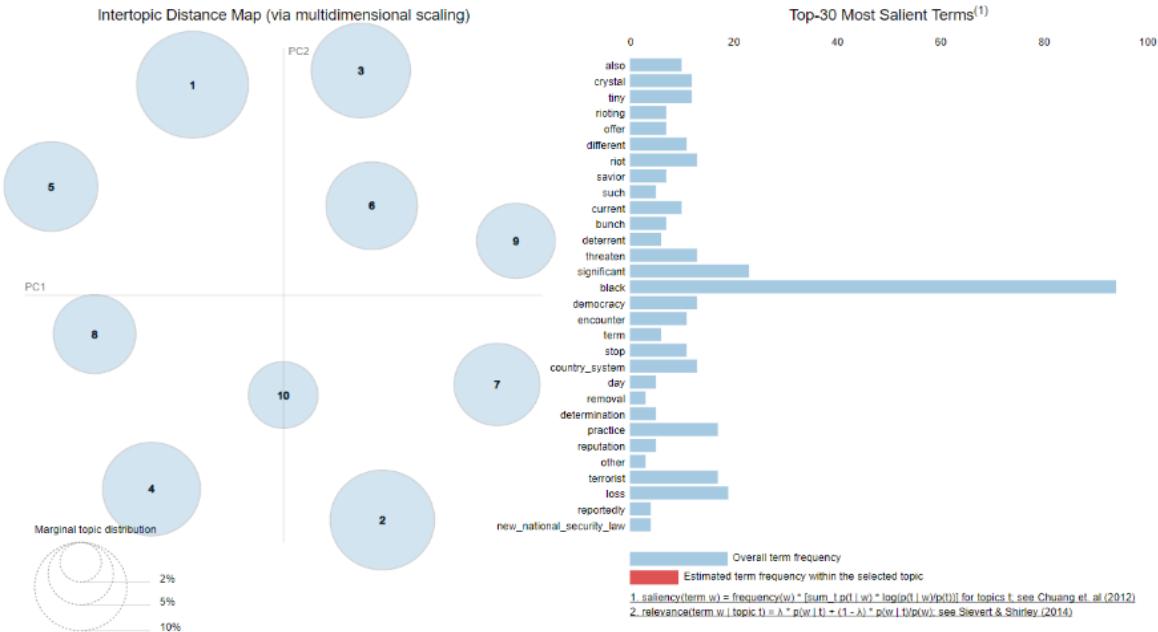
---



# NEWS | CHINA NEGATIVE

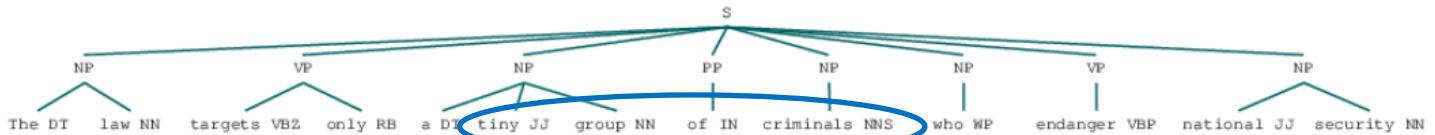
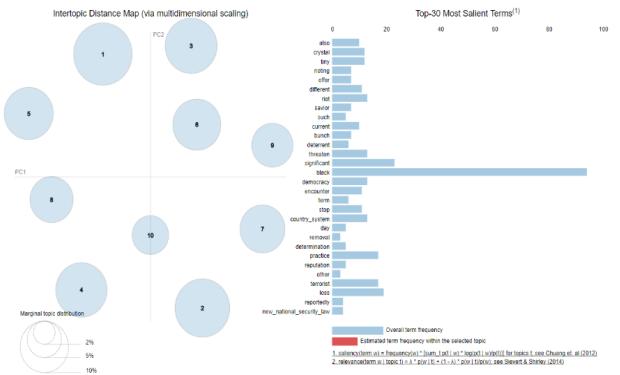
---

- The Coherence Score had **10 Topics**



# NEWS | CHINA NEGATIVE

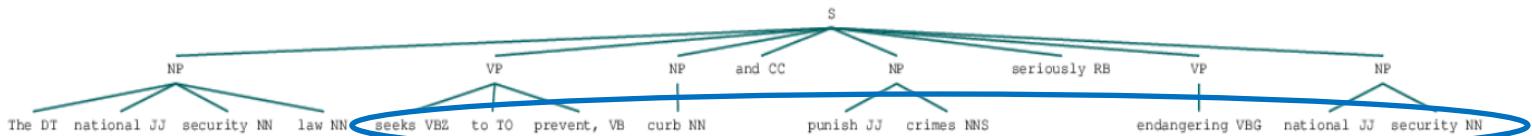
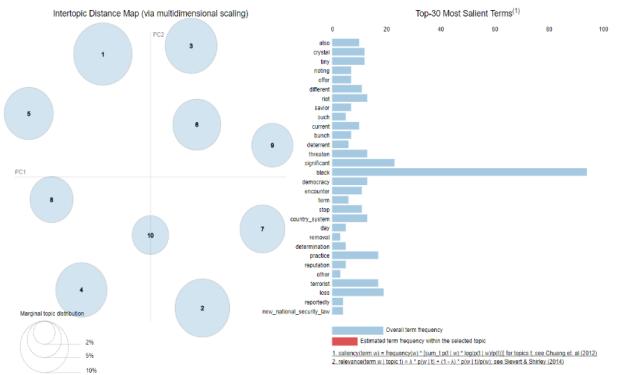
- The Coherence Score had 10 Topics
- A summary\* is highlighted below:
  - The NSL intents is to target a **small group** of criminals



\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA NEGATIVE

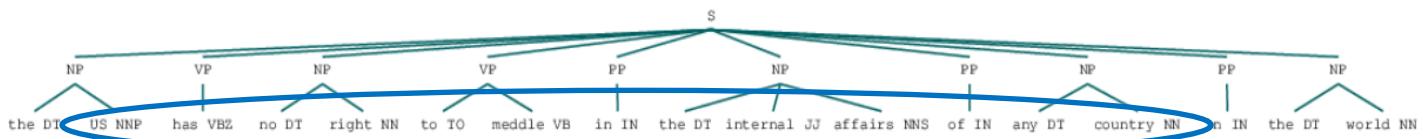
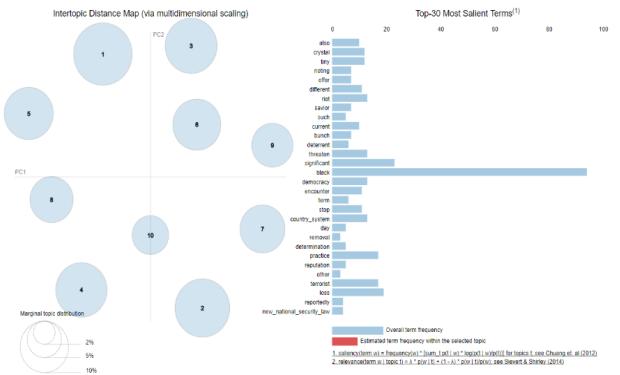
- The Coherence Score had 10 Topics
- A summary\* is highlighted below:
  - The NSL intents is to target a small group of criminals
  - **NSL's intent is to heal HK from previous violence**



\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA NEGATIVE

- The Coherence Score had 10 Topics
- A summary\* is highlighted below:
  - The NSL intents is to target a small group of criminals
  - NSL's intent is to heal HK from previous violence
  - **Discontent of the US for meddling in their affairs**

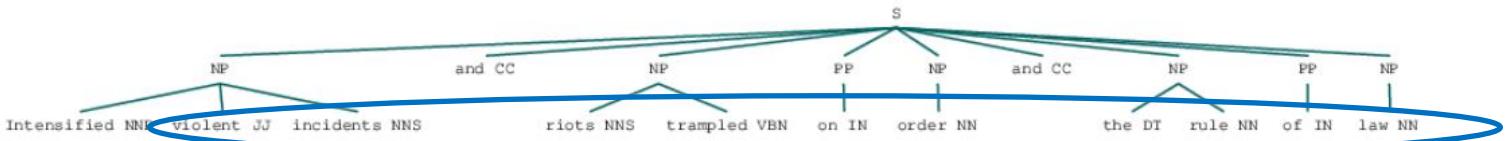
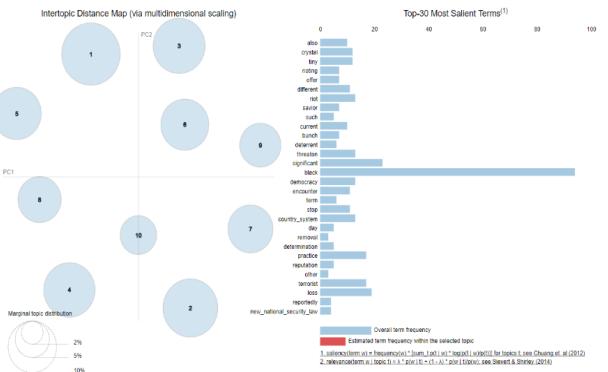


