2021-06-15

Understanding Quote-Tweet Usage During the COVID-19 Pandemic

David Hardy Bean
Brigham Young University

Understanding Quote-Tweet Usage

During the COVID-19 Pandemic

David Hardy Bean

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Amanda L. Hughes, Chair
Derek L. Hansen
Justin S. Giboney

School of Technology

Brigham Young University

ABSTRACT

Understanding Quote-Tweet Usage
During the COVID-19 Pandemic

David Hardy Bean
School of Technology, BYU
Master of Science

The COVID-19 pandemic first entered the international news cycle with mixed levels of concern. How did people across the globe react to first encounters with this virus? For many it was like seeing for the first time the spewing ash of a volcano, or the receding tides of a tsunami. Many reacted with disbelief, not knowing the proper course of action for themselves or for their community.

This study explores the topics of discussion and reactions to the pandemic through the lens of quote-tweets—from the initial confusion and disbelief to the eventual politicization, economic closures, and reopening events. Quote-tweets are reactions to other tweets. This makes them idea to study opinions on various topics. If a tweet covers something interesting, often a quote-tweet will follow, displaying a reaction message tacked under the original message. This generates discussion about the topic in the original tweet.

We gathered tweets from five of the first months of the pandemic and found several trends. Early on, collected quote-tweets were much more likely to discuss health-related topics such as symptoms, demographic information, or death. Conspiracy theories and disinformation also abounded during this time. Quote-tweet reactions were often short, simple, and expressing disbelief. Quote-tweets made up between 30 and 40 percent of all tweets streamed from twitter using search terms of Coronavirus and COVID-19.

Later in our collected data, quote-tweets discussing the economy in relation to COVID-19 began to appear. They also grew more critical and political, often directing criticisms toward local or foreign government officials. Quote-tweet reactions followed suit and more often expressed criticisms and opinions of their own. Both agreement and disagreement increased over time.

Although disasters often generate political debate, online discussion about the COVID-19 pandemic shifted dramatically over the course of this study. The trends of topics and opinions that make up these online discussions via quote-tweets and original tweets can inform health and emergency officials on trends to be found in pandemics and disasters to come.

Keywords: twitter, disaster, COVID-19, response, APA

ACKNOWLEDGEMENTS

First, praise and thanks be given to God, who giveth to all liberally and upbraideth not, especially to those who ask. Second, praise be given unto those who lack wisdom and ask—even that certain lawyer who asked, "and who is my neighbor?". May our research of digital communications help reveal new insights to this fundamental question.

I'm grateful to Lucia Mata, who forged a path to access hundreds of non-English tweets within this research, as well as executed the bulk of the English coding.

I'm grateful to the School of Technology, which not only funded a large part of my course work but has taught me basically everything I know about the world of computers in the last six years.

Lastly, I'm grateful to my committee chair professor Amanda Hughes, who pioneered much research in the space of digital communications during disasters, and provided me with vital direction, confidence, and review.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 The Role of Social Media During a Disaster

Social media platforms can be a reliable place for those affected by disasters to find and share information (Palen and Hughes, 2018; Reuter and Hughes, 2018). These platforms host an ever-growing body of knowledge that can be useful to emergency responders. For example, Acar and Muraki (2011) and many other researchers have found patterns in disaster related posts on social media that can help with identifying victim needs, assessing damage, finding misinformation, etc., resulting in a clearer image of what happens on the ground during a disaster. However, there is still much that we do not know about how the features of different social media platforms support disaster communication (Palen and Hughes, 2018; Reuter and Hughes, 2018) and how these features might help emergency responders extract useful information. Knowledge of how social media features affect online discourse can help people like first responders and researchers to find better ways of filtering and finding the information they are seeking during a disaster.

Emergency responders are important to consider when attempting to uncover new information within social media during a disaster because their work is so closely tied to public health and what is happening in disaster-affected communities. For example, in 2020, response teams within the US government were very interested in learning about the spread of COVID-19. In March of 2020, there was a shortage of tests, making the true spread of the virus impossible to predict. The Center for Disease Control (CDC) took to other methods of learning about what was happening on the ground as well as communicating vital information to the public (https://www.cdc.gov/socialmedia/tools/Twitter.html). With the right filters in place, organizations like the CDC could find detailed data about how diseases are spreading within small communities. The ultimate goal of this and other disaster related research is to improve the ability of first responders and organizations like the CDC to find useful information through social media. Much research has found interesting trends and methods of filtering data within social media (Palen and Hughes, 2018; Reuter and Hughes, 2018), but unless this research is mapped back to a specific objective within disaster response then the responders' ability to react to and manage disasters within a community isn't improved.

