COVID in Canada

Prabhsimran Singh, Anthony Quint, Guangyi Zhang

McGill university
Montreal, Canada

Introduction

Coronavirus(COVID) had an life changing effect on every one of us around the world (WHO 2019). The way in which people communicated and socialized, as well as their topics of discussion, totally changed. Once the COVID vaccinations were rolled out, different people reacted in different ways, some backed the vaccines while others questioned it. Therefore, to study how people reacted to COVID, we decided to perform an analysis of the topics discussed by the people related to COVID, including their concerns with the vaccination and the sentiment related to their views. The study presented in this report focuses on the views of the Canadian people only. Our multi-step analysis is shown in the Figure 1.

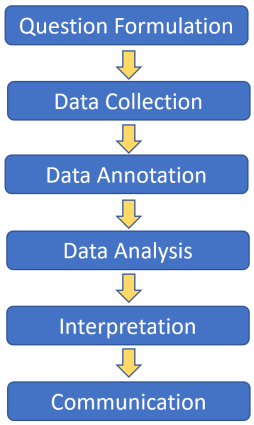

Figure 1: Project phases

For this project, the question of ***"What Canadian people think of COVID vaccination and how COVID Affected the Canadian Society ?"*** sums up what we want to discover from this study. For the data collection phase, we collected tweets related to COVID or COVID vaccination over 4 days from 15th Nov 2021 to 19th Nov 2021(Ben Huberman February). Then, data annotation was performed. During the data annotation phase, we classified eight topics discussed in the people's tweets involved in this study. We also distinguished the sentiment of each tweet whether it was of a positive, negative or neutral sentiment during annotation phase.

In the data analysis and interpretation phase, we found that a majority of tweets were related to vaccination and they were mostly positive. This showed us that a large amount of Canadians supported COVID vaccinations and they believed in it. Moreover, our analysis also found that almost every Canadian whose tweets related to spread of COVID in society were genuinely concerned for the rising COVID cases, the spread of virus and the fatalities due to it. Canadians were also very vocal about the failures of the government and political leaders in handling and controlling the COVID situation. Apart from this, people also tweeted about the Canadian healthcare systems, how COVID effected their health, news related to COVID and COVID restrictions. The detailed explanation related to the phases discussed in Figure 1 are presented in next sections of the report.

Keywords	COVID, COIVD-19, covid, covid-19, #COVID, #COIVD-19, #covid, #covid-19, coronavirus, vaccination, VACCINATION, #vaccination, #VACCINATION, Moderna, Pfizer, Janssen, AstraZeneca
Start time	2021-11-15T00:00:00.000Z
End time	2021-11-19T00:00:00.000Z
Max results	500
language	lang:en
Location	place_country:CA
Retweet	-is:retweet

Table 1: Query parameters

Data

For data collection phase, we collected 1000 English tweets related to COVID or COVID vaccination from the Canadian region using Twitter API v2(Twitter 2020) (O'Mahony 2021), excluding retweets. We got the twitter academic access which allowed us to filter and collect the tweets based on the location. To get the tweets we used the search query "`https://api.twitter.com/2/tweets/search/all`" with query parameters that contained keywords, location info, start time, end time and max results. The parameters used in the query are listed in Table 1.

For the keywords we used all the possible combinations we could think that would allow us to get the most relevant results. For one query request, twitter only returned a maximum of 500 tweets, therefore we collected tweets in 2 requests, one request from `2021-11-15T00:00:00.000Z` to `2021-11-17T00:00:00.000Z` and other request from `2021-11-17T13:00:00.000Z` to `2021-11-19T23:00:00.000Z` in Coordinated Universal Time (UTC). All the other parameters remained unchanged. Afterwards we merged the collected tweets in one file to get the final 1000 tweets. The query parameter `lang:en` allowed us to get the tweets with English language only. To get only the Canadian region tweets we used `place country:CA` parameter. And finally, to collect only original tweets and filter out the retweets we used `-is:retweet`.