\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA NEGATIVE

---

- The Coherence Score had 10 Topics
- A summary\* is highlighted below:
  - The NSL intents is to target a small group of criminals
  - NSL's intent is to heal HK from previous violence
  - Discontent of the US for meddling in their affairs
  - Protestors **criminal activity**

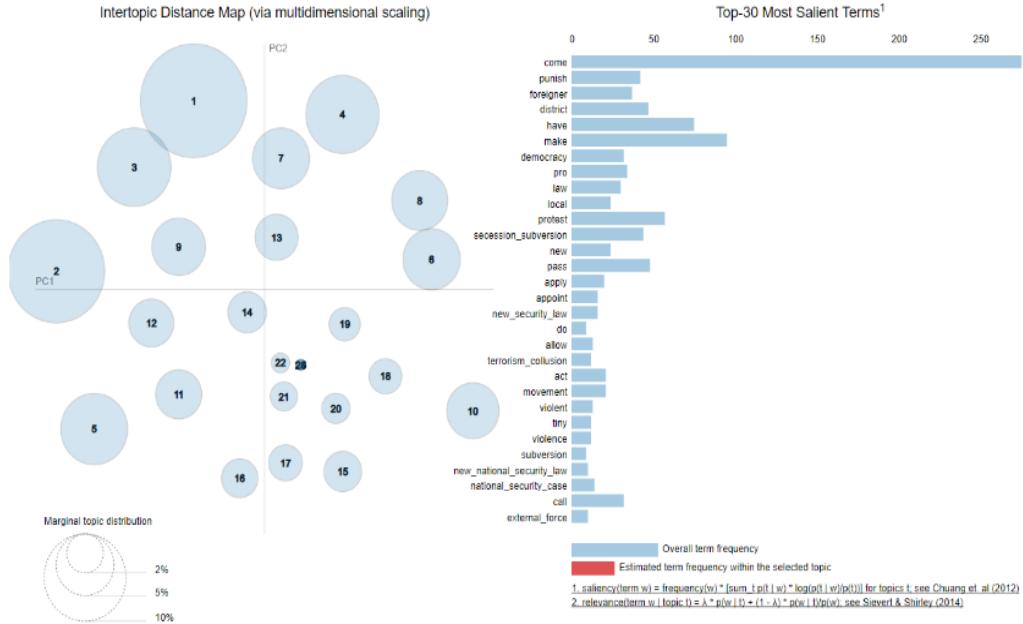


\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L NEGATIVE

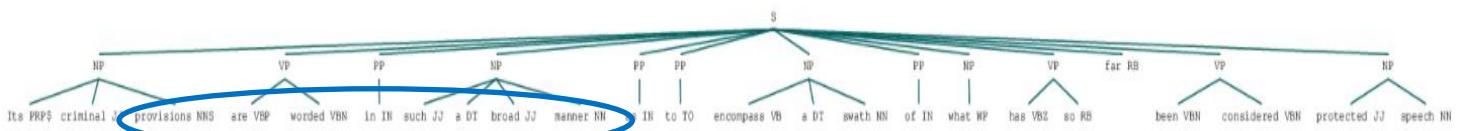
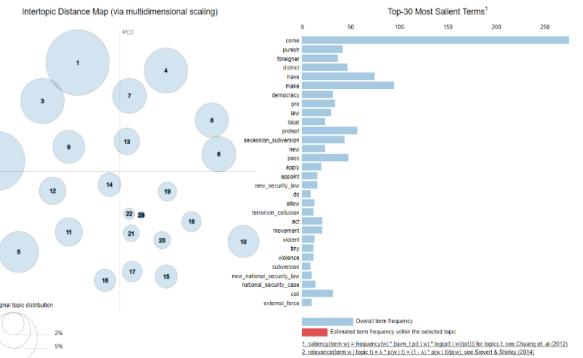
---

- The Coherence Score had 47 Topics
  - **10 that had 7+ observations shortlisted**



# NEWS | INT'L NEGATIVE

- The Coherence Score had 47 Topics
  - 10 that had 7+ observations shortlisted
- The summary is highlighted below:
  - The **NSL is vaguely worded**

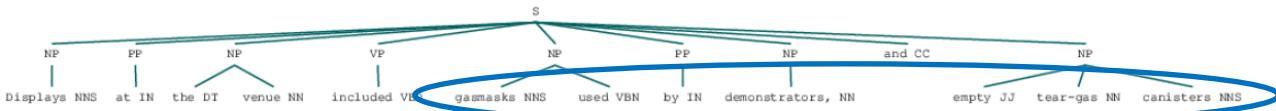
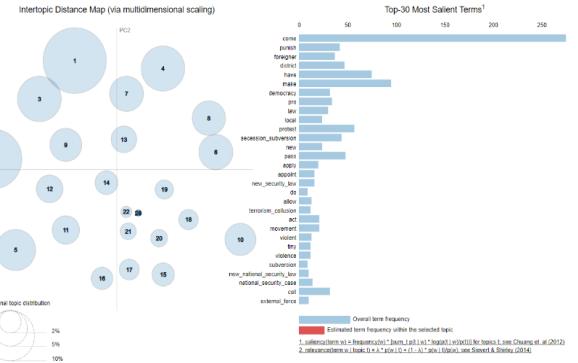


\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L NEGATIVE

---

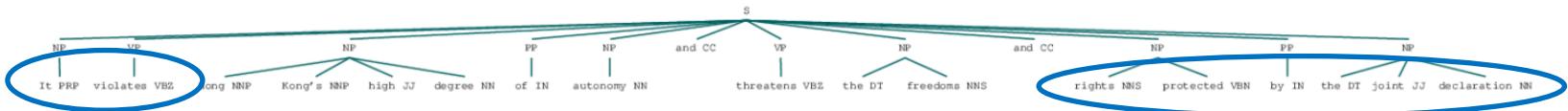
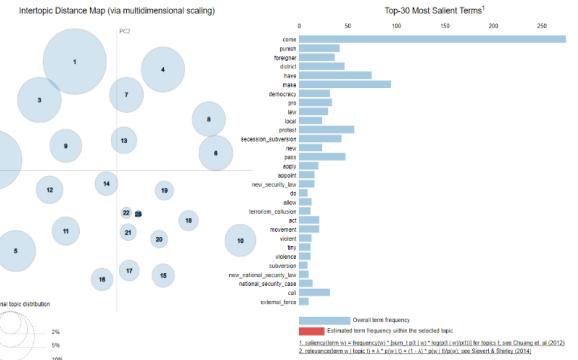
- The Coherence Score had 47 Topics
  - 10 that had 7+ observations shortlisted
- The summary is highlighted below:
  - The NSL is vaguely worded
  - **Tactics of the protesters**



\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L NEGATIVE

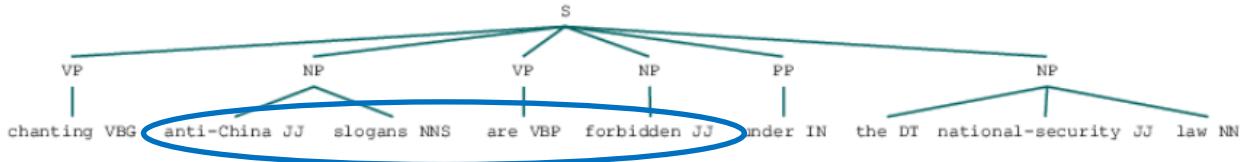
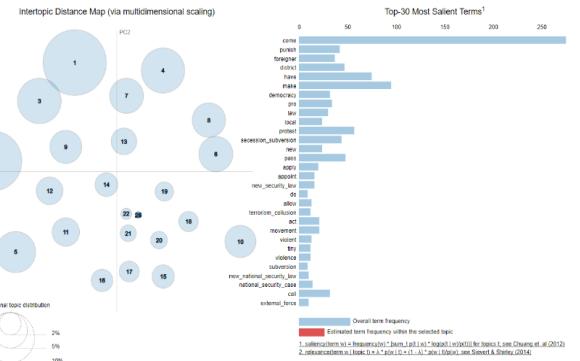
- The Coherence Score had 47 Topics
  - 10 that had 7+ observations shortlisted
- The summary is highlighted below:
  - The NSL is vaguely worded
  - Tactics of the protesters
  - The **NSL undermines the OcTs system**



\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L NEGATIVE

- The Coherence Score had 47 Topics
  - 10 that had 7+ observations shortlisted
- The summary is highlighted below:
  - The NSL is vaguely worded
  - Tactics of the protestors
  - The NSL undermines the OcTs system
  - Possible **suppression of free speech**

