INFO 550 Write up

COVID-19 Tracker

COVID-19 has become a critical issue towards global public health since the beginning of 2020. In this project, we tracked the spread of this pandemic over time and explore some interesting issues as follow: What is Linguistic Variation of COVID-19 on Twitter and Google? How do the public feel towards COVID-19 as well as Trump's handling? Any further social impacts were brough by COVID-19?

COVID-19 Global Spread

How does COVID-19 spread from China to the opposite of world, United States over time? See the attached animated gif! You can see that at beginning only China has the most confirmed cases, gradually Iran, Iraq, United Kingdom and United States! And United States turned to be the country that has the most confirmed cases compared to others finally!
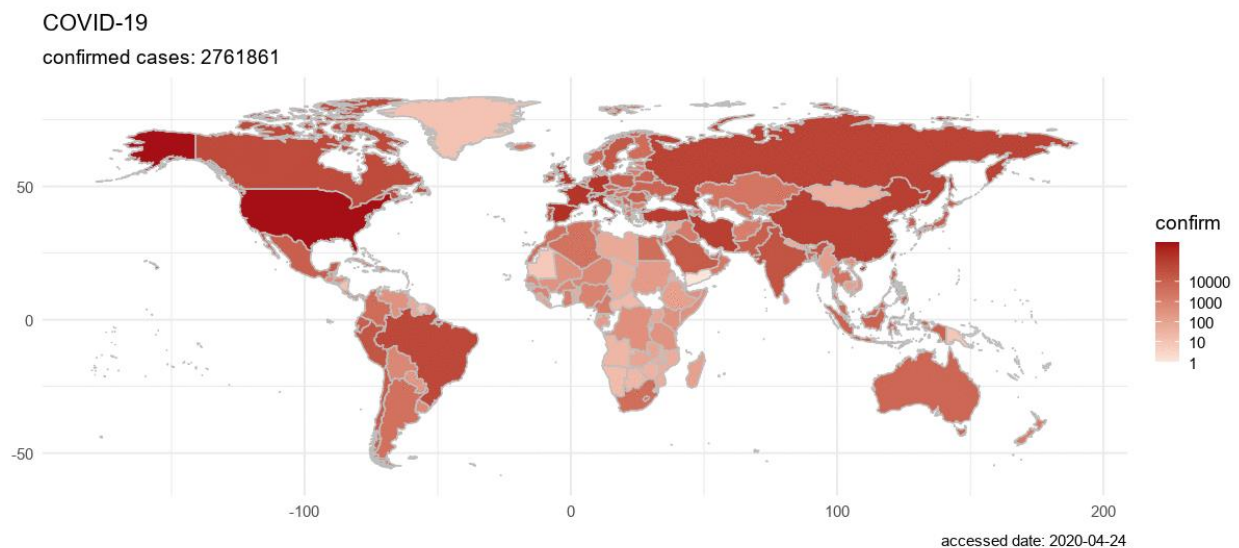

Figure 1 Current confirmed cases in global

So how many confirmed cases would be in the next ten days? From the Figure 2, we can see the linear relationship between the time and the log of confirmed cases in the period of 15th April to 24th April. We can use this model to predict the number of confirmed cases in the next 10 days. And after the day I made prediction, the number of confirmed cases reaches one million, just as this forecast model suggested. Data is from R package "nCov2019".
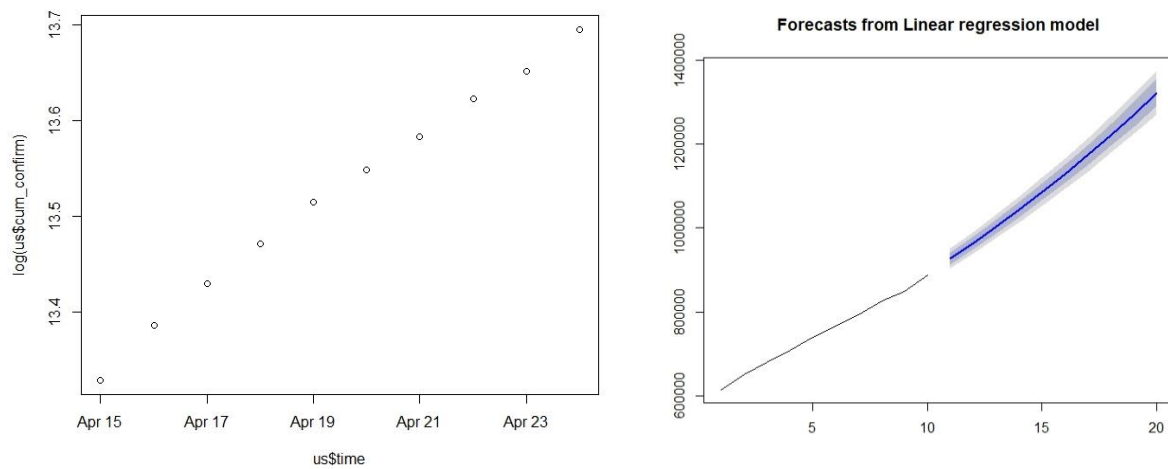
Figure 2 Cases prediction in next 10 days