1.2 The Role of Twitter During a Disaster

This research seeks information derived from very specific features within a particular social media platform, Twitter. The users of these features are only a small fraction of the population that disaster responders intend to help, but data derived from these features can inform behavior for a broader audience.

Twitter, a commonly used platforms in disaster, has several qualities that make it a great candidate for research within disaster response. First, posts on Twitter tend to spread more quickly than on other platforms. Many organizations don't have the consumer network infrastructure to spread news and make announcements on their own, so they rely on Twitter to get the word out. For example, many public transit systems will announce delays via Twitter. Second, Twitter has a public, well-documented, API which has made it a popular subject for many previous studies. Lastly, like many social media platforms, Twitter evolves with public discussion. On any given day the most widely discussed topics are current events such as the Coronavirus pandemic which became our focus for this study.

Twitter has evolved to host unique classes of interactions, such as reply-tweets and quote-tweets. These types of tweets broadly reflect the verb they are named after. Reply-tweets are tweets generally directed at the writer of another tweet (a reply to the original tweet), but sometimes are used as a continuation of a series of tweets from a single author forming a story. Reply-tweets were introduced in Twitter's initial release. Quote-tweets are tweets that reference another tweet's text and contain a comment that is usually about that text. Quote-tweets were introduced in early 2015. It's possible, but uncommon, for a tweet to be a hybrid of these two types by quoting one tweet and replying to another. Figure 1-2 contains a screenshot example of each of these interactions.

While both types of tweets are interesting, quote-tweets are the focus of this research. Quote-tweets are the newest class of tweet with few previously published studies singling them out (Garimella et al, 2015). Quote-tweets are also more common. When sorting our collected tweets about COVID-19, our quote-tweet collection gathered 10 times the number of tweets gathered by the reply-tweet collection. Despite the interface displaying the button for each of these interactions with equal prominence, it seemed that either quote-tweets are a very popular interaction, or the Twitter streaming API favored quote-tweets in its responses.

This research seeks to understand how quote-tweets, along with their self-supplied original tweets, communicated about the COVID-19 pandemic. We want to understand the patterns found in quote-tweets and how these patterns changed over time and determine whether this information can be used by health officials or emergency responders. COVID-19, unlike natural disasters and other pandemics in recent history, was a global, lifechanging event drastically impacting social and economic activity nearly everywhere. Though the event itself was unique, the communication patterns we discover can likely be applied to a variety of crisis events and the national and international level.

Figure 1-2 Left: Critic Uses Reply-Tweet to Respond to President Trump. Right: User Uses Quote-Tweet as Citation for Remarks About Senior Trump Officials

The quote-interaction type is important to gain a better understanding of the general discourse of Twitter. In the preliminary research for this study, it was found that a third of the generated tweets in a given time period were of the quote-tweet type. This represents a large number of tweets about which little is known.

Being a relatively new feature, quote-tweets have not been studied much in—or outside of—the context of disasters, yet some early researchers produced some exemplary work. Garimella et al. (2015) studied samples of quote-tweets and reply-tweets in the context of U.S politics and found that Twitter users tended to use higher sentiment polarity when reply-tweeting than quote-tweeting, meaning reply-tweets contained diction reflecting more emotions such as anger and gratitude than quote-tweets. This was probably due to the way the structures are presented on the user interface. Reply-tweets don't appear within the timeline of the author, while quote-tweets do. This means that reply-tweets are easier to forget and more hidden from view to the author's followers. This additional anonymity seems to allow authors to express their emotions more freely within reply-tweets.

So why are quote-tweets so much more popular than reply-tweets which provide this additional freedom of expression? We don't know with certainty, but it might be that users want to broadcast information found in their quote-tweets. After all, nobody wants to log onto social media to write posts that no one will read. While reply-tweets contain more emotional language, maybe there's something in quote-tweets that makes them just as genuine. Analysis of quote-tweets written during the pandemic may reveal what it is about quote-tweets that makes them the more common way to express opinions and reactions toward topics found on Twitter.