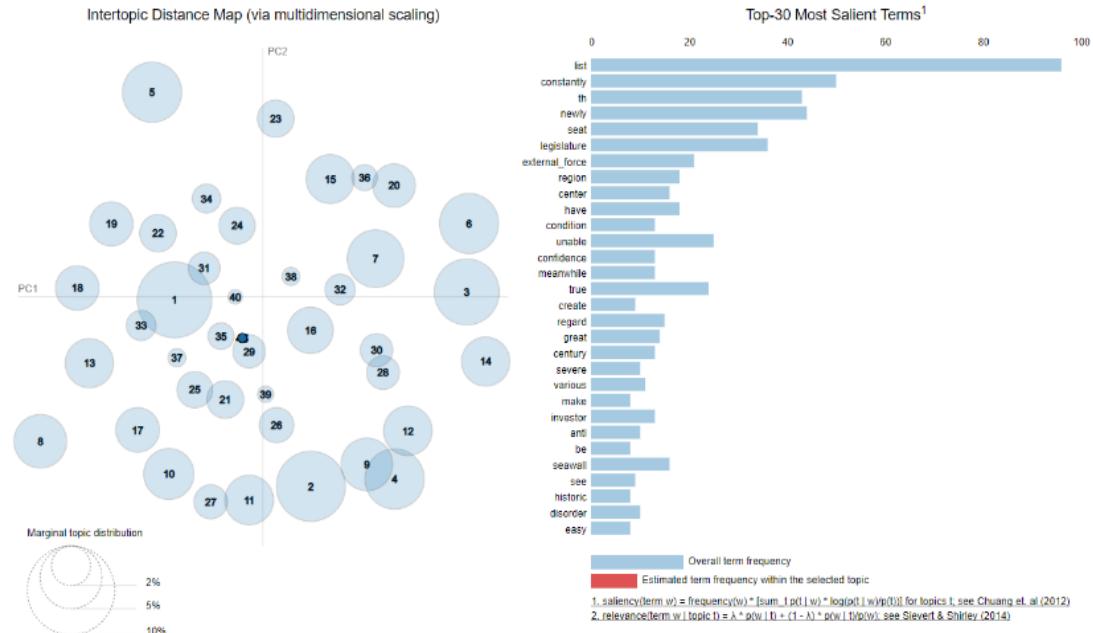


\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA POSITIVE

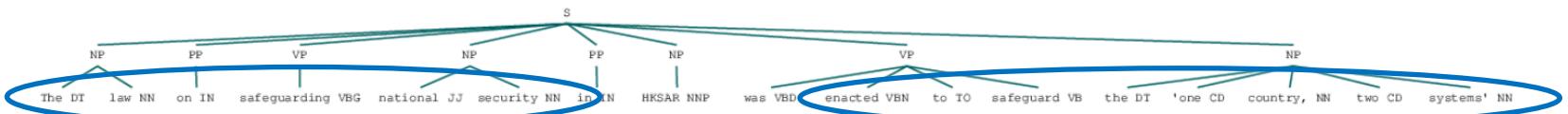
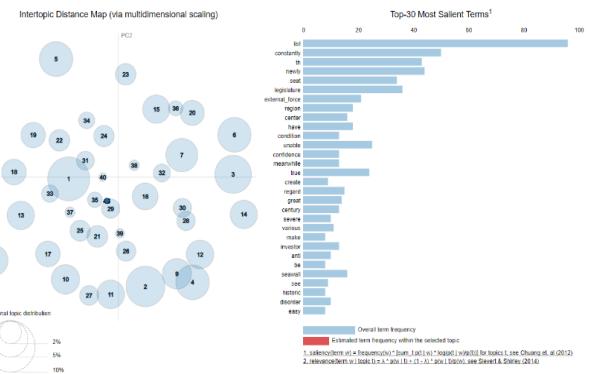
---

- The Coherence Score had **48 Topics**
  - Top 25% contributors were **shortlisted: 12**



# NEWS | CHINA POSITIVE

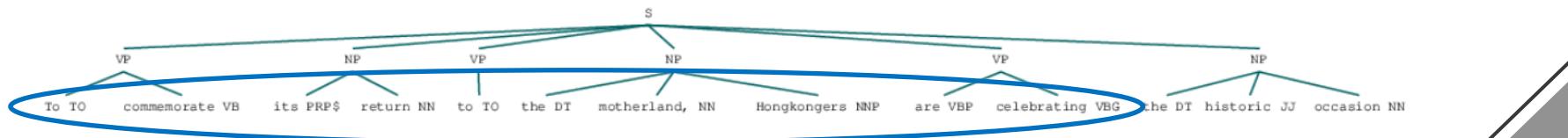
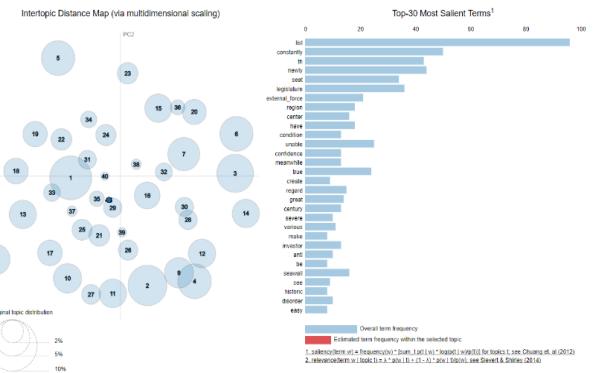
- The Coherence Score had 48 Topics
  - Top 25% contributors were shortlisted: 12
- The summary is highlighted below:
  - The NSL is about **safeguarding national security**



\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA POSITIVE

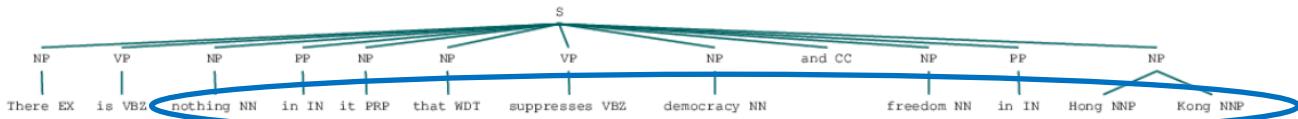
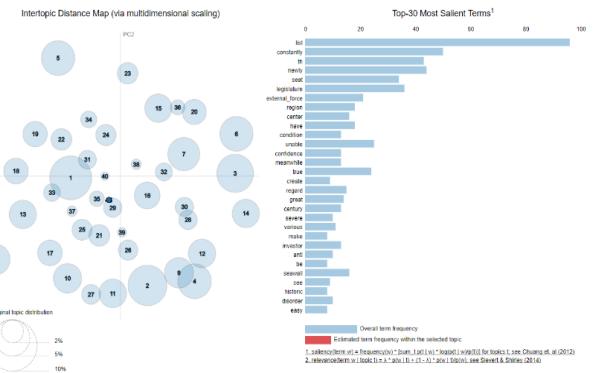
- The Coherence Score had 48 Topics
  - Top 25% contributors were shortlisted: 12
- The summary is highlighted below:
  - The NSL is about safeguarding national security
  - **Rejoiceful mood** on HK's return to China



\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA POSITIVE

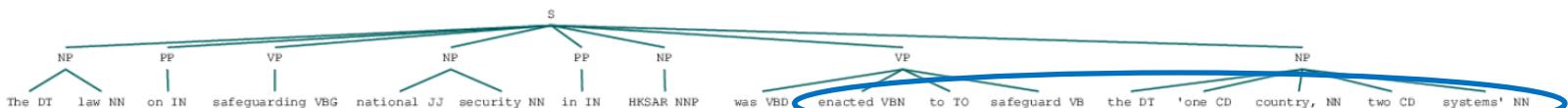
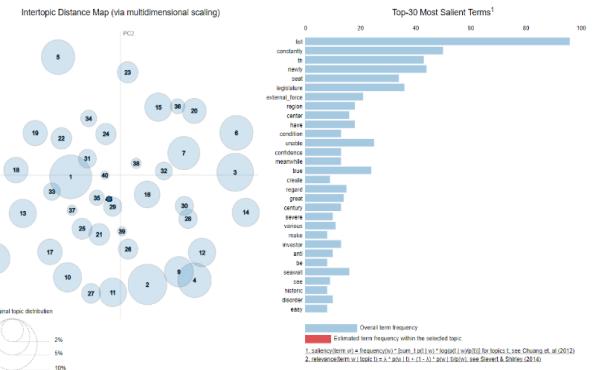
- The Coherence Score had 48 Topics
  - Top 25% contributors were shortlisted: 12
- The summary is highlighted below:
  - The NSL is about safeguarding national security
  - Rejoiceful mood on HK's return to China
  - The **NSL has nothing to do with freedom of speech**



\* A larger subset of examples in both the Report & Source Code

# NEWS | CHINA POSITIVE

- The Coherence Score had 48 Topics
  - Top 25% contributors were shortlisted: 12
- The summary is highlighted below:
  - The NSL is about safeguarding national security
  - Rejoiceful mood on HK's return to China
  - The NSL has nothing to do with freedom of speech
  - The **NSL will not infringe on OcTs**

