

# Homework 2

## Math 189Z – Covid-19: Data Analytics and Machine Learning

Nico Espinosa Dice

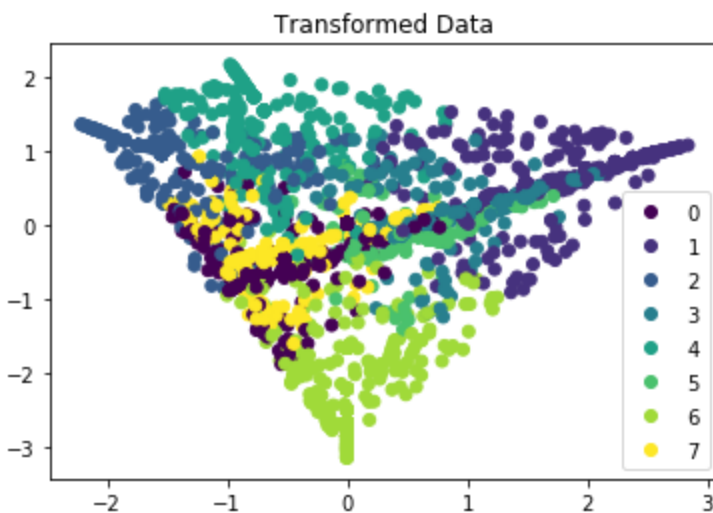
April 16, 2020

### Task 1: List of Stop Words

#### TASK 1: add stop words

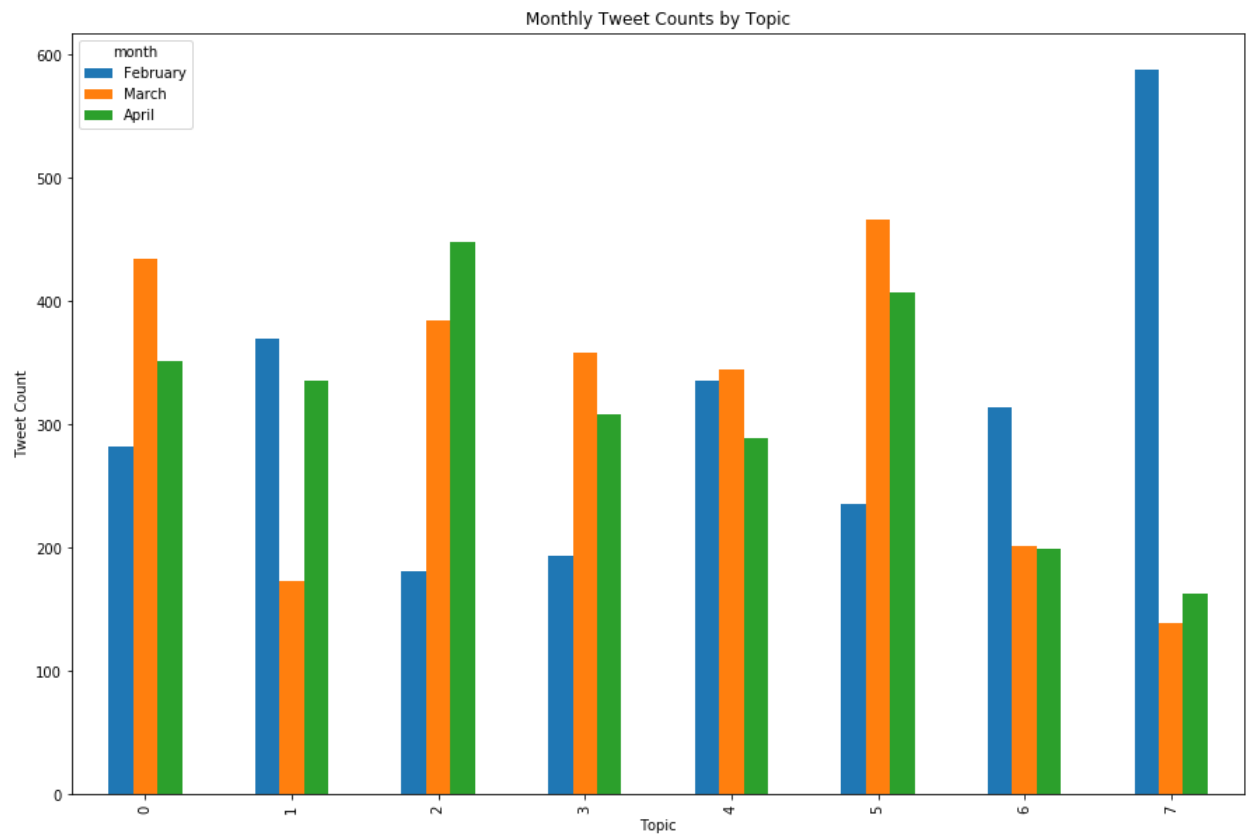
```
In [5]: my_stops = text.ENGLISH_STOP_WORDS.union(
    ["com", "www", "https", "COVID-19", "Covid-19", "Coronavirus", "virus",
     "corona", "coronavirus", "covid", "covid-19", "19", "covid 19", "COVID 19", "Covid 19", "covid19",
     "Covid19", "COVID19"]
)
```

### Task 2 and 3: Image and Analysis of PCA Results on Tweet Topics



The plot above shows us that topic models are partially structured, as parts of the graph share similarities to the previous PCA plot of structured data. In particular, the "edges" created by varying the number of dimensions is partially present in the Tweet PCA plot, but it is noticeably less clear than the structured data's plot. Nevertheless, there is clearly more structure to this plot than the unstructured data's plot, indicating that the topics do have structure, suggesting that we will be able to gain useful information from a topic model.