In this research we compare quote-tweets with the tweets they respond to, which we also refer to as original tweets. These original tweets are not initiated in connection with other users, but are comprised of text, photos, video, or links to broadcast to the posting user's followers. The examples of reply-tweets, (Figure 1-2) were tweeted in conjunction with a regular tweet which contained text and an embedded news article link. By studying the relationship between quote-tweets and embedded original tweets, we hope to find patterns of behavior that are useful from the perspective of health officials and emergency responders. These patterns might provide a foundation for future research to identify individual needs that health officials can address or reveal that quote-tweets seldom reflect these needs and they could be filtered from view to the benefit of emergency responders.

Quote-tweets are important constructs within Twitter. Together, they host the bulk of person-to-person communication on Twitter's platform. Yet they also encourage and support different types of communication, observed by Garimella et al. (2015), that have not been studied in the context of a pandemic. In this research, we will explore the relationship between quote-tweets and their corresponding original tweets, so we can better understand how this form of communication is used during a pandemic and how it has changed over time. It is expected that this understanding will inform strategies for health officials or emergency responders to more easily find tweets posted by those affected by a pandemic, filter irrelevant tweets, and/or post tweets that are more likely to help. The research could also be valuable to Twitter itself because it will inform platform engineers of communication structures that enable clarity, visibility, ease of use, and other desirable outcomes. COVID-19 wasn't an ordinary disaster event. Its once-in-a-lifetime nature invoked patterns of communication on Twitter and elsewhere that might not apply directly to other events, disasters, or pandemics in the near future. Instead, it gave a glimpse of how an unprepared world responds to change.

2 LITERATURE REVIEW

This literature review outlines a broad perspective of existing disaster communication research. It also describes the research, unrelated to disaster communication, that has been done on quote-tweets and reply-tweets.

Garimella et al. (2015) were some of the first researchers to study the communication outcomes of quote-tweets in comparison with reply-tweets. This was because quote-tweets were a feature introduced in early 2015. They found that negative sentiment was more common in replies than in quote-tweets among a sample of political tweets gathered in 2015. They also found that quote-tweets tended to have further reach within user networks, probably partially due to the mechanic that records quote-tweets and not reply tweets in a user's timeline. This research showed that the differing post structure between quote-tweets and reply-tweets resulted in different discourse outcomes. These results might partially be due to a limitation, in that quote-tweets were a new feature when the research was conducted. Early adopters of quote-tweets were long-time Twitter users who had more experience with the platform. Reply-tweet users may have had a significantly higher number of first-time posters lacking the etiquette needed to raise the average sentiment polarity. A similar explanation could describe the larger quote-tweet user network reach.

The age-gap between now, 2021, and research of Garimella et al in 2015 should be enough to justify revisiting quote-tweet usage patterns. Now that quote-tweets are a longstanding addition to Twitter posting structure, behavior may have changed and leveled out from the time when quote-tweets were new. Social cognition theory suggests that most learned behavior comes from users mimicking others' behavior (Bandura, 1986). If this is the case, then quote-tweets may have been too new at the time of the study to gain an accurate perspective about how they differed from older structures like reply tweets. Many of the users were mimicking the behavior of the first adopters while usage innovation had not yet leveled out. Given that Garimella et al.'s (2015) research took place when the feature was less than one year old, it may be that quote-tweets and reply-tweets have matured to reflect different user behavior.

Our research adopted a similar methodology to Garimella et al. (2015) using both natural language processing and human coding to map features of sentiment and observed use case to each tweet object. We also focused on the relationship between quote-tweet and original-tweet on a deeper level. Rather than recording the quote-tweet behaviors noted by Garimella et al.: forwarding, public replies, and whether opinion existed; we instead developed a methodology that focused on the content of the quoting user's opinion along with the overall reaction toward the original tweet.

In other research, Acar and Muraki (2011) found that before, during, and after a tsunami the content of the tweets changed. Content also varied greatly between geographical regions. They concluded that this was because of differing user population priorities surfacing as a result of the disaster. This research outlines the differences in victim tweet use cases, and maps these use cases over time and region. This research, along with other research conducted on Twitter usage during a disaster (Palen, L., & Hughes, A. L. (2018); Reuter, C., Hughes, A. L., & Kaufhold, M.-A. (2018); Hughes, A. L. (2019), paints a picture of how those affected by a disaster behave online. We aim to expand this knowledge by discovering user behavior during a long-term, worldwide pandemic. Rather than looking at the differences in user behavior between two cities over a period of weeks, our research focused on global differences over a period of months. COVID-19 hosted many different factors not found in a tsunami, including a timeframe and world-wide demographic not often found in other studies measuring how tweets change across differing regions and times. Takahashi and colleagues (2015) conducted a similar study, part of which measured how tweets changed during the lifetime of a hurricane in the Philippines. They found that mentions of relief efforts increased as time progressed, which made sense given that the first tweets were gathered before the storm had reached the coastline.