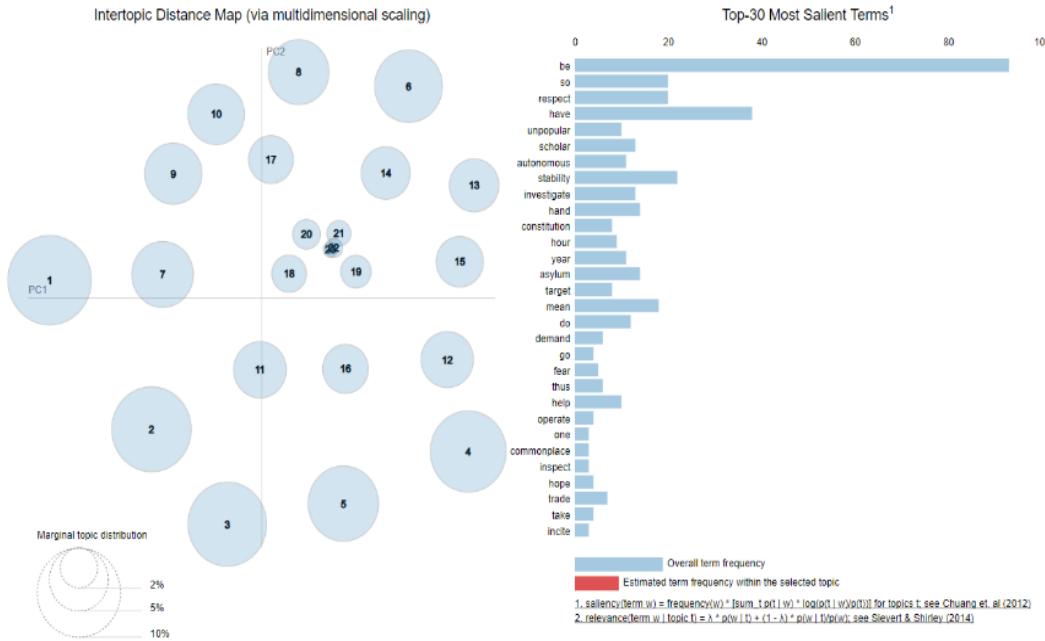


\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L POSITIVE

---

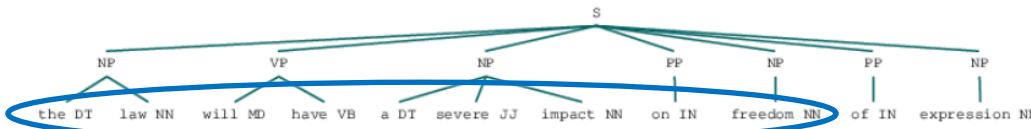
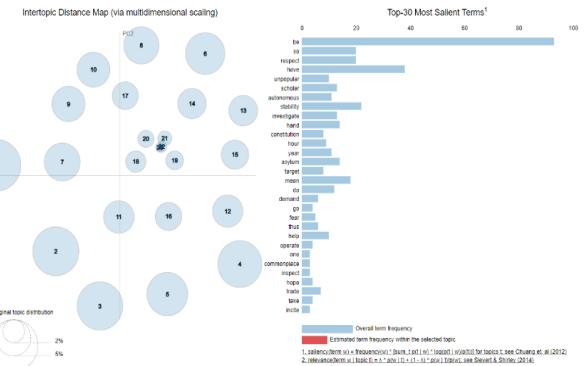
- The Coherence Score had **23 Topics**
  - **17** that had **10+** observations shortlisted



# NEWS | INT'L POSITIVE

---

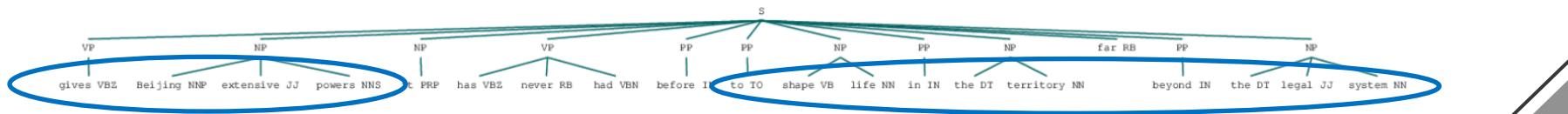
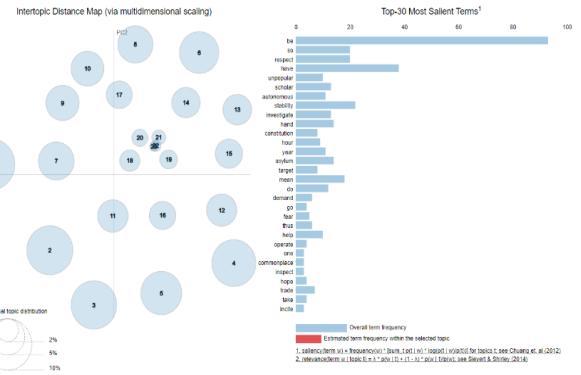
- The Coherence Score had 23 Topics
  - 17 that had 10+ observations shortlisted
- The summary is highlighted below:
  - **Usage of positive words:** seen a notable amount of the time in Int'l Positive albeit they don't come across with Positive intent



# NEWS | INT'L POSITIVE

---

- The Coherence Score had 23 Topics
  - 17 that had 10+ observations shortlisted
- The summary is highlighted below:
  - **Usage of positive words:** seen a notable amount of the time in Int'l Positive albeit they don't come across with Positive intent

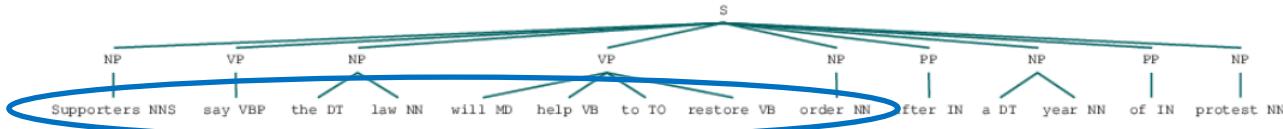
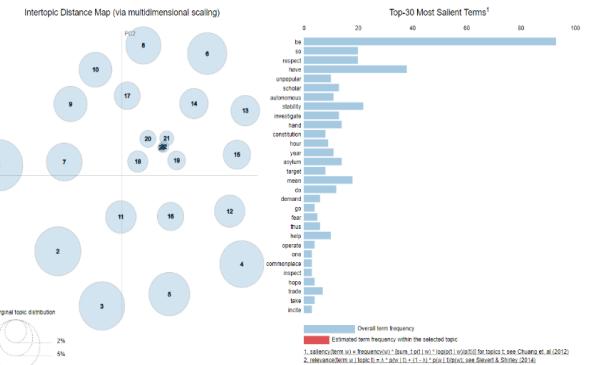


\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L POSITIVE

---

- The Coherence Score had 23 Topics
  - 17 that had 10+ observations shortlisted
- The summary is highlighted below:
  - Usage of positive words
  - **Positive mentions towards supporters**

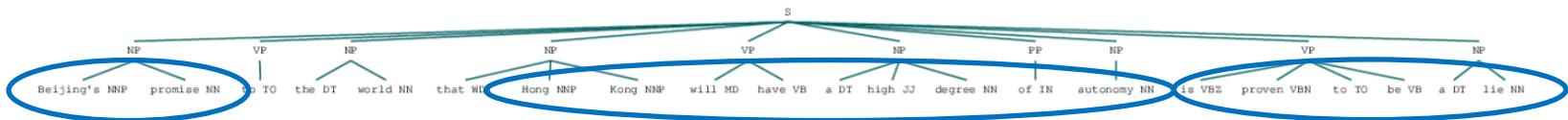
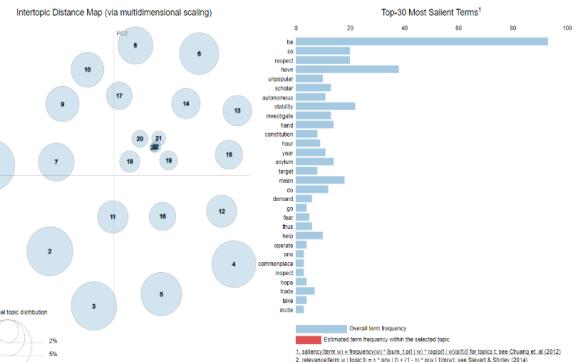


\* A larger subset of examples in both the Report & Source Code

# NEWS | INT'L POSITIVE

---

- The Coherence Score had 23 Topics
  - 17 that had 10+ observations shortlisted
- The summary is highlighted below:
  - Usage of positive words
  - Positive mentions towards supporters
  - **Mentions of Beijing's position** of safeguarding security