# Task 4: Monthly Tweet Counts by Topic



# Task 5: Discussion of Results

## TASK 5: Discuss Results

Now it is time to do qualitative analysis of your topic modeling. Your final task is to look at the graph above and choose a topic with a majority from each month and a topic that seems to be present in all months (4 total topics) and assign a semantic meaning to each based on the words associated with that topic (when you print the topics using `print_topics()` they are printed in the same order as the graph). For example if the words were ['test', 'positive', 'negative'...] you could say 'Testing for COVID-19'. Comment on whether the results on the bar graph make sense after assigning the semantic meaning to the topic.

### Present in all four months

**Topic 4:** ['china', 'people', 'chinese', 'amp', 'world', 'government', 'response', 'health', 'media', 'spread', 'trump', 'scientists']

**Semantic Meaning:** During February, March, and April, there was clearly focus on *world governments* and *scientists* and their *responses* to combat the *spread* of Covid-19.

**Interpretation:** The results make sense because it is likely that in the development of a global pandemic, many people would focus on the response from governments and scientists, as well as developments in the spread of the virus.

### Majority in February

**Topic 7:** ['china', 'chinese', 'new', 'outbreak', 'global', 'impact', 'apple', 'hong', 'kong', 'reut', 'rs', 'amid']

**Semantic Meaning:** During February, there was focus on the *outbreak* in *China* that could – and has had – *global impact*.

**Interpretation:** Because the virus developed in China, it makes sense that in the early to mid development of the virus, the focus of Twitter users would be on China. In particular, because the virus was still in the early to mid stage in February, many people used the phrase "outbreak" to describe the virus, where they (led by the WHO) have since moved to classifying the virus as a "pandemic."

### Majority in March

**Topic 5:** ['like', 'people', 'just', 'flu', 'amp', 'going', 'don', 'know', 'really', 'crisis', 'doesn', 'safe']

**Semantic Meaning and Interpretation:** During March, many more people began worrying about their *safety*, which makes sense given that the virus really started to spread during the month of March. Furthermore, as more news and research came out about the virus, many individuals compared the virus to the *flu*, either because of similar symptoms or because they underestimated the severity of the virus. The comparisons of Covid-19 to the flu have since dwindled in the month of April because the death rate has risen, and many people have been affected by the death of friends or family. Furthermore, as the virus began spreading around the world in March, people began to use the term *crisis* to describe the situation (increasing the severity of the description from "outbreak").

### Majority in April

**Topic 2:** ['time', 'people', 'amp', 'help', 'thank', 'stay', 'social', 'spread', 'doctors', 'workers', 'home', 'let']

**Semantic Meaning and Interpretation:** During April, there has been a major movement to thank doctors and health care workers, among others, for their sacrifices and help combating the virus. This is reflected in the topic, as *thank*, *doctors*, *workers*, and *help* are all in the topic. Furthermore, many US states have imposed shelter-in-place orders that strongly suggest or mandate that citizens *stay home*.

## Topics:

```
[ 'patients', 'hospital', 'deaths', 'use', 'new', 'cases', 'died', 'week', 'april', 'today', 'number', 'city' ]
[ 'president', 'trump', 'like', 'government', 'people', 'negative', 'pandemic', 'response', 'lockdown', 'days', 'queue', 'face' ]
[ 'free', 'china', 'people', 'hope', 'pandemic', 'support', 'let', 'fight', 'death', 'help', 'doesn', 'narendramodi' ]
[ 'news', 'trump', 'deaths', 'ago', 'cases', '2020', '04', 'away', 'health', 'disease', 'amp', 'know' ]
[ 'hydroxychloroquine', 'trump', 'says', 'people', 'just', 'getting', 'pic', 'twitter', 'dr', 'fauci', 'really', 'pers on' ]
[ 'hospital', 'tiger', 'bronx', 'zoo', 'tests', 'positive', 'tested', 'minister', 'symptoms', 'breaking', 'confirmed', 'ik_savioursofnation' ]
[ 'time', 'covid_19', 'home', 'like', 'people', 'just', 'right', 'amp', 'stay', 'make', 'safe', 'country' ]
[ 'news', 'world', 'new', 'pandemic', 'stop', 'uk', 'spread', 'testing', '5g', 'help', 'west', 'india' ]
```