Linguistic Variation of COVID-19 on Twitter

Coronavirus? Pandemic? Covid-19? There are multiple terms we used to describe this critical issue until WHO officially termed it as COVID-19. So we can track the term change on Twitter, seeing the pattern how the public usually call COVID-19. Here, we parsed the tweets from 100 US Senators and 435 members of the House of Representatives and scraped the last 200 tweets for each of the 535 members of Congress. Since Trump delivered the COVID-19 address on 5th Feb, there is a huge peak at that day on tweets. Democrat politicians tend to tweet more recently. Data was parsed from R package "rtweet".
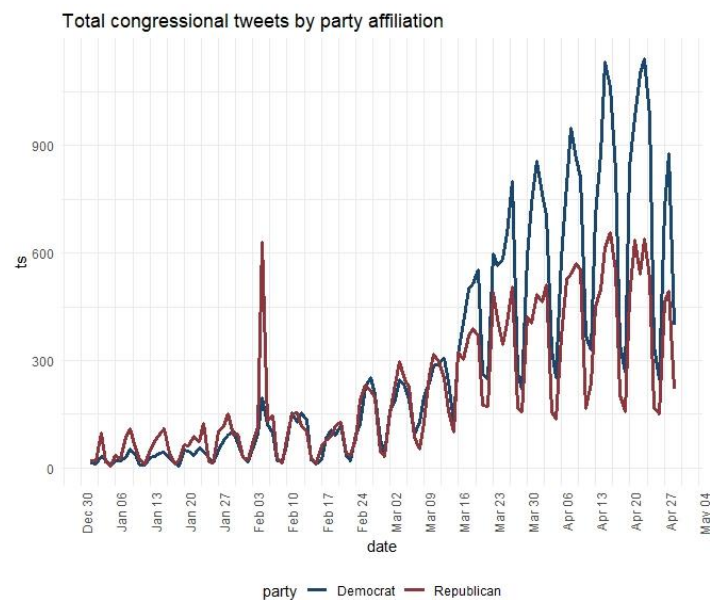

Figure 3 Total congressional tweets by Party

The following two figures show the same pattern referring to 2019 novel coronavirus among US Senators and House Representatives. It's easy for us to find that originally 'coronavirus' was widely used in tweets but more recently, 'pandemic', 'covid19' and 'coronaviruses become equally used to refer. Data is parsed from package 'rtweet'.
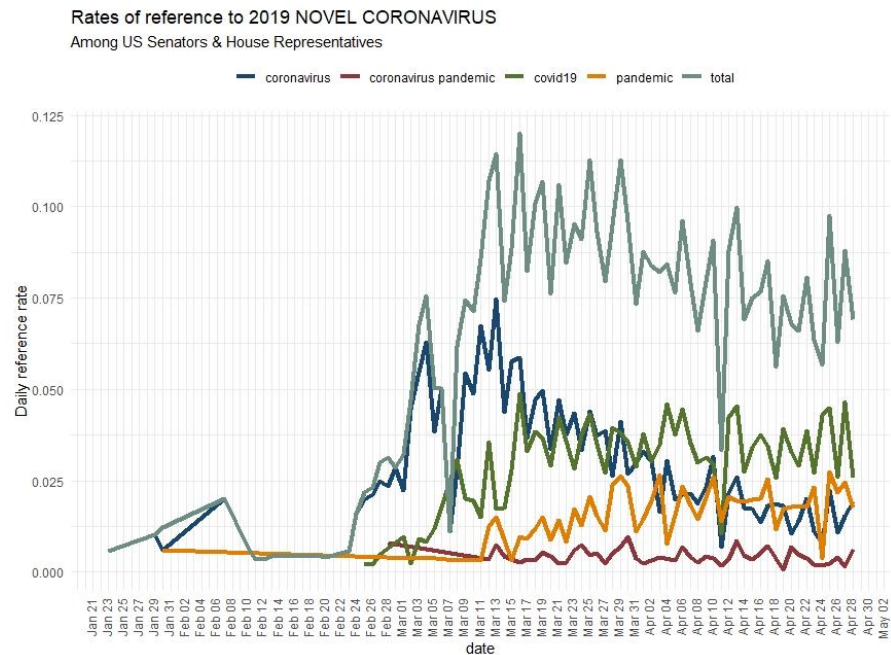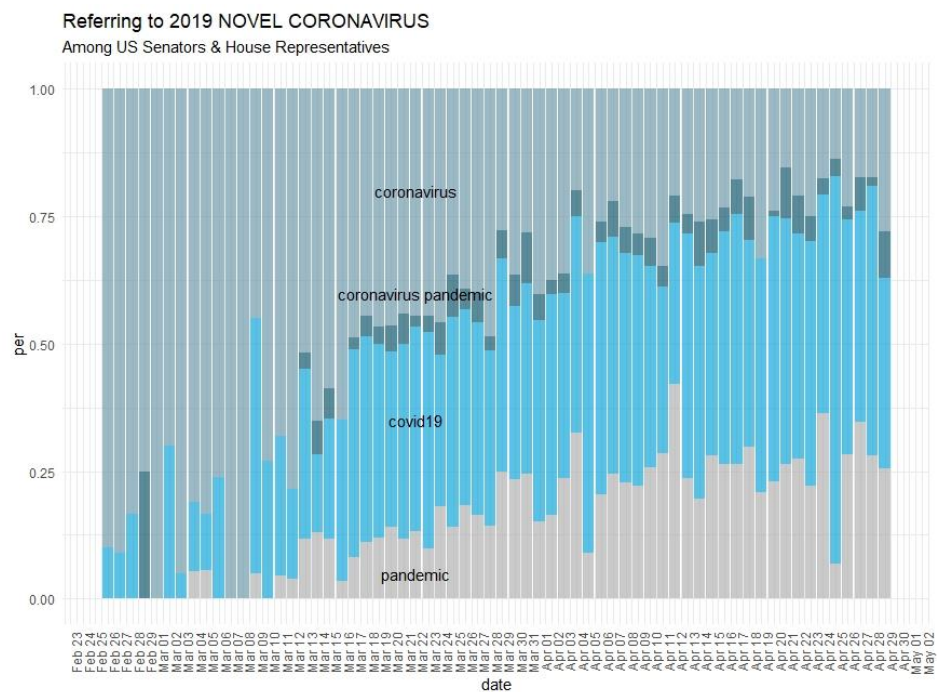

