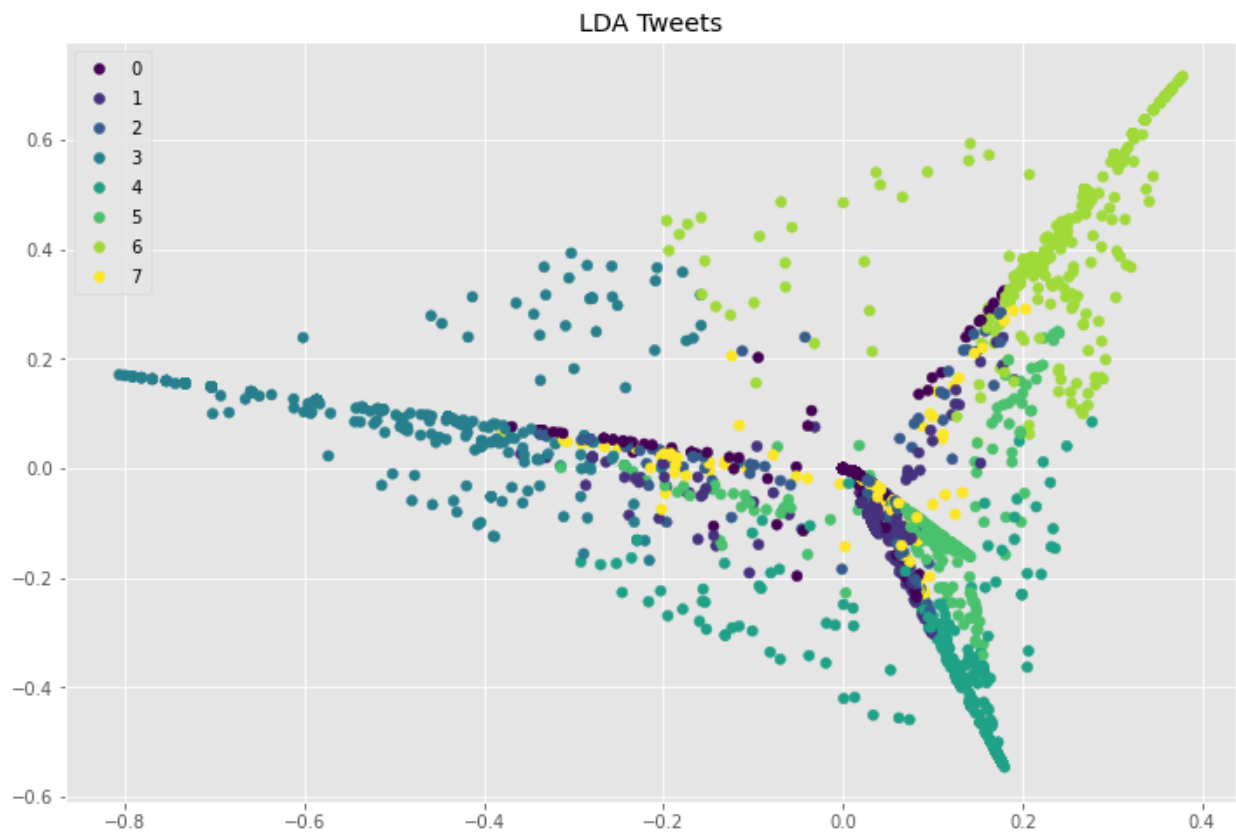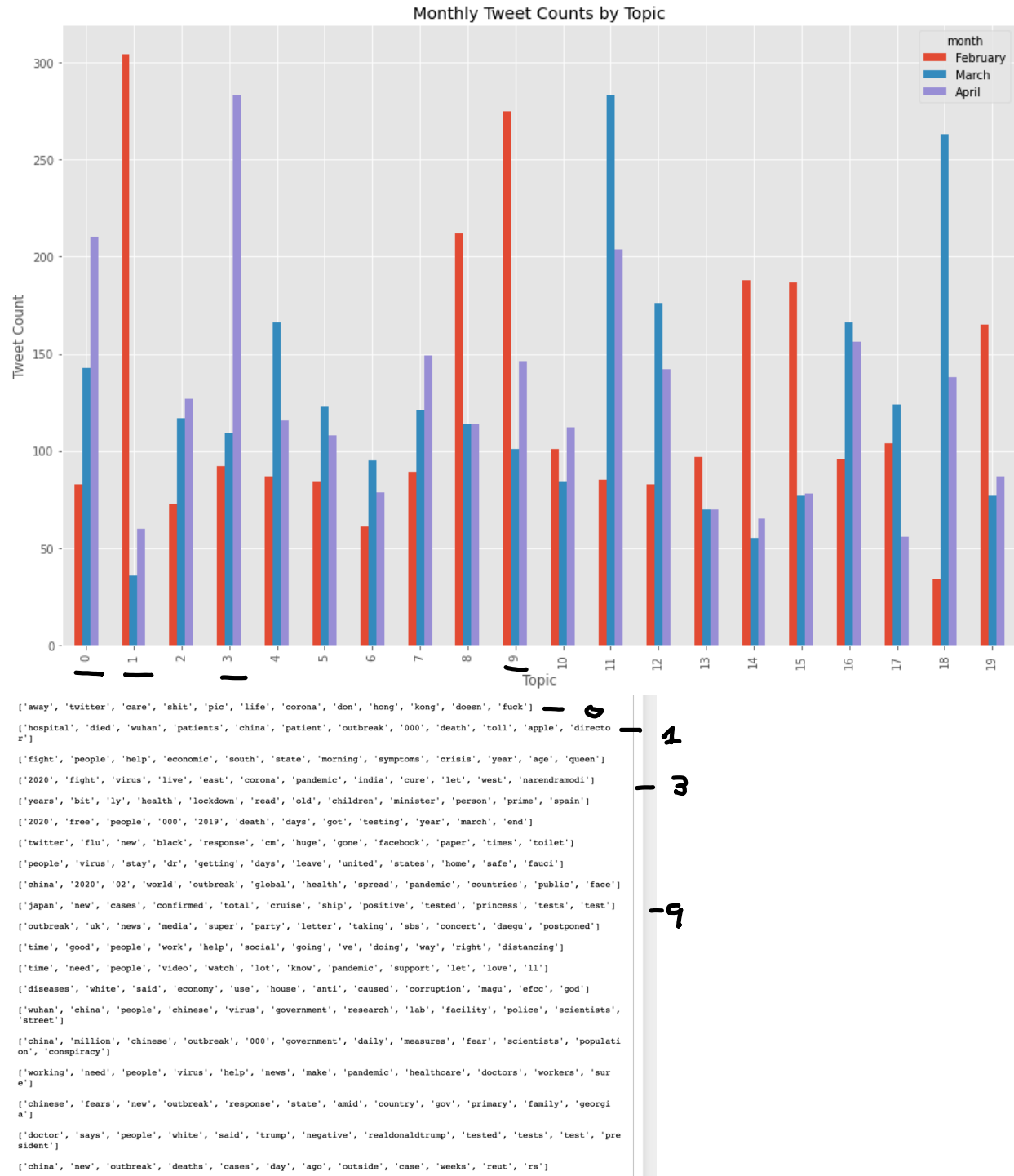['we',' has', 'is', 'can', 'just', 'a', 'like', 'http', 'the', 'if', 'and', 'that', 'almost', 'every', 'can', 'com', 'https', 'coronavirus', 'covid_19','covid19','covid', "19", 'www', 'breaking', 'amp']