Methods

In the data collection phase, raw tweets were analyzed to develop an appropriate typology. Each group member conducted open coding on the same set of 200 tweets to develop a preliminary categorization system. Next, each of our preliminary categories was interrogated and compared against each of the group members. A more adept categorization system was chosen such that it addressed the aforementioned research question, had an objective definition, and ensured that each category appeared in a roughly uniform distribution. To validate our new categories, each group member coded an additional 50 tweets; the typology was deemed satisfactory.

For the annotation phase, each group member annotated the full 1000 tweets. The final typology for a given tweet was chosen if at least $2/3$ of the group members agreed. Otherwise, a case-by-case approach was taken to investigate the tweets with a three-way split.

To analyze the data, the frequencies/proportions of each of the topics was computed, as well as their respective sentiment frequencies. Finally, a TF-IDF analysis was conducted to determine the 10 words in each category with the highest TF-IDF scores (REZGUI 2007). All words were separated by their punctuation, such that candidate words had to be exclusively alphabetic characters. The formula used is as follows:

$$\text{TF-IDF}(w, \text{topic}, \text{script}) = TF(w, topic) \times \text{IDF}(w, \text{script})$$

$$\text{TF}(w, \text{topic}) = \text{number of times given topic uses word } w$$

$$\text{IDF}(w, \text{script}) = \log\left[\frac{\text{total number of topics}}{\text{number of topics use word } w}\right]$$

where w refers to a word in a tweet and $topic$ refers to the category we defined. Whereas, $script$ refers to the whole data of 1000 tweets which was collected for this study.

To produce a more relevant TF-IDF analysis of the tweets, any word that matched a list of stop words (like a, an, and, the) was eliminated from consideration. Further, words that appeared infrequently (less than 5 times), were also eliminated. This helped ensure that the final subset of words was appropriately representative of the tweet data, yet unique enough to provide interesting discussions.

	Positive	Negative	Neutral	Total
Vaccination	130	79	15	224
Covid Restrictions	42	51	22	115
Political	10	74	9	93
Health Effect of Covid	16	65	12	93
News	9	45	122	176
Spread of Covid in Society	64	103	16	183
Covid Test Healthcare	41	43	13	97
Spam	5	7	7	19
Total	317	467	216	1000

Table 2: Frequencies of the categorization of tweets and corresponding sentiments

Results

This section defines the selected categories (or topics), their definitions, characterization and engagement.

Topic definition

1. Vaccination : This category includes all the tweets that were related to vaccination. To differentiate between the tweets that support or reject the vaccination, we used sentiment analysis. Every tweet that was positive towards the ability of vaccination to fight COVID, or that talked about vaccination in a good light, was marked as positive. The tweets that opposed vaccination or were pessimistic about vaccination were considered as tweets with negative sentiment. The tweets which were uncertain about vaccination or had mixed sentiments were considered neutral.

2. COVID Restrictions : This category includes tweets that discussed anything related to the COVID restrictions set up in society, such as a curfew, travel restrictions or proof of vaccination. In this case, positive sentiment referred to when people were in agreement with the restrictions. Contrarily, negative sentiment was defined when individuals were opposed to the restrictions; for example, if a person believes that the restriction is not necessary. Finally, a neutral sentiment defined a tweet that was impartial about COVID restrictions.

3. Political : The tweets in which Canadians expressed their thoughts and opinions about the government, political leaders or any political action related to COVID were categorized as political. This included tweets that talked about advancing political agenda, people blaming or congratulating leaders for COVID response or opinions about steps taken by the government related to COVID. Tweets which praised the political actions to COVID response were defined as positive whereas tweets condemning those actions were annotated as negative. Tweets which neither praised or condemned the actions were considered as neutral.