\* A larger subset of examples in both the Report & Source Code



## CHINA | POSITIVE

- The NSL is to **safeguard** the majority of HK
- **HK people celebrate** the return to China
- The NSL will **not infringe on OcTs**



## INT'L | POSITIVE

- **Great deal of positive words** without a possible positive intent
- Mentions of **NSL supporters in HK**
- Mentions of **Beijing's position**

---

## NEWS MODELING TAKE-AWAYS

---



## CHINA | NEGATIVE

- The **violence of protestors**
- **Interference in China** & HK's internal **affairs**
- China's belief that the **NSL will lead to long term stability & prosperity** in HK



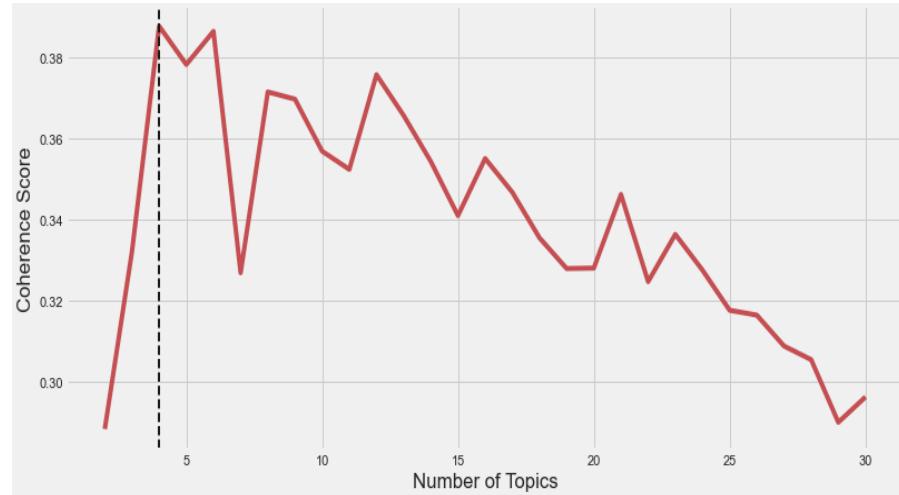
## INT'L | NEGATIVE

- NSL's **infringement on Common Law**
- Credence towards the **vagueness of the language** used in the NSL
- **Actions & tactics of protestors**

# TWITTER | INITIAL FINDINGS

---

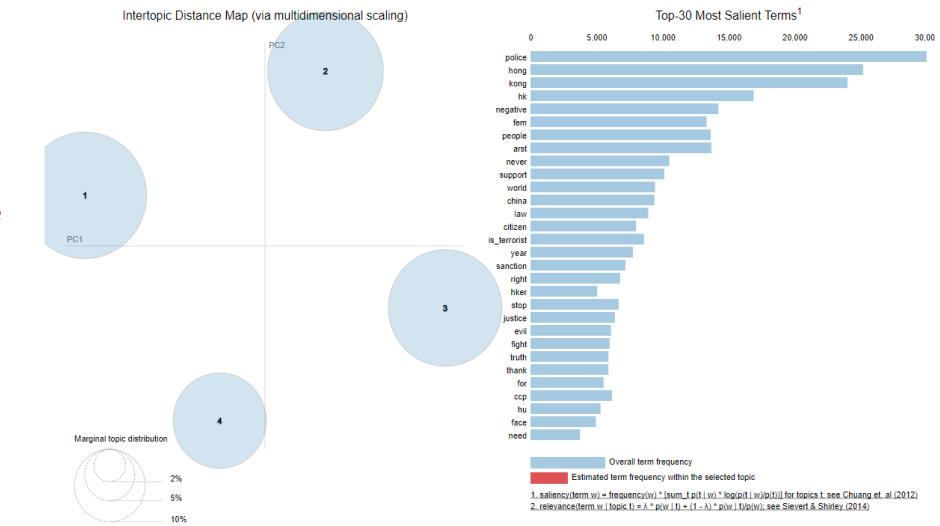
- The 163,699 Tweets presented 4 Topics as the ideal breakdown



# TWITTER | INITIAL FINDINGS

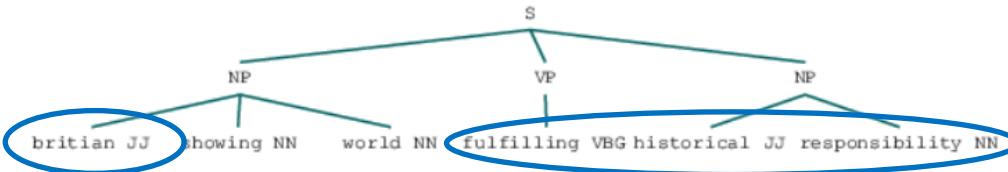
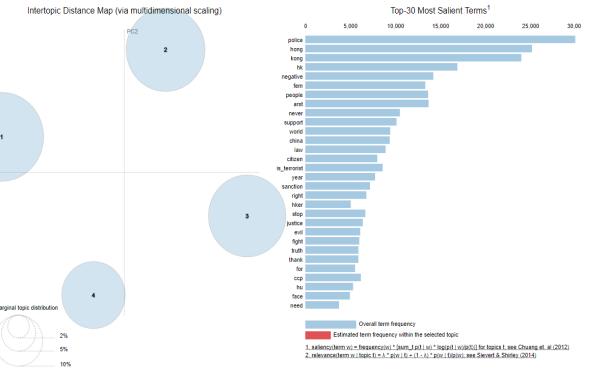
---

- The 163,699 Tweets presented 4 Topics as the ideal breakdown
  - The sizes **appear similar in size**
  - Distance to the center suggest unique**
  - No overlap** observed



# TWITTER | TOPIC ONE

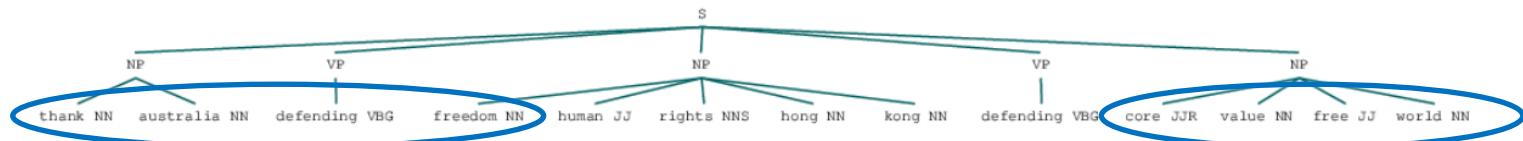
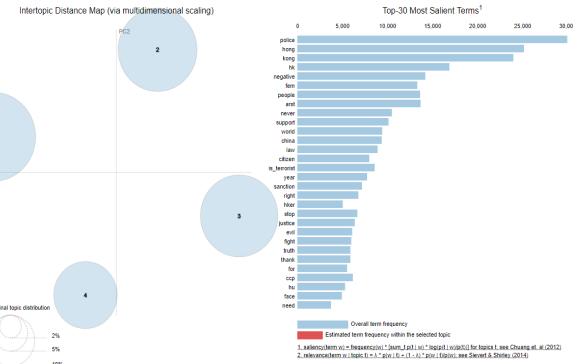
- The summary\* is highlighted below:
  - **Appreciation for support from the world**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC ONE

- The summary\* is highlighted below:
  - Appreciation for support from the world
  - Mentions of **support from the free world**

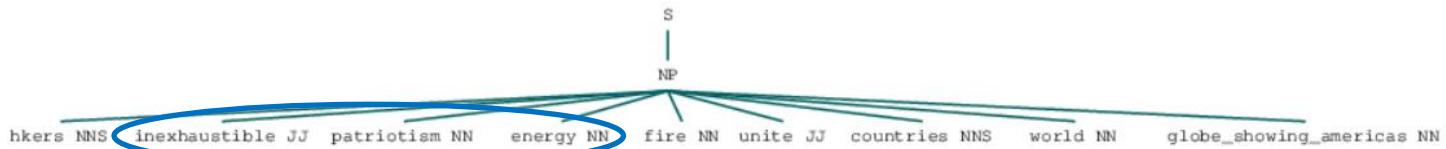
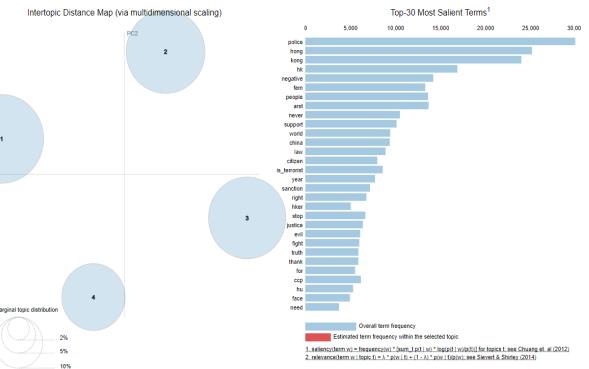