The data appears to be structured, albeit less than the generated data. We see that there are a significant number of tweets that do not seem to belong to a topic, the majority of them are structured around the principal topics that we found using LDA. This means that there is a trend within this dataset to be clustered around specific topics. Basically, this data tells us what we already know: people on twitter love recycling ideas.

Monthly Tweet Counts by Topic

['away', 'twitter', 'care', 'shit', 'pic', 'life', 'corona', 'don', 'hong', 'kong', 'doesn', 'fuck'] — 0

['hospital', 'died', 'wuhan', 'patients', 'china', 'patient', 'outbreak', '000', 'death', 'toll', 'apple', 'director'] — 1

['fight', 'people', 'help', 'economic', 'south', 'state', 'morning', 'symptoms', 'crisis', 'year', 'age', 'queen']

['2020', 'fight', 'virus', 'live', 'east', 'corona', 'pandemic', 'india', 'cure', 'let', 'west', 'narendramodi'] — 3

['years', 'bit', 'ly', 'health', 'lockdown', 'read', 'old', 'children', 'minister', 'person', 'prime', 'spain']

['2020', 'free', 'people', '000', '2019', 'death', 'days', 'got', 'testing', 'year', 'march', 'end']

['twitter', 'flu', 'new', 'black', 'response', 'cm', 'huge', 'gone', 'facebook', 'paper', 'times', 'toilet']

['people', 'virus', 'stay', 'dr', 'getting', 'days', 'leave', 'united', 'states', 'home', 'safe', 'fauci']

['china', '2020', '02', 'world', 'outbreak', 'global', 'health', 'spread', 'pandemic', 'countries', 'public', 'face']

['japan', 'new', 'cases', 'confirmed', 'total', 'cruise', 'ship', 'positive', 'tested', 'princess', 'tests', 'test'] — 9

['outbreak', 'uk', 'news', 'media', 'super', 'party', 'letter', 'taking', 'sbs', 'concert', 'daegu', 'postponed']

['time', 'good', 'people', 'work', 'help', 'social', 'going', 've', 'doing', 'way', 'right', 'distancing']

['time', 'need', 'people', 'video', 'watch', 'lot', 'know', 'pandemic', 'support', 'let', 'love', 'll']

['diseases', 'white', 'said', 'economy', 'use', 'house', 'anti', 'caused', 'corruption', 'magu', 'efcc', 'god']

['wuhan', 'china', 'people', 'chinese', 'virus', 'government', 'research', 'lab', 'facility', 'police', 'scientists', 'street']

['china', 'million', 'chinese', 'outbreak', '000', 'government', 'daily', 'measures', 'fear', 'scientists', 'population', 'conspiracy']

['working', 'need', 'people', 'virus', 'help', 'news', 'make', 'pandemic', 'healthcare', 'doctors', 'workers', 'sure']

['chinese', 'fears', 'new', 'outbreak', 'response', 'state', 'amid', 'country', 'gov', 'primary', 'family', 'georgia']

['doctor', 'says', 'people', 'white', 'said', 'trump', 'negative', 'realdonaldtrump', 'tested', 'tests', 'test', 'president']

['china', 'new', 'outbreak', 'deaths', 'cases', 'day', 'ago', 'outside', 'case', 'weeks', 'reut', 'rs']

I'd like to do something slightly different than what the discussion guidelines say, as there isn't a clear winner among topics. We see that topic 0 steadily increases every month, a collection of words that seem rather pessimistic. It contains the only swear words among topics, and stubs of "don't", "doesn't", "care", "away", and seems to have a meta streak, containing "twitter" as well. Obviously, people are getting more and more fed up with shelter-in-place and are turning

to twitter to vent. Hong Kong likely got grouped up with this as it is an emotionally charged topic, prone to similar language. I call this the **"Table flip"** topic.

Topic 1 peaked in February, back when this seemed only like China's (and companies that manufacture there) problem. This is the **"What's going on over there?"** topic. This explains the rapid decline as Covid-19 became a pandemic and is supported by the presence of "Wuhan", "patients", and "outbreak", among others.

Let's transition from the most populous country to the second most.

Topic 3 has peaked recently, as the spread of Corona has hit India. Despite being less populous than China, it is more densely populated, and this poses significant challenges for their health care system. This is the **"This could be really bad…"** topic. This topic contains "india" and the twitter handle of their Prime Minister.

Finally let's look at Topic 9: **"I'd take my chances with the Titanic over this."**
This one's cut and dry. This is about the Diamond Princess, the cruise ship that got quarantined in Japanese waters at the start of February. Every word in the vector is clearly a common word from the constant updates on the spread of the virus on the ship.

```
['japan', 'new', 'cases', 'confirmed', 'total', 'cruise', 'ship', 'positive', 'tested', 'princess', 'tests', 'test'].
```

Overall, some very cool trends to visualize.