4. Health Effect of COVID : This category includes tweets that discussed the physiological impact that the COVID virus has on the human body, such as if somebody men-

	Vaccination	COVID Restrictions	Political	Health Effect of Covid	News	Spread of Covid in Society	COVID Test Healthcare	Spam
1	received	qr	workers	infection	india	freedom	air	immune
2	choice	university	treatments	fine	update	air	book	reported
3	vaccinate	travel	hospitals	im	cases	vaccinate	flight	season
4	imagine	tonight	zero	complications	nova	premier	wait	reactions
5	shot	restrictions	continue	biggest	reported	infect	kid	olds
6	vaccination	flight	money	lack	zone	mask	tested	wait
7	moderna	christmas	economy	wearing	active	sick	tests	stopped
8	pfizer	masks	save	underlying	brunswick	choice	pcr	probably
9	approved	required	alberta	cold	scotia	outbreak	test	adverse
10	booster	proof	meeting	pain	november	distance	rapid	yeg

Table 3: TF-IDF result

tioned how sick they felt, or if they were asymptomatic. Tweets were labelled with positive sentiment if they mentioned that the effects of COVID were mild, while the tweets which indicated that severe effects of COVID were labelled with negative sentiment. Lastly, a neutral sentiment was used if the tweet was impartial to the health effects of COVID.

5. News : The news category included tweets that were published directly from a news source, or from an individual journalists. News tweets were labelled with neutral sentiment if they offered impartial information/data, i.e., they were not emotionally charged. In the cases where they were emotionally charged, such as in OP-ED (opinion) articles, a positive or negative sentiment was used to clarify this.

6. Concerned about Spread of COVID in Society : This category defined tweets that discussed the concerns regarding the spread of COVID in society. In this case, positive sentiment referred to tweets that were not concerned with the spread of COVID (for example: tweets discussing the benefits of remote working), while negative sentiment referred to tweets that are worried. Neutral sentiment is used when a tweet is impartial about the spread of COVID.

7. COVID Test / Healthcare : This category was related to tweets about COVID tests, test results, clinics and hospitals. If people praised the current medical facility and found the COVID test arrangements sufficient, then that was considered as positive sentiment. In contrast, if people complained that it was not convenient to take COVID test or the test was not reliable or healthcare facility is not good, then this was a negative sentiment. If people just tweeted about where and how to take a COVID test, the estimated time to get the result, this was considered as a neutral sentiment.

8. Spam : Every tweet that was unrelated to COVID or the COVID vaccination was categorized as spam, e.g. a tweet that used a *#COVID* but it is about something else. Therefore, in order to check their sentiment, we focused on the tone of words they used. If a tweet contained enthusiastic, happy or excited words, it was considered positive whereas it was considered negative if it contained hateful, bad words. In the end, if a tweet is missing sentiment then it was labelled as a neutral sentiment tweet.

Topic characterization and engagement

As we obtained the dataset from Twitter, we selected 1000 tweets in total and conduct in-depth analysis to make a deeper understanding of the data. As illustrated in Table 2, we summarized the content of the tweets according to 8 topics, including vaccination, COVID restrictions, political, health effect of COVID, news, spread of COVID in society, COVID test healthcare, and spam. Furthermore, we categorized the content into “positive”, “negative”, and “neutral” and would like to make a better understanding of the sentiments of every tweets.

Table 2 summarizes that a majority of tweets were related to vaccination (224) followed by Spread of COVID in Society (183) whereas other topics were almost distributed uniformly($\approx$100). In terms of sentiments most of tweets (467) were of negative sentiment, followed by positive sentiment tweets(317).

For a more straightforward understanding of the data, we decided to visualized it. Firstly, we plotted a bar graph containing numbers of tweets with different sentiments in each category, as shown in Figure 2. Most people showed positive sentiment towards the vaccination whereas for political, spread of COVID, health effect of COVID people were mostly negative. News had majority of Neutral sentiment whereas the other topics showed a mix of all sentiments.

Next, we made a pie chart Figure 3 to analyse the topic engagement by showing the percentage distribution of every topic. We observed that vaccination was the hottest topic with 22.4%, followed by spread of COVID in society with 18.3% and News with 17.6%. The spam constituted a tiny 2.9% among the rest.

Lastly, a TF-IDF analysis was completed to develop a stronger understanding of the content of the categorized tweets; the results are displayed in Table 3, and are discussed further in the discussion section. The specific words provide context on the themes of discussion within each category, and indicate the primary area’s of people’s concerns. For instance, a popular word that appeared under Vaccination and Spread of COVID in Society is ‘choice’, which emphasized

the conflict between freedom of choice and public health safety which many people had.

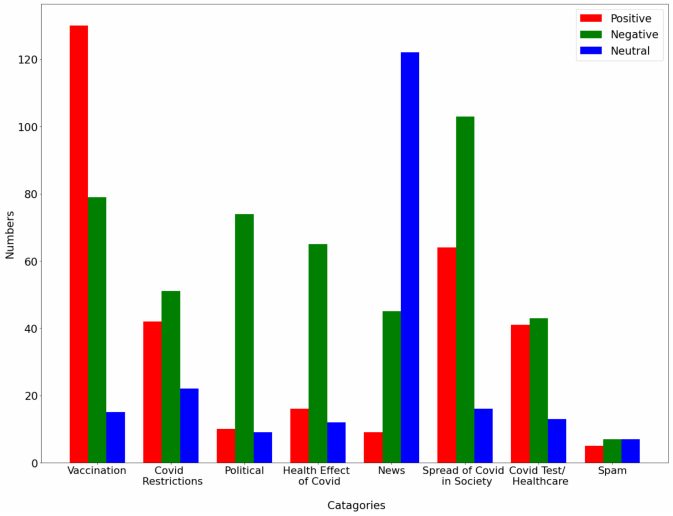

Figure 2: Frequencies vs. Topics sentiments

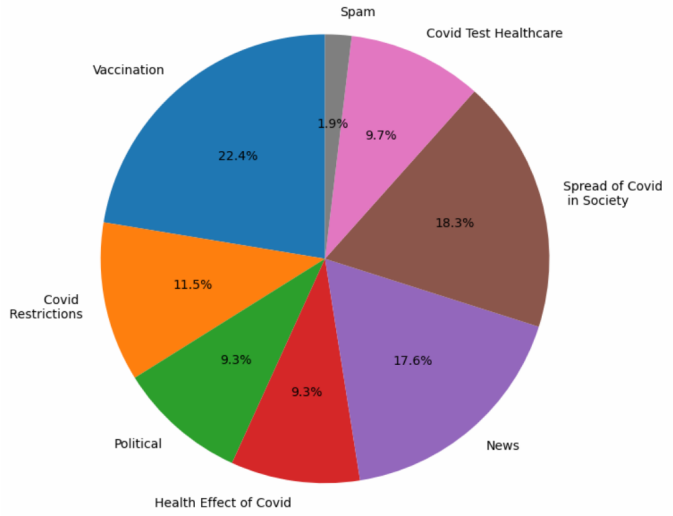

Figure 3: Proportions of the Tweet Categories

Discussion

Since it's conception in late 2019, COVID-19 was declared as a long-term global pandemic. COVID has dominated people's topics of conversations and their concerns. This motivated the analysis of tweets as a proxy for understanding Canadians opinions on COVID and how COVID impacted their lives.

In many countries, imposing restrictions is treated as an effective method to cut down the spread of COVID. An typical example was in Wuhan, China. Wuhan was locked down from January 23, 2020 to April 8, 2020 with 76 days in duration. Despite the general lock down in China was painful and detrimental to local economics, people's daily life has come back to normal since April 2020 whereas, some countries are still recovering from the loses they faced due to COVID restrictions. Many countries imitated this action and imposed COVID restrictions to try to reduce the spread of COVID. It was reasonable that 11.5% of the tweets are coming from the topic of COVID restrictions. This is because while some people believe it was an effective method to reduce the spread of COVID, others may think that COVID restrictions restrict their freedom. Government imposed mask mandates are common restriction. One example tweet from our data set is: "*wish i died of covid i hate the vax mask world*". This tweet is a typical negative sentiment, in that the person was trying to protest the mandatory mask order and the COVID restrictions. Travel restriction was also widely discussed. One example tweet was:"*Travel planning is much more bewildering this Christmas... Looking forward to seeing my parents for the first time in over two years*";this is an neutral sentiment. Evidently, the TF-IDF analysis in Table 3 affirms that many of the common points of discussion are around 'travel', and 'masks'. Further, the TF-IDF analysis suggested that people discussed 'proof's of vaccination ('qr' codes), how restrictions affected their 'university' studies, and how restrictions may affect their holiday plans like for 'Christmas'.

If we look at the number of positive sentiment tweets related to COVID restrictions we can argue that some Canadians believe that COVID restrictions will probably be required in the short term, to reduce transmission and fix ineffective pandemic response systems, in order to prevent future lockdowns. The goal of these restrictions is to effectively suppress COVID infections to low levels, allowing for early detection of localized outbreaks and prompt response through efficient and comprehensive find, test, trace, isolate, and support systems, allowing life to return to near-normal without the need for broad restrictions. These thoughts are very similar to current government of Canada guidelines (Canada 2020).To mention a few nations, Japan, Vietnam, and New Zealand have demonstrated that strong public health interventions can restrict transmission and allow life to return to normal, and there are several other success stories(Alwan et al. 2020). However, for USA, by relaxing the restrictions in a rush has incurred a dramatic increase in COVID cases, especially in late 2020(CDC 2020b).It seems to be necessary to relax the restriction to recover the economy when the local COVID situation is somehow controllable.

To deal with COVID, many countries tried their best in developing useful vaccines to fight against COVID. Starting at the beginning of 2021, COVID vaccinations were gradually made available to the general public. It was evident that vaccination made up the majority of the tweets in our analysis, about 224 tweets out of 1000. This suggested that vaccination was a strong point of conversation, moreover in that there was a strong divide between people for or against vaccination. The preceding sentiment analysis determined that the majority of people are pro-vaccination, which is likely due to the efficacy of previous vaccines, government and healthcare recommendations, and people's desire

to feel safe (CDC 2020a). Despite this, a significant portion of the vaccination tweets had a negative sentiment. One example tweet was as follows: *"Can Anything About the Pfizer Vaccine Trial be Trusted?"*. This tweet illustrates a common trend with the negative vaccination tweets, in that they question the legitimacy of the vaccine as an effective health measure, and suggest that there are other incentives at play. This idealogy is not entirely unfounded, as controversies have occurred in the past that question the healthcare system/government as a reliable source. One infamous case is the Tuskegee Syphilis Study, which was an ethically abusive study conducted between 1932 and 1972 by the United States Public Health Service and the Centers for Disease Control and Prevention on a group of nearly 400 African Americans with syphilis (Tus 2021). In this trial, no treatment was provided to the individuals despite the common use of penicillin at the time. Another infamous example, which sparked the anti-vax movement in the first place, was the UK doctor who falsely published a paper stating that vaccinated babies are more likely to develop autism (Wakefield A 1998).This paper was later retracted. These examples highlight the dangers of government misconduct and of misinformation when it comes to public health emergencies. Referring to the TF-IDF analysis in Table 3,the common buzzwords related to the process of vaccination, such as 'pfizer' and 'moderna', which are the prominent vaccine producers, and 'recieved' and 'booster', which referred to individuals receiving their vaccines and also obtaining a booster shot dose. Most interestingly, one of the words was 'choice'. This relates to the aforementioned discussion on the conflicting nature of the vaccine, in that people may choose to get vaccinated to protect themselves, or people may reject the vaccine because it is within their freedom to act as they wish based on their beliefs.

COVID test and healthcare is another important area in this pandemic. The public health sector enunciated the importance of conducting COVID tests to verify whether a person is infected or not. Therefore the Canadian government had arranged many COVID testing spots, vaccination clinics within the reach of people. Moreover, hospitals played a major role in treating COVID patients and saving a lot of lives. Because a vast majority of people relied on hospitals, clinics, COVID testing centers and vaccination centers (sur 2020), the tweets of this category comprised 9.7% . One example tweet was:*"To help with the backlog caused by COVID-19, our government has invested over $100,000 to ensure building a #HealthCare system to support the needs of all Ontarians."* this is a positive sentiment, as it talks about the steps taken by government to strengthen the healthcare system. However, one negative sentiment example was:*"Any other Canadians going to #MTGVegas please DM me, am tryin.g to. work out getting covid test appt and not finding any easy time slot to do it."*. Although the COVID test is beneficial to all of us, limited testing resource cannot fulfill the testing needs. This is reflected in some of the public's discourse with the healthcare system, as they demand government an increase in COVID testing and healthcare resources. According to the TF-IDF result in Table 3, words 'pcr','rapid' and 'test' showed that people talked about the COVID test categories. The words 'air','flight', in this category suggested that people made or discussed COVID test prior to their travel, one possible reason of this can be because of some countries requiring negative COVID testing result upon arrival. A possible explanation for the 'wait' and 'book' keyword can be, it showed that the current testing resources were still limited to fulfil the testing need of Canadians and they had to wait for test or results and were also worried about time taken to book an appointment to have a test.

COVID has become a global pandemic, and the spread of COVID has provided many challenges. A majority of people were more concerned about the spread of COVID in the society, and how COVID was impacting their lives. This was reflected in our data, where tweets regarding the spread of COVID in society comprised the largest proportion and magnitude of negative sentiments. For example, *"People are walking out of the COVID testing centre and removing their masks before they get outside. AND THIS IS WHY WE ARE NEVER GETTING OUT OF THIS MESS"*, this was a typical negative sentiment towards COVID as person reacted with angry and hate. However, not all people approached COVID with the same pessimism. For example:*"PEOPLE stop wearing masks!!! I want work from home"*, this was a positive sentiment. Many positive tweets emphasized social distancing, wearing masks, getting vaccination and suggested following safety protocols to protect the community. Indeed, the spread of COVID had brought us much panic and made our lives difficult. However, it is still possible to reduce the spread by safety measures. The TF-IDF results in Table 3 suggested that people talked about risk of 'infection' and getting 'sick', which could be caused by 'outbreaks'. Social 'distancing' was a common suggestion to try and mitigate this. Interestingly, a popular theme in this category relates to the words: 'freedom' and 'choice'. Evidently, the freedom of choice was a major factor in the spread of COVID; as determined if people choose to behave in a manner recommended by health authorities, or whether they believe and act otherwise.

Discussions around COVID has presided over the news every day since 2019. The news mostly reported the updates on number of new cases, and what actions are given. For example, *"Nova Scotia detected 31 new COVID-19 cases/253 active cases/16 in hospital, 7 of whom are in ICU"*. Also,*"Health Canada receives submission to authorize Moderna's COVID-19 vaccine for kids"*. These are two typical neutral sentiments. News are generally informative, it let people know what is happening around them. According to Figure 2, majority of the news tweets were neutral. During the COVID situation, the news has become a key method to inform people about their relative risks of danger from COVID, and whether they will need to take more actions to protect themselves. According to the TF-IDF result in Table 3, tweets from news sources were predominately providing 'update's about COVID 'cases' that were 'reported' in particular locations, such as in 'India', 'Nova Scotia', and in New 'Brunswick'. These were likely discussing outbreaks; for example, *"India has recently seen a large surge in positive cases."*

Concluding Remarks

Overall, Canadian people's response to COVID vaccines was mostly positive. This showed us that a large amount of Canadians supported the COVID vaccinations and they believed in it. Canadians demonstrated a strong concern for the spread of COVID in society, where they desired the safety and well-being of everyone involved. Further, Canadians actively voiced their expectations for political and governmental interventions when it came to appropriately and effectively handle the pandemic; and they were critical when measures were inadequate. Apart from this, people also tweeted about Canadian healthcare systems, how COVID effected their health, news related to COVID and COVID restrictions.