\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC TWO

---

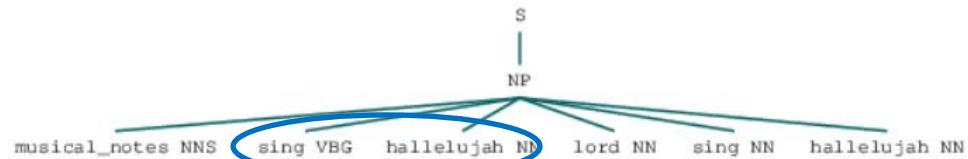
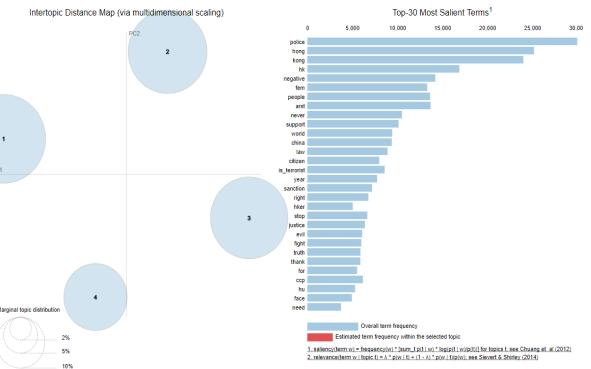
- The summary\* is highlighted below:
  - Protestors **sense of comradery**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC TWO

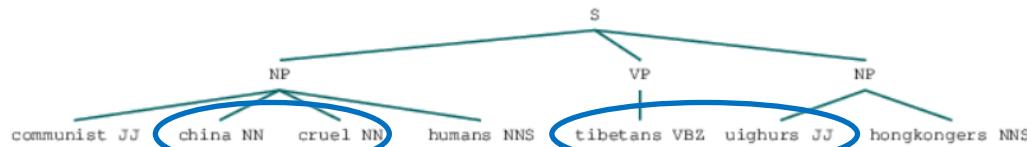
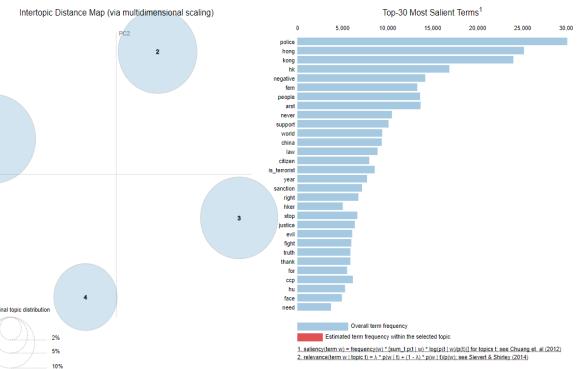
- The summary\* is highlighted below:
  - Protestors sense of comradery
  - **Usage of positivity** to achieve their goals



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC THREE

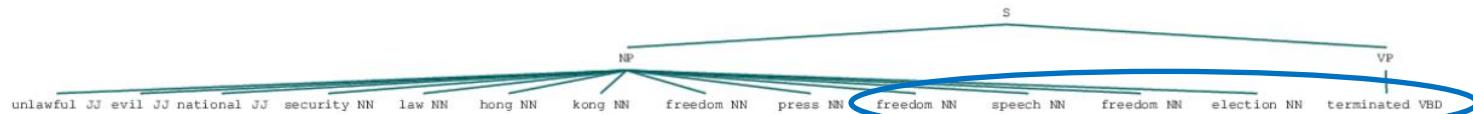
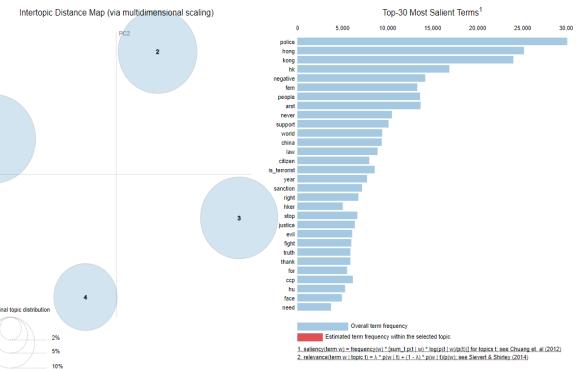
- The summary\* is highlighted below:
  - Vocalize **their opinion** towards what they **believe China's goal is**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC THREE

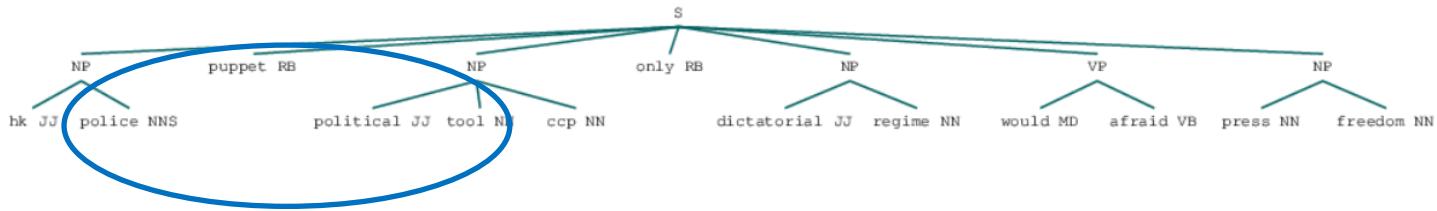
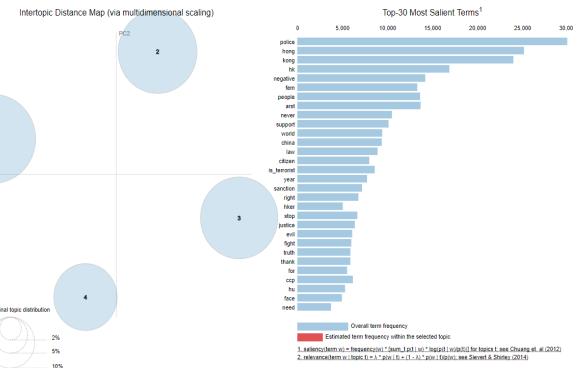
- The summary\* is highlighted below:
  - Vocalize their opinion towards what they believe China's goal is
  - **Belief that the NSL will set limitations on their freedom**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC THREE

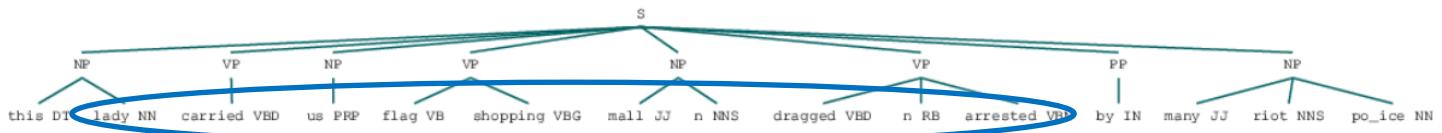
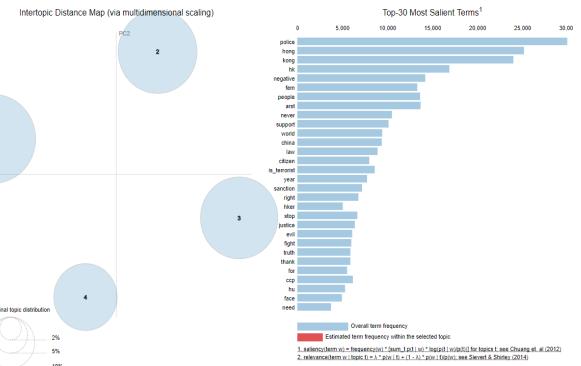
- The summary\* is highlighted below:
  - Vocalize their opinion towards what they believe China's goal is
  - Belief that the NSL will set limitations on their freedom
  - **Belief that the HK police are puppets**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC FOUR

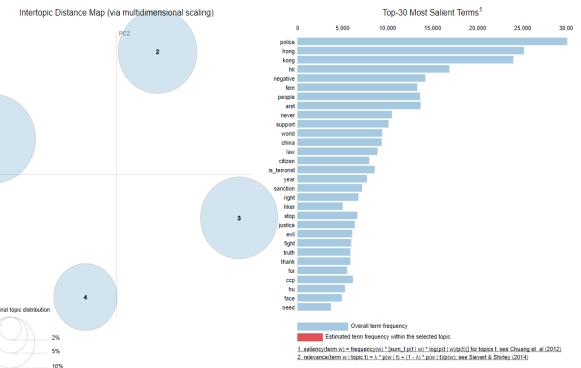
- The summary\* is highlighted below:
  - Arrests