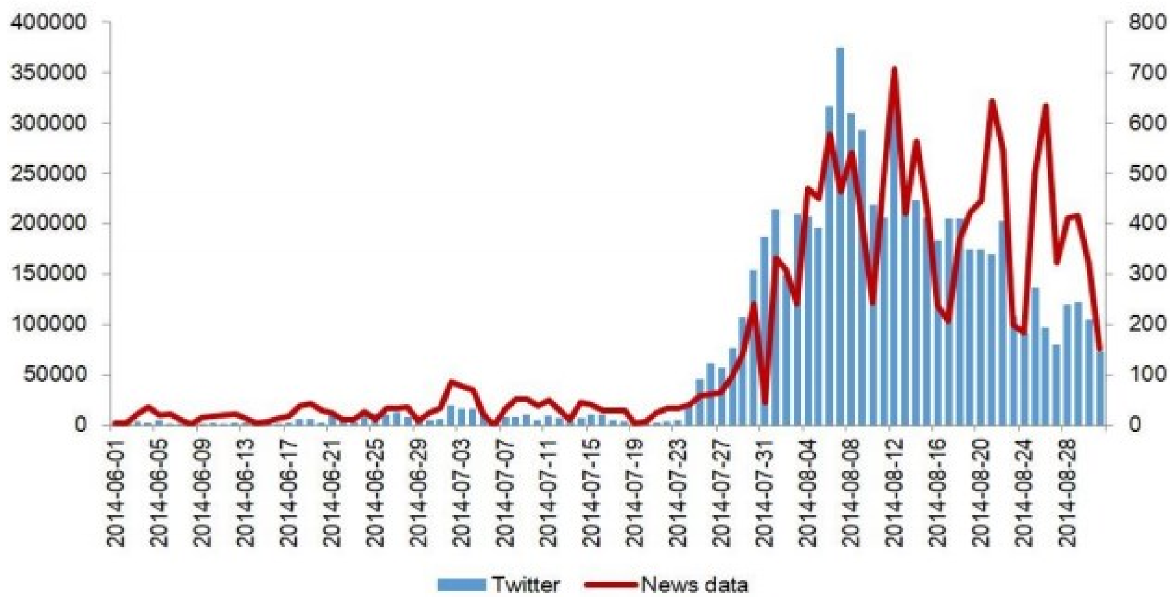

Figure 2-1: Kim et al (2016) Compared Tweet Numbers (Seen Y Axis Left) with News Media Mentions (Seen Y Axis Right) Regarding the Ebola Outbreak of 2014

COVID-19 was a pandemic, not a natural disaster. One of the closest resemblances in recent history is the Ebola outbreak of 2014, which gathered some interest on Twitter and news feeds when it occurred. Kim et al (2016), studied sentiment and content of tweets covering this outbreak. The bulk of the tweets gathered in this study were gathered during the month of July

2014, which can be seen in Figure 2-1, which shows the frequency of mentions of Ebola within Twitter and news media for part of 2014.

Kim et al (2016) measured the outbreak conversations in terms of sentiment and keyword occurrences on Twitter in comparison with traditional news media. Our research differed from this in focus. We sought to measure topics from a disaster response perspective. We sought to define themes for each window of tweets gathered while comparing each set with its predecessor. The Ebola outbreak presented a set of tweets larger than most natural disasters, but the overall world impact and timeframe of Ebola are small in comparison to that of COVID-19.

How do quote-tweets add to the observations of each of these studies? They allow us to capture reactions, how communication is received, about the same kinds of posts that were found in these previous works. Most research, including the studies cited thus far, gathered original tweets which in our research is the source of *topics.* But in addition to *topic,* our research categorizes the original tweet's linked quote-tweets with a *share-type,* which allowed us to map trends such as whether users tended to agree or disagree with the *topics* as they were presented.