One interesting theme in this discussion was one about freedom and choice. While it appeared that most of Canadians were pro-vax and believed in COVID as a legitimate pandemic, there was a sub population of Canadians who believed their freedom was being restricted, and who felt that measures have gone too far. Overall, in a democratic society, there will always be two sides to an argument, but it appears that most of Canadians genuinely are in the fight against the pandemic.

Group Member contributions

At the beginning of the project, the work was partitioned such that each group member would have an equal amount of work to complete. All the group members participated in weekly group meetings where we discussed future plans, and the progression of the data collection, results, and the writing of final paper.

Firstly, Prabhsimran was responsible for the data collection phase. He accessed Twitter's API, obtained academic permission to download tweets, and then obtained the tweets that satisfied our inclusion and exclusion criteria. He cleaned and prepared the data such that it was ready for the annotation phase.

The annotation phase was equally completed by all group members. We each dummy coded a small portion of the tweets in order to decide our final categorization system, and then we each fully annotated the full 1000 tweets. Prabhsimran then wrote a script which provided a brief analysis of the annotated data.

Next, Anthony was responsible for the TF-IDF analysis. He wrote the scripts which calculated the word frequencies, and then determined the best criteria and equations for the TF-IDF scores to be calculated on. Then the TF-IDF scores were calculated with another script, which saved the top 10 highest scored words in each category.

Finally, while all group members participated in writing the final paper, Guangyi took an emphasized role, where he was also responsible for formatting the paper, producing the figures and tables, and also completing the bibliography which is included at the end of the paper.

Overall, all group members communicated efficiently, completed their tasks in stride, and acted in a professional and academic manner.

References

2020. "Worst surge we've seen": Some hospitals in delta hot spots close to breaking point. https://www.aamc.org/news-insights/worst-surge-we-ve-seen-some-hospitals-delta-hot-spots-close-breaking-point.

2021. Tuskegee Study of Untreated Syphilis in the Negro Male. https://en.wikipedia.org/w/index.php?title=Tuskegee_Syphilis_Study&oldid=1056017065.

Alwan, N. A.; Burgess, R. A.; Ashworth, S.; Beale, R.; Bhadelia, N.; Bogaert, D.; Dowd, J.; Eckerle, I.; Goldman, L. R.; Greenhalgh, T.; Gurdasani, D.; Hamdy, A.; Hanage, W. P.; Hodcroft, E. B.; Hyde, Z.; Kellam, P.; Kelly-Irving, M.; Krammer, F.; Lipsitch, M.; McNally, A.; McKee, M.; Nouri, A.; Pimenta, D.; Priesemann, V.; Rutter, H.; Silver, J.; Sridhar, D.; Swanton, C.; Walensky, R. P.; Yamey, G.; and Ziauddeen, H. 2020. Scientific consensus on the COVID-19 pandemic: we need to act now. *The Lancet*, 396(10260): e71–e72.

Ben Huberman. February. How Do Data Scientists Use Twitter? Let Us Count the Ways. https://towardsdatascience.com/how-do-data-scientists-use-twitter-let-us-count-the-ways-50494e2a95c8.

Canada, P. H. A. o. 2020. Coronavirus disease (COVID-19): Prevention and risks. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks.html.

CDC. 2020a. COVID-19 Vaccination. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/How-Do-I-Get-a-COVID-19-Vaccine.html.

CDC. 2020b. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker.

O'Mahony, L. 2021. Getting Started with Data Collection using Twitter API v2 in less than an hour. https://towardsdatascience.com/getting-started-with-data-collection-using-twitter-api-v2-in-less-than-an-hour-600fbd5b5558.

REZGUI, Y. 2007. Text-based domain ontology building using Tf-Idf and metric clusters techniques. *The Knowledge Engineering Review*, 22(4): 379–403.

Twitter. 2020. Twitter API. https://developer.twitter.com/en/docs/twitter-api.

Wakefield A, A. A. e. a., Murch S. 1998. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*.

WHO. 2019. Coronavirus disease (COVID-19). https://www.who.int/health-topics/coronavirus#tab=tab_1. Accessed: 2021-12-10.