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC FOUR

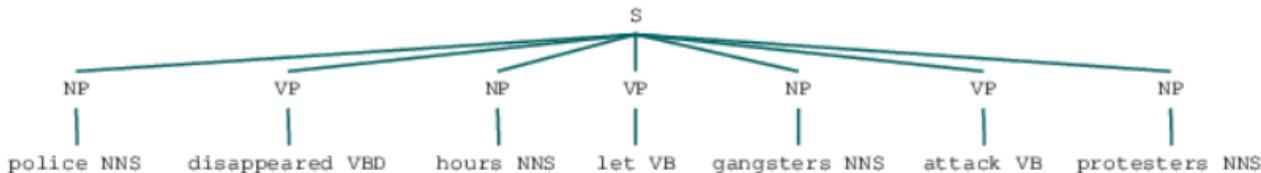
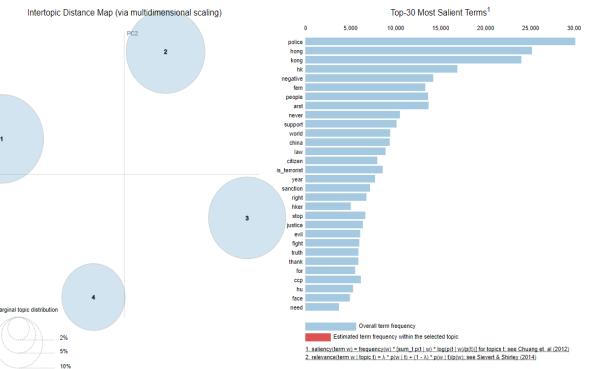
- The summary\* is highlighted below:
  - Arrests
  - **Confrontations with legal authorities** with charges



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC FOUR

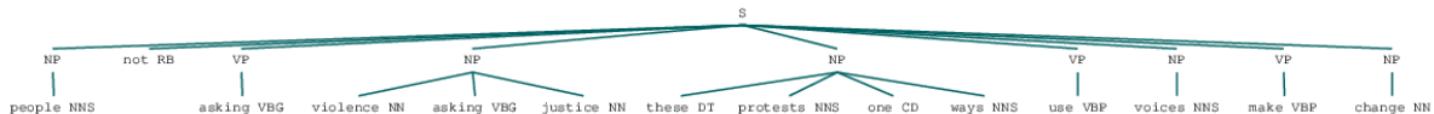
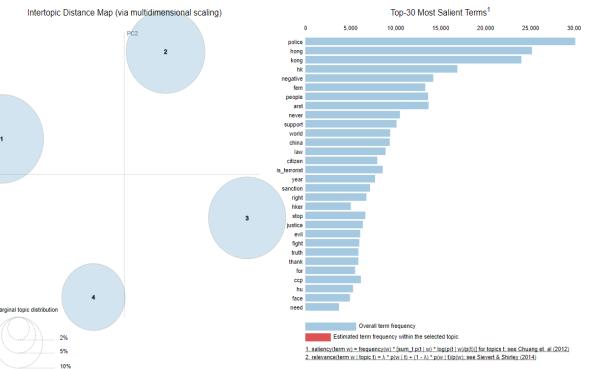
- The summary\* is highlighted below:
  - Arrests
  - Confrontations with legal authorities with charges
  - Police left while **gangsters came in**



\* A larger subset of examples in both the Report & Source Code

# TWITTER | TOPIC FOUR

- The summary\* is highlighted below:
  - Arrests
  - Confrontations with legal authorities with charges
  - Police left while gangsters came in
  - Protestors belief that **the media confuses their intentions**



\* A larger subset of examples in both the Report & Source Code



## TOPIC 1

- Appreciation for support from abroad
- Helping their perceived fight for freedom



## TOPIC 2

- Comradery towards their fight
- Usage of positive Emoji's

---

## TWITTER MODELING TAKE-AWAYS

---



## TOPIC 3

- Their negative opinion towards what they believe China's goal with the NSL is
- Belief that the HK police are puppets



## TOPIC 4

- Police & legal authorities confrontations
- The protestors belief that the media confuses their intentions

# 06

## READ BETWEEN THE LINES ("RBL")

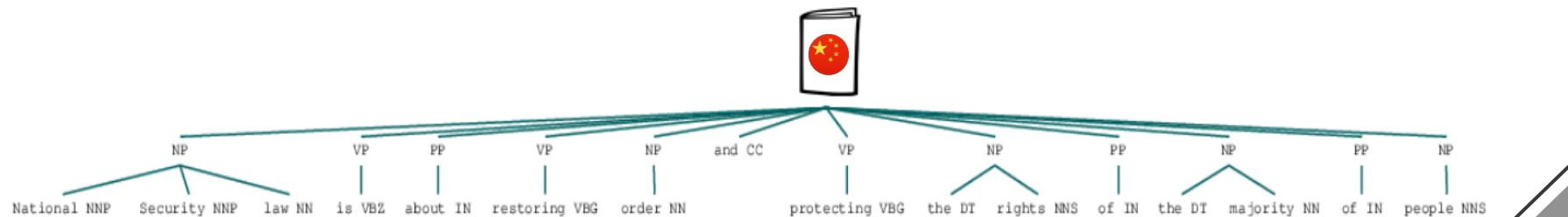
---



# RBL | BRINGING THEM ALL TOGETHER

---

- The narrative which connects them all may be summarized below:
  - **China believes that the NSL will safeguard the people of HK & bring long-term prosperity**

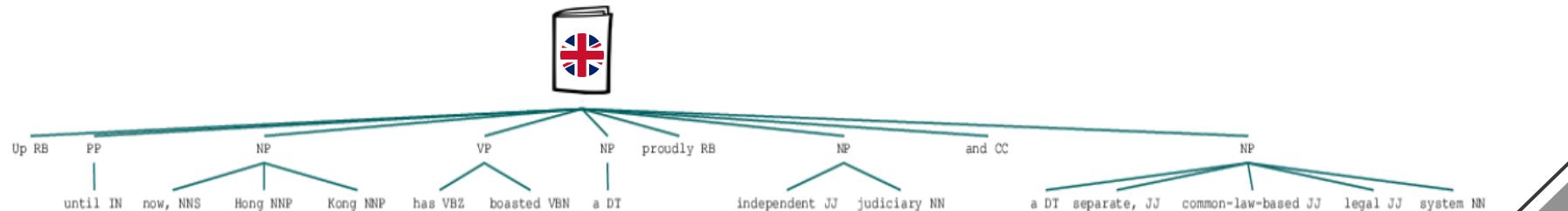


\* A larger subset of examples in both the Report & Source Code

# RBL | BRINGING THEM ALL TOGETHER

---

- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the **Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom**

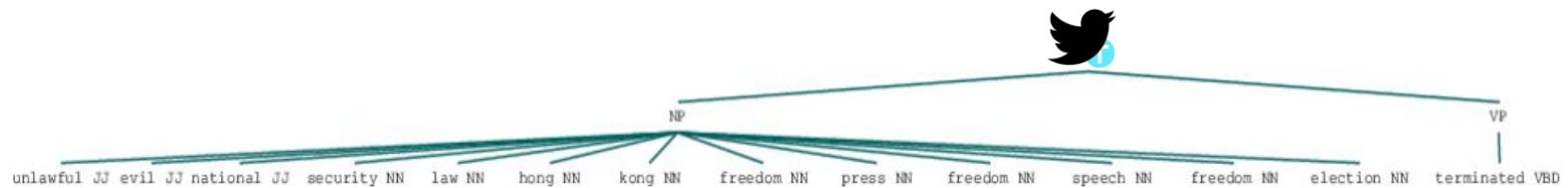


\* A larger subset of examples in both the Report & Source Code

# RBL | BRINGING THEM ALL TOGETHER

---

- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; **including the Tweeters**

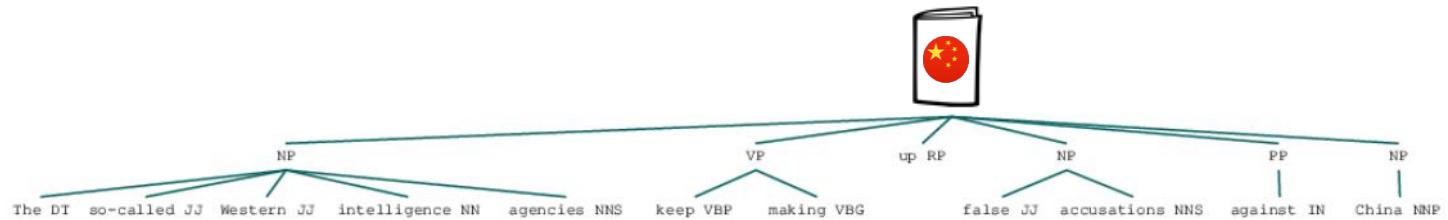


\* A larger subset of examples in both the Report & Source Code

# RBL | BRINGING THEM ALL TOGETHER

---

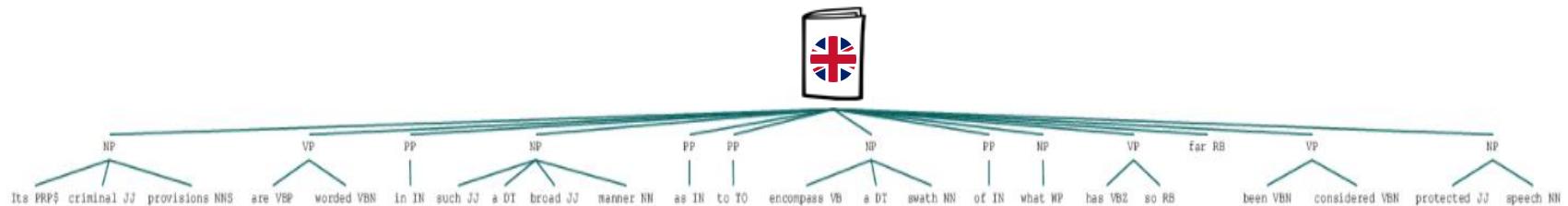
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - **China believes the Int'l community have a misunderstanding of the pure intent of the NSL.**