Hossman et al. (2011) designed a module to be added to Twitter that might help the Twitter platform become more useful to end users during a disaster event. This research adopted a top-down approach by creating a Twitter disaster mode with the purpose of connecting victims with responders in the event of a disaster. This research outlined the gap between Twitter and a hypothetical ideal platform designed for disaster response. This knowledge was a useful starting point as we attempted to define how quote-tweets fit into this gap. Hossman et al.'s (2011) platform redesign only indirectly relates to our research purpose. This redesign only applied broadly to communication via original tweets during a disaster, but the study occurred before quote-tweets were invented, and long before the COVID-19 pandemic.

3 METHODOLOGY

Here, we discuss the strategy we used to measure outcomes in quote-tweets and their corresponding original tweets. Also discussed is the process by which this strategy evolved throughout the research. After some trial and error, we settled on two coding operations: one focused on original tweets, and the other on quote-tweets. This schema would highlight some of the differences between quote and original tweets as well as link sharing behavior with topic usage. Two questions we sought to answer with this strategy were, 'How are different sub-topics within COVID-19 tweets being shared?' and 'How did topics and sharing change over time?'.

3.1 Development of a Qualitative Approach to Quote-Tweet Categorization

The first step of the qualitative portion of this research was to gather an initial set of quote-tweets spread over the course of the COVID-19 pandemic. We wrote a script to gather tweets using the Twitter Streaming API, which used search terms of *COVID-19* and *Coronavirus*. This same script also stored the data received from the API in a Mongo database. Data stored includes the username, description, and text of both the original tweet and the person quoting it. This script was initiated on four separate dates spaced approximately one month apart starting in late January of 2020 and ending in May of 2020. The result was five separate database collections with a combined 1.3 million tweets.

Coronavirus as a topic fluctuated in popularity and quantity over the months of collection. This meant that how long the script ran each time would fluctuate because we gathered approximately 100,000 tweets in each pass. The time required to gather 100,000 tweets ranged from 2 hours to 11 hours for each set. The demographics of user nationality might have been affected by this, as a wider time-lapse would encompass high traffic periods for more regions than just the western hemisphere. Each month the script was started between the hours of 4:00 and 8:00 pm MST.

We conducted an initial study of the dataset with the goal of identifying the topics found in the tweet text. We gathered random samples of 1000 tweets from each of the first four collections (the 5th set had yet to be collected). We removed duplicate retweets and non-English tweets, which cut the document down to about 1600 rows with disproportionately fewer tweets from the first two collection dates. There were more non-English tweets in the early datasets. This was likely because Asian countries were more heavily affected by the virus in January and February and it wasn't until March that the pandemic took the world stage. We only looked at the non-English tweets on our first analytical pass of the dataset. The purpose of this first pass was to start deriving categories that would apply to tweets of any language. We used these categories to generate the *tweet_topic* field which we would relate to each row of our next and final set of tweets.

Our second goal was to identify categories that described the relationship between original and quoting users. We wanted to answer questions such as: How often does the quoting user agree with the subject matter presented by the quoted user? How often are the words of the quoted user inspiration for a joke from a quoting user? We used questions like this along with the

1600 rows of sample tweets to generate the *share_type* field which we would relate to each row of our final set of tweets.

Deriving the final list of *tweet_topics* and *share_types* was an iterative process. During this process, we read samples of tweets from each month and labeled the tweets with words that best described the content for each. Once we had applied descriptive words to several hundred tweets, we filtered the list of descriptive words—combining the least common words on the list with others to form more general descriptors. For example, the topic of *politics* originally consisted of several other *topics* describing what kind of politics the tweet exemplified: right-wing, bipartisan, local, American, etc. Hsieh, H.-F., & Shannon, S. E. (2005) described this approach as a conventional content analysis, which means all the topics were derived from the tweet text itself. The same process was applied for *share_type*, which resulted in fewer outcomes than *topic.* We wrote detailed instructions and definitions to document and track our decisions for the *tweet_topic* and *share_type*.

3.1.1 Tweet Topic Coding Scheme

The final coding scheme for the *tweet_topic* category can be found in the table below. The relationship between tweet and topic was many-to-many, meaning that each tweet could have more than one topic, and each topic applied to many tweets. On average each tweet had about 3 *topics.* The average *topic* applied to about 15 percent of the tweets. This ranged from smaