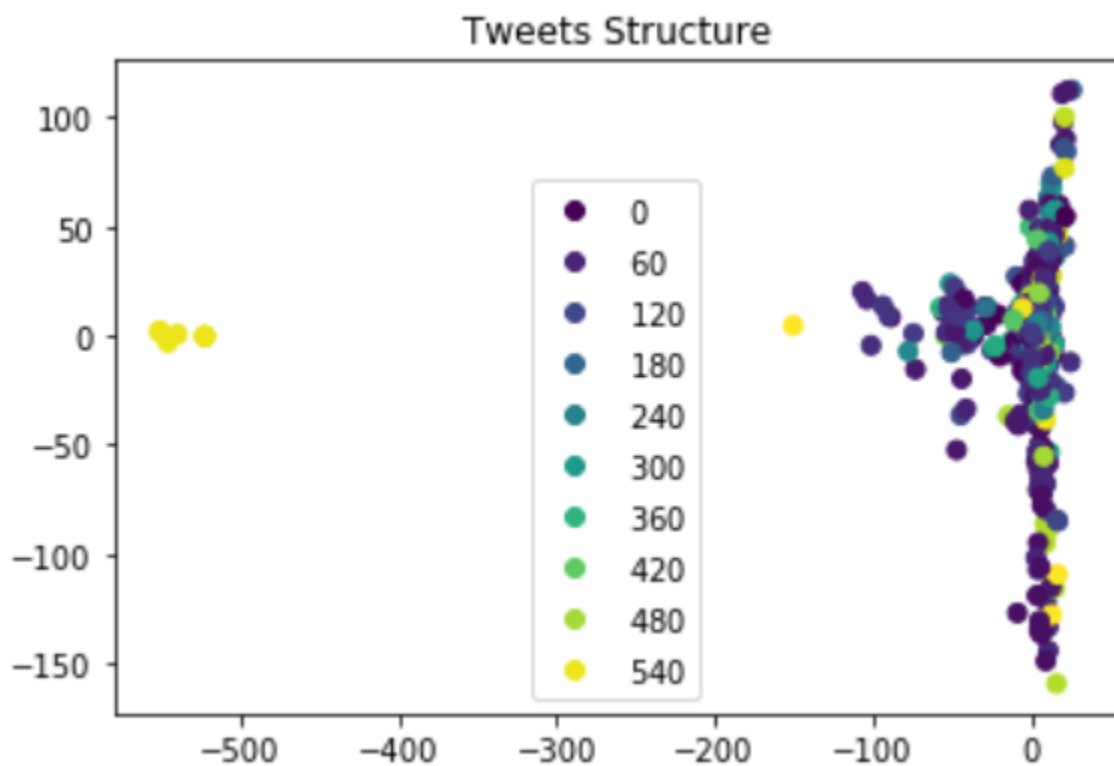MATH 189Z Homework 2

Jeremy Tsai
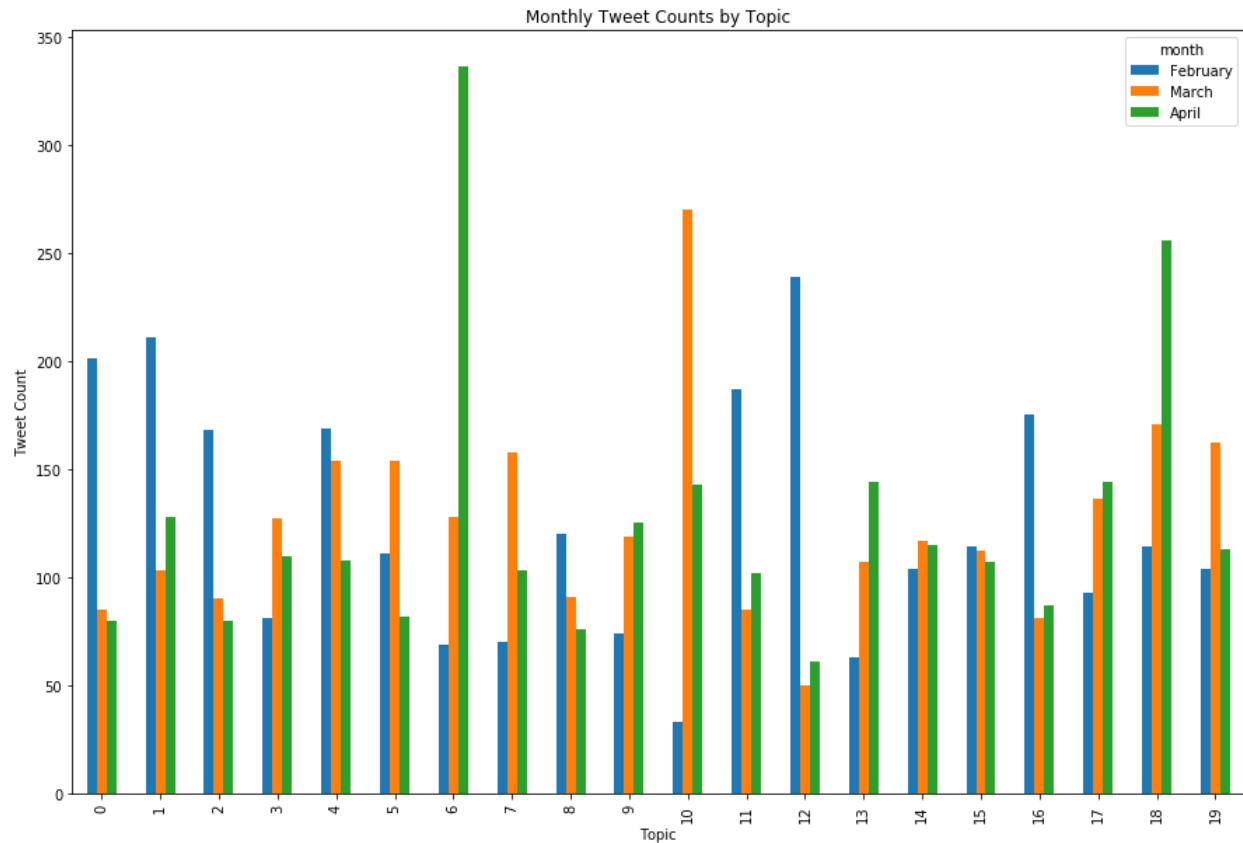
Task 1:

List of stop words:

['COVID-19', 'covid-19', 'COVID', 'covid', 'Coronavirus', 'coronavirus', 'http', 'make', '19', 'like', 'doesn',

'com', 'https', 'amid', 'covid19', 'bit', 'ly', 'www', 'reut', 'rs', 'let', 'don', 'just', 'covid_19', 'says', 'said',

'did', 'covid—19', '2020', '20', 'thing']

Task 2 and 3:

From the image of our PCA results, we can see that most tweets are clustered in a straight line, with a lone group on the left side of the graph. If the tweets were structured, the colors would be bunched up together, as shown in the structured data example graph. However, in our case, the colors are all jumbled together except for a small group of yellow dots on the left side. This leads me to believe that there is not a lot of organized structure within the tweets regarding the Coronavirus. This means that although the tweets themselves are about the Coronavirus, the words and themes within the tweets are varied and diverse.

Task 4:

Monthly Tweet Counts by Topic

Task 5:

Topic 6 (majority in April):

['fight', 'people', 'tonight', 'virus', 'help', 'lockdown', 'corona', 'india', 'ready', 'fighting', 'pm', 'narendramodi']

The semantic meaning of this topic seems to be centered around the Coronavirus situation that has led to fighting in India and what India's Prime minister Narendra Modi has done regarding the situation. This makes sense since currently in India, there has been waves of violence due to government officials claiming that Coronavirus has been spread by Muslims, leading to various hate crimes across the country. The Prime Minister Narendra Modi leads the government which is known to have a bold Hindu nationalist stance, so there has been blame towards him and his role in increasing violence.

Topic 10 (majority in March):

['white', 'trump', 'negative', 'breaking', 'realdonaldtrump', 'house', 'tested', 'administration', 'tests', 'test', 'president', 'donald']

The semantic meaning of this topic revolves around President of the U.S. Donald Trump and his claims regarding Coronavirus tests, as well as his own testing for Coronavirus which turned out to be negative. This makes sense since Donald Trump has a tendency of tweeting everyday and talking about the Coronavirus, which then prompts many other users to react to them. Additionally, Donald Trump was

tested for Coronavirus in the middle of March and the test came out negative, which was relayed many times throughout Twitter.

Topic 12 (majority in February):

['china', 'virus', 'outbreak', 'global', 'impact', 'apple', 'corona', 'warns', 'demand', 'market', 'production', 'meet']

The semantic meaning of this topic is about the initial outbreak of the Coronavirus around the world and the negative effect it began to have on the global market. This makes sense since during February, the world was starting to realize how big of an issue the Coronavirus would become. The Coronavirus started in China but cases were popping up around the world and stock markets began to suffer. The effects were becoming global and various businesses began to suffer. Large companies such as Apple and Nike began to close down their retail stores.

Topic 18 (prevalent in all months):

['time', 'twitter', 'people', 'virus', 'pic', 'stay', 'social', 'corona', 'really', 'home', 'safe', 'distancing']

The semantic meaning of this topic seems to be that people should stay home and practice social distancing in order to stay safe from the Coronavirus. It makes sense that this topic has been prevalent within the past 3 months on Twitter, since the Coronavirus has had such a large impact for everyone around the world. It has caused schools and businesses to shut down and government mandates for people to quarantine at home. Social distancing began to be enforced so that there would be less large gatherings that would facilitate the spread of the virus. With such large effects to our lifestyles, it was inevitable that the Coronavirus would be heavily mentioned and discussed on social media like Twitter.