Figure 4 Rates of reference to coronavirus

Figure 5 Probability distribution of referring to coronavirus

Governmental Intervention

What policies your local government publicize to address COVID-19? Chinese government lock down the Wuhan city. American study and work from home… etc. All of these policies impose great effects on citizen' lifestyle. The data was obtained from https://joachim-gassen.github.io/
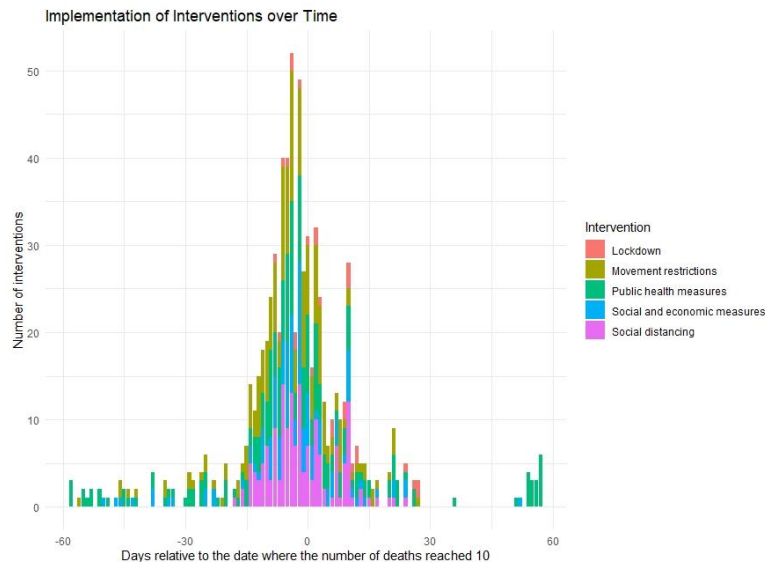
and then visualized.

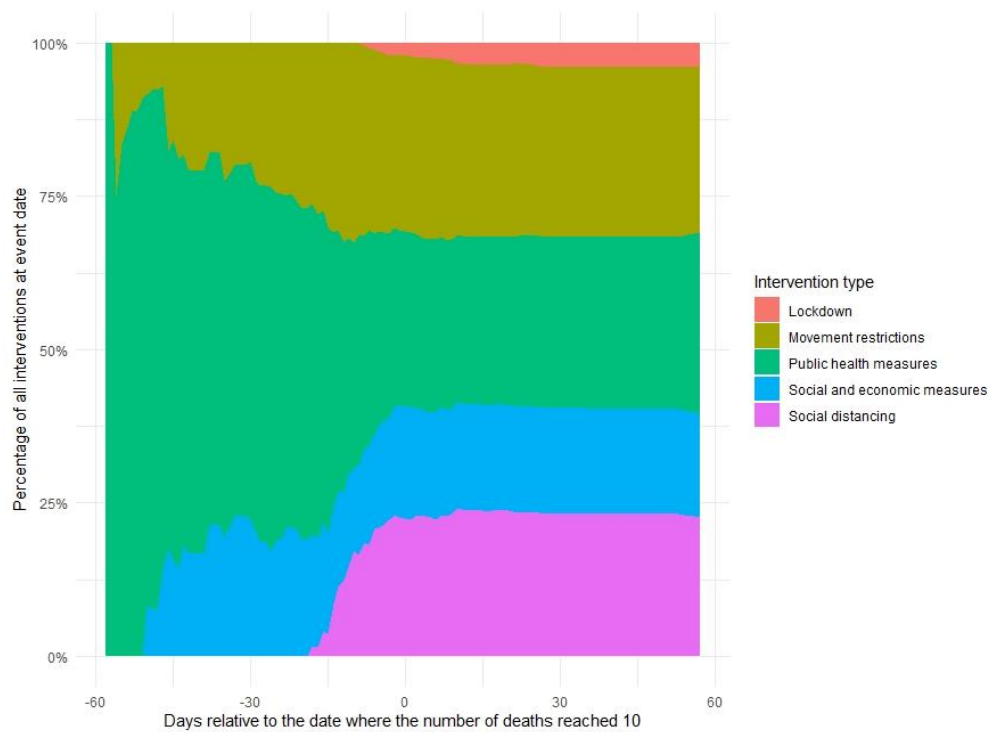Figure 6 Implementation of Interventions over Time

Figure 7 Percentage of all intervention at event date

From the figure above, we can see the pattern that public health measures were first adopted in COVID-19 surveillance and then followed by social distancing. However, for the consideration of economic impacts, lockdown is seldomly used. It should be praised that most governments around world tried immediate multiple interventions before the pandemic started.

Sentiment Analysis on Twitter

What is your feeling towards toward COVID-19? Fear? Happy to stay in health? No sense? Do you trust President Trump that he did a great job on handling COVID-19? The following sentiment analysis with Twitter will tell you what Twitter users are thinking. In the first search, keyword 'COVID19' was used while in the second time, both keyword 'COVID19' and 'Trump' were used.

From Figure 8-9, no significant difference between these two searches, suggesting that the public doesn't feel that President Trump made a lot progress on handling crisis. One thing to point out that most people still hold in positive attitudes and trust the government while relatively less people hold negative attitude and fear COVID-19.
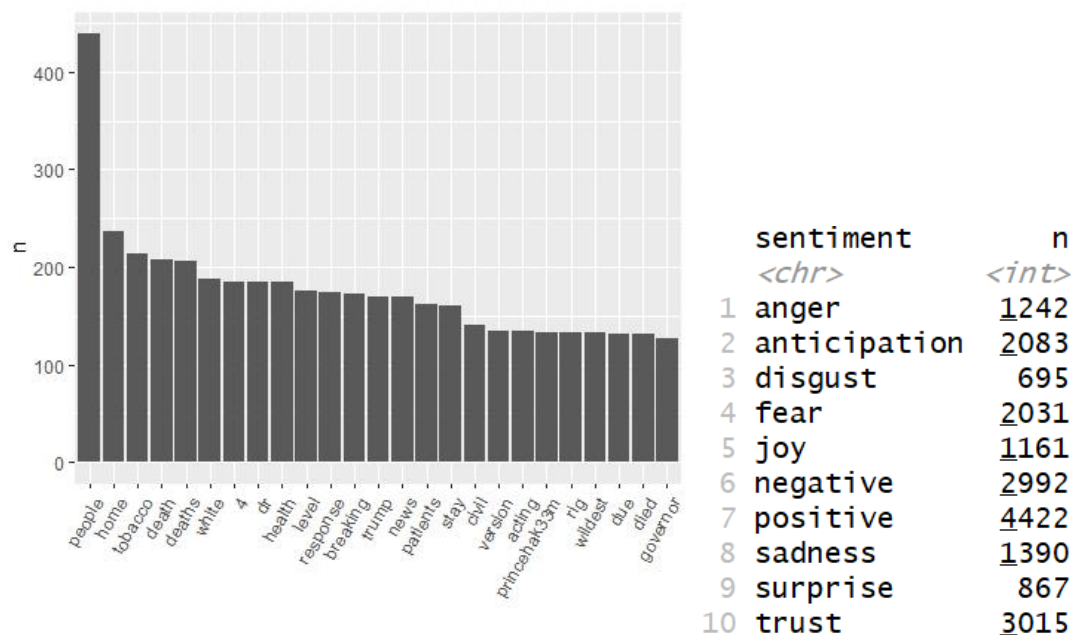

	sentiment	n
	<chr>	<int>
1	anger	1242
2	anticipation	2083
3	disgust	695
4	fear	2031
5	joy	1161
6	negative	2992
7	positive	4422
8	sadness	1390
9	surprise	867
10	trust	3015

Figure 8 The public's feeling to COVID19

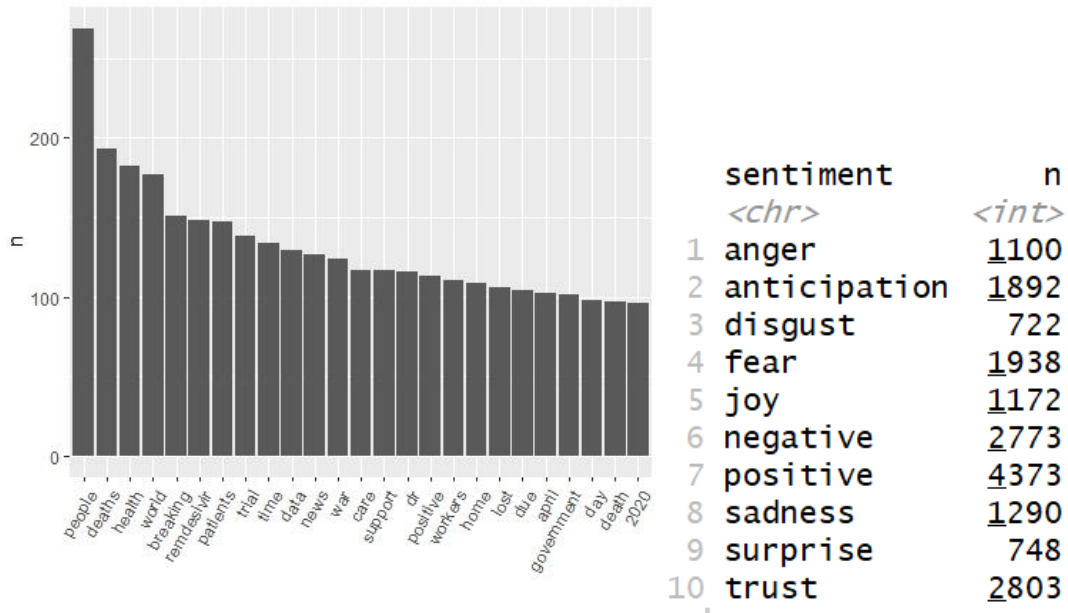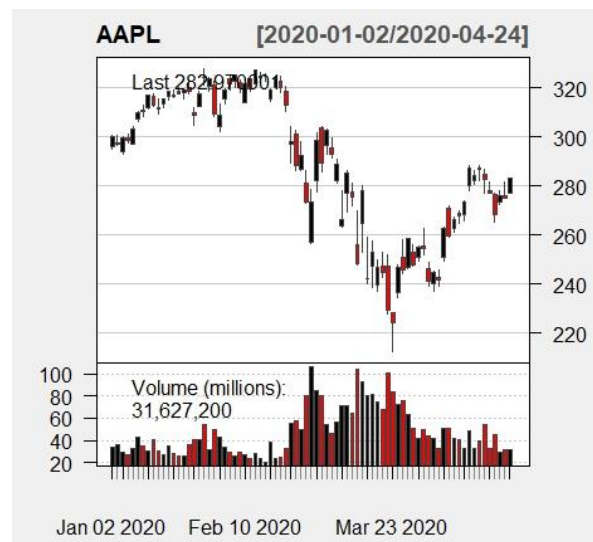|	sentiment	n
	<chr>	<int>
1	anger	1100
2	anticipation	1892
3	disgust	722
4	fear	1938
5	joy	1172
6	negative	2773
7	positive	4373
8	sadness	1290
9	surprise	748
10	trust	2803

Figure 9 The public's feeling to Trump's Handling on COVI10

Social Impacts

Four times circuit breaker within one months, lacking in respiratory maching, unexpected unemployment rate,...etc. Without a doubt, COVID-19 bringswa unprecedented problems to us. In our study, we pick up four typical campanies, ranged from high techinologies, travel and food companies, to show their stock price change through this year. Data was parsed from R package 'quanmod' and then be visualized.
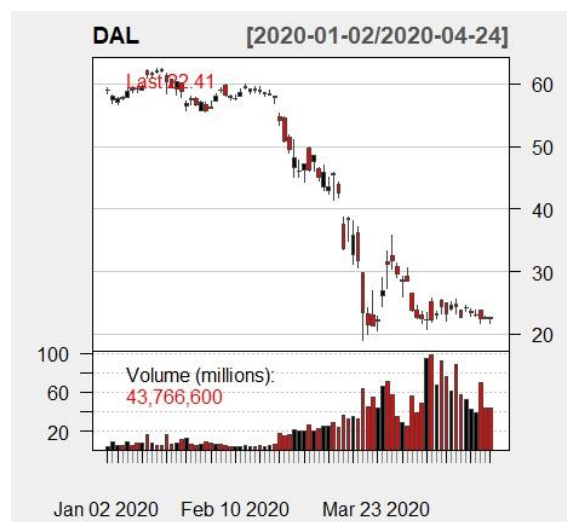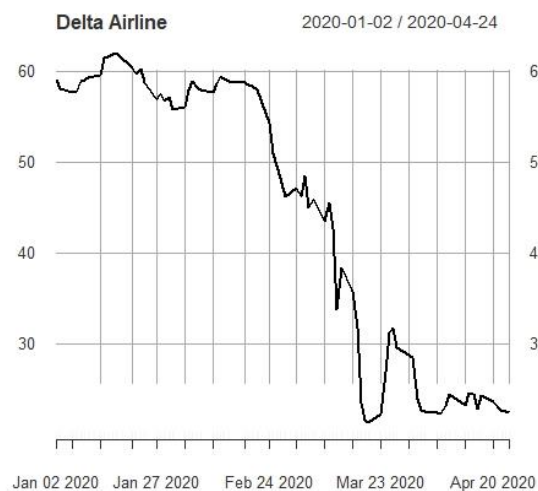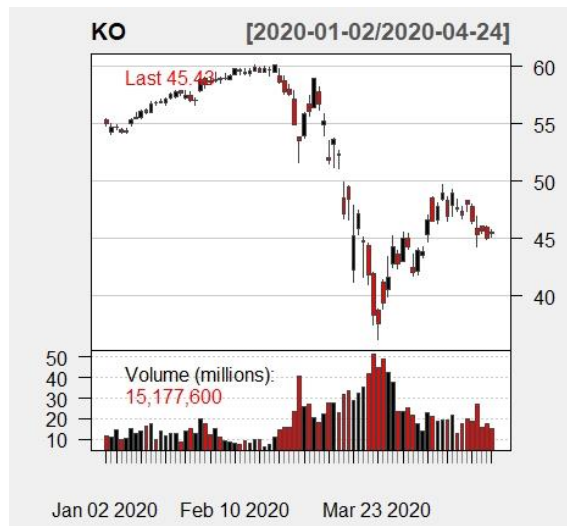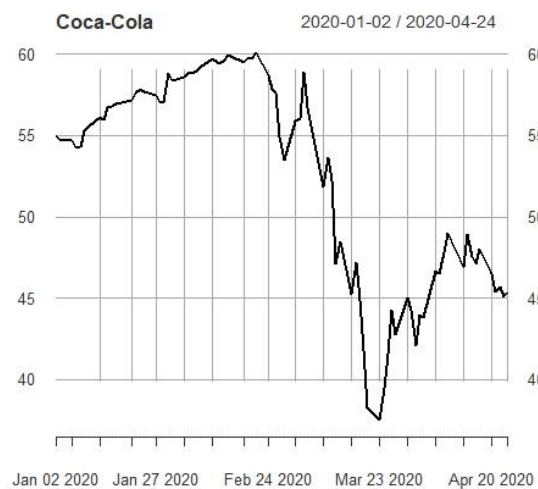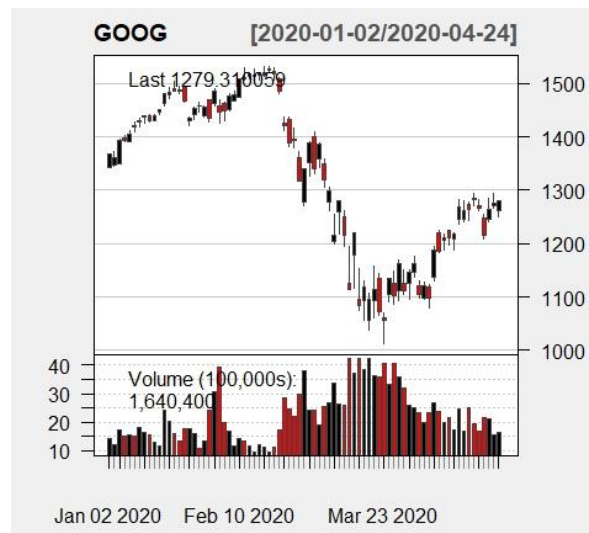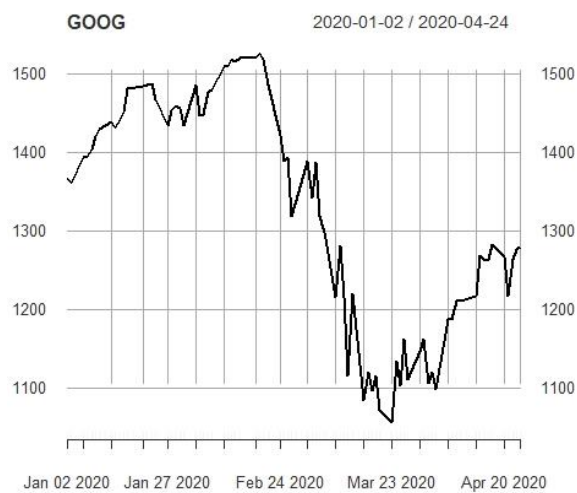
Figure 10 Stock chart in 2020

From the chart plots, we can see stock prices for all four companies suffered from huge decrease starting in the mid of February and recently begun to increase back after unlimited QE plans were established. Unfortunately, Delta Airline remains at low prices. It is reasonable because of working from home, the demands for flight are not back to normal status. Besides, we can see several red charts in candle chart plot, suggesting a dramatically decrease (possibly circuit breaker) of stock price on that day.

Another impact is unprecedented unemployment rate, which could be reflected by United States Unemployment Insurance Claim in Figure 11. Data comes from latest release report of Department of Labor (https://www.dol.gov/ui/data.pdf). On average, last year, the number of UI claims is 264 every week before April, 2020. However, as you can see in Figure 11, more than 6000 claims were filed during the unemployment peak.
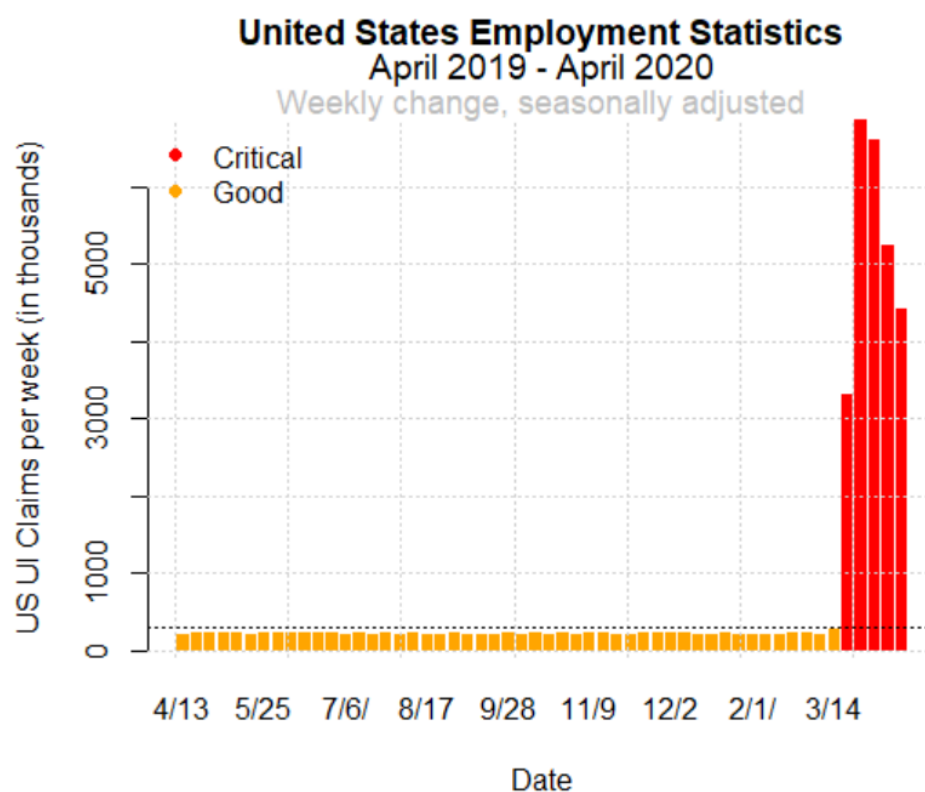

Figure 11 Unemployment insurance claim per week

Trend Analysis

We have investigated the linguistic variance on Twitter, but we would like to know whether it shares the same pattern on Google? What are the people most concerned about in terms of COVID-19? The Figure 12-14 show the answers.
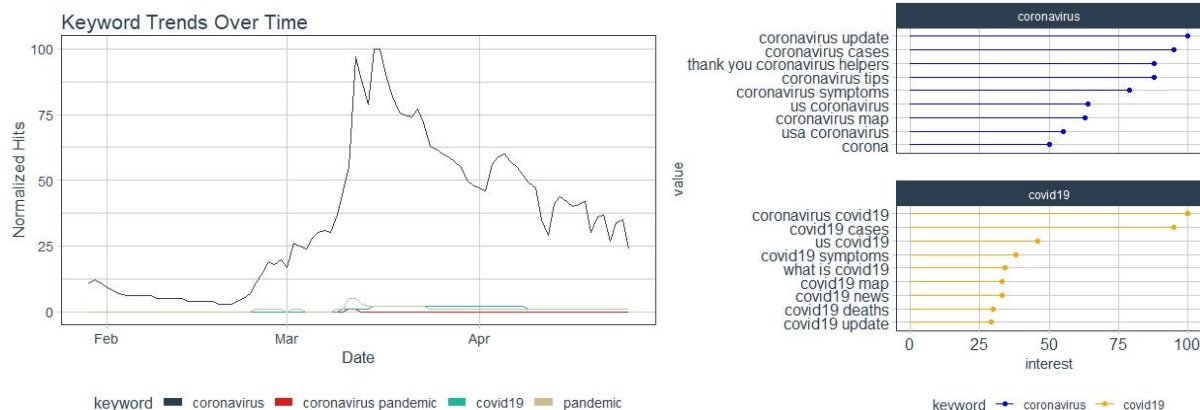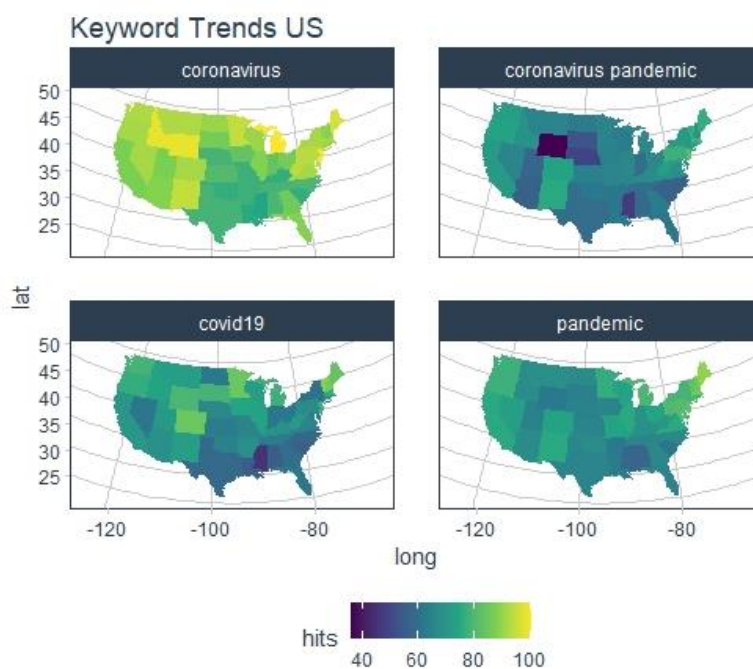
Figure 12 Keyword Trends Over Time

Figure 13 Geographical Keyword Trends in US

Differently, searching pattern on Google is more centered on coronavirus and relevant topics also suggest that Google users are more familiar with coronavirus while covid-19 not. Google users mainly search for coronavirus and covid19 for conditions updates, symptoms as well as tips to address COVID19. People in northwest, northeast, California, New Mexica and Wyoming tend to be more concerned with COVID19.

Summary

'Coronavirus' was originally widely used on Tweeter before WHO officially named it 'COVID19'. Recently, senators and House Representatives tend to equally use 'pandemic', 'coronavirus' and 'COVID19' on Twitter. However, people still tend to use coronavirus on Google search. Most governments around world took timely actions like public health measures, social distancing as well as public transportation restriction to stop the spread of COVID19 but lock down is less frequent adopted.

The public sit on the fence towards Trump's handling on COVID 19. Most people still hold in positive attitudes and trust the government while relatively less people hold negative attitude and fear COVID-19.

Stock market suffered from unprecedented circuit breaker in March 2020 and some companies' stock price increased back gradually after unlimited QE plans were publicized. However, some travel-related companies' stock prices stay in low price. Unemployment Status is also under its worst condition.

My Thoughts over Time

At beginning, I planned to make dashboard showing the outbreak spread over time. However, I found some good examples online and think it useless to replicate other people's work. So, I decided to make an animated photo instead. I also planned to predict the outcome of COVID19 with comparison of H1N1, but the fact that the number of COVID19 is much higher than H1N1 changes my mind. Moreover, I tried to parse data from Reddit, but I am unfamiliar with XML language and no such r package found useful.

The reason for me to investigate the reference for 2019-nCov comes from President Trump's deliberate use of 'Chinese virus'. Since I am afraid of it becoming a start of racism, I decided to look at whether other senators did the same thing. Fortunately, they did not. Moreover, I see the huge difference lying the attitudes towards to COVID19 between American and Chinese at the early stage, in which no one worn masks on the street in US. This makes me determined to check the public's attitudes towards coronavirus and which aspects the public care most.