# RBL | BRINGING THEM ALL TOGETHER

---

- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. **Int'l News believes the NSL has vague wordage to possibly serve other intentions.**

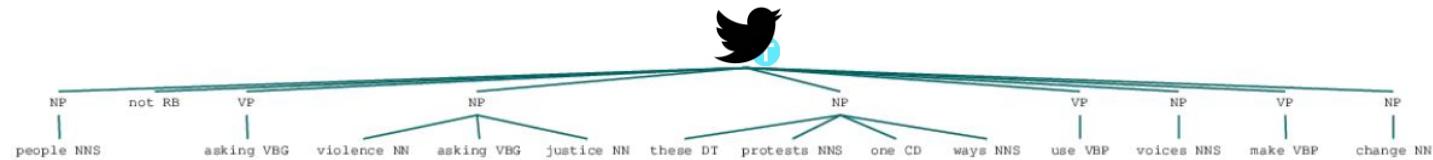


\* A larger subset of examples in both the Report & Source Code

# RBL | BRINGING THEM ALL TOGETHER

---

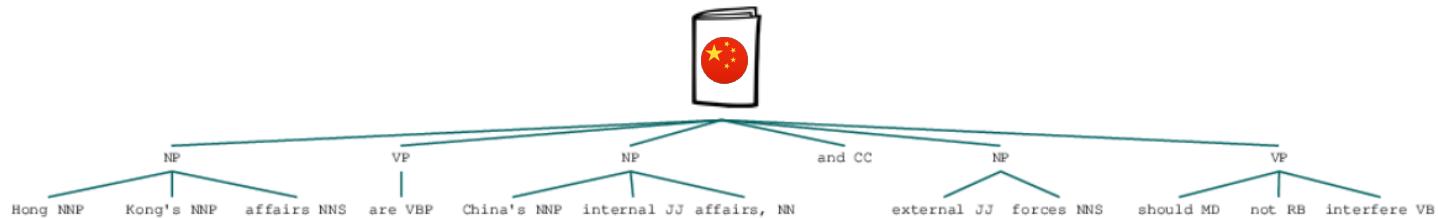
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. **The Tweeters believe protests are their only way to achieve the society they wish for**



# RBL | BRINGING THEM ALL TOGETHER

---

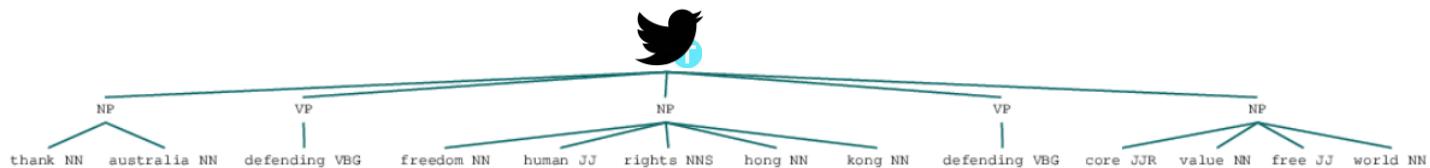
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. The Tweeters believe protests are their only way to achieve the society they wish for
  - **China believes that the Int'l community should stay out of Chinese affairs**



# RBL | BRINGING THEM ALL TOGETHER

---

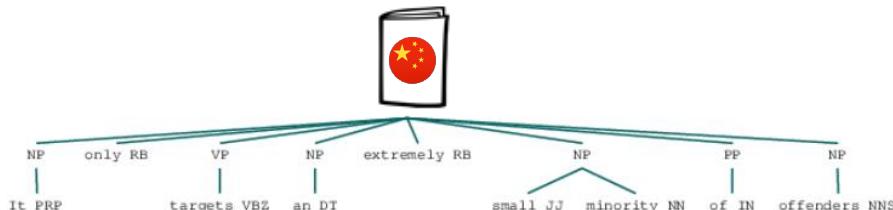
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. The Tweeters believe protests are their only way to achieve the society they wish for
  - China believes that the Int'l community should stay out of Chinese affairs & **the Tweeters are thankful for the support from the Int'l community**



# RBL | BRINGING THEM ALL TOGETHER

---

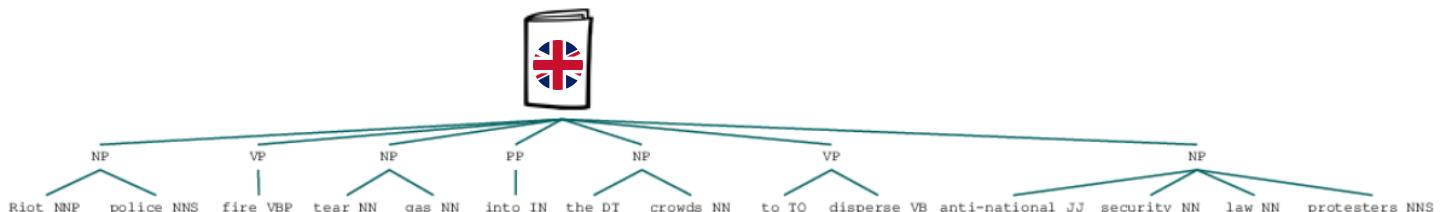
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. The Tweeters believe protests are their only way to achieve the society they wish for
  - China believes that the Int'l community should stay out of Chinese affairs & the Tweeters are thankful for the support from the Int'l community
  - **China's position on police engagement is justified & only targets a minority of criminals.**



# RBL | BRINGING THEM ALL TOGETHER

---

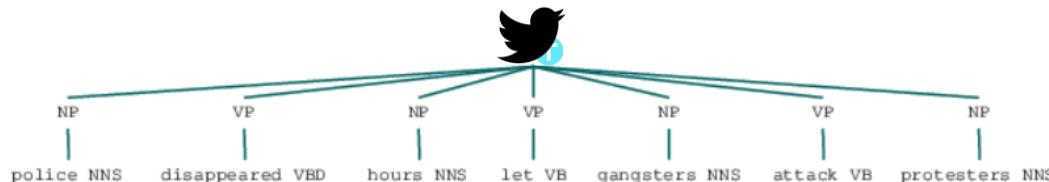
- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. The Tweeters believe protests are their only way to achieve the society they wish for
  - China believes that the Int'l community should stay out of Chinese affairs & the Tweeters are thankful for the support from the Int'l community
  - China's position on police engagement is justified & only targets a minority of criminals. **Int'l News highlights the engagements**



# RBL | BRINGING THEM ALL TOGETHER

---

- The narrative which connects them all may be summarized below:
  - China believes that the NSL will safeguard the people of HK & bring long-term prosperity while conversely the Int'l community believe the NSL infringes on Common Law & will infringe on HK's freedom; including the Tweeters
  - China believes the Int'l community have a misunderstanding of the pure intent of the NSL. Int'l News believes the NSL has vague wordage to possibly serve other intentions. The Tweeters believe protests are their only way to achieve the society they wish for
  - China believes that the Int'l community should stay out of Chinese affairs & the Tweeters are thankful for the support from the Int'l community
  - China's position on police engagement is justified & only targets a minority of criminals. Int'l News highlights the engagements while **Tweeters convey a belief that the police are going above their call of duty or not on duty**



\* A larger subset of examples in both the Report & Source Code

A photograph of a woman with short dark hair and glasses, wearing a purple jacket, speaking into a microphone. She is positioned in front of a large, brightly lit city skyline at night, with numerous skyscrapers and a prominent Ferris wheel reflected in the water in the foreground.

# 07

---

## NEXT STEPS

# NEXT STEPS

---



## TRANSLATION

Google Translate sent many errors for Twitter, we had to limit our data to English

## FORMATTING

Minor formatting issues that didn't present issues for Sentiment Scores but should be addressed

## COUNT VECTORIZER

Re-examining the data frames with attention to Count Vectorizer may help

# THANK YOU

谢谢 | 多谢

By, Rand Sobczak Jr.  
[rand.sobczak@gmail.com](mailto:rand.sobczak@gmail.com)  
+1 313 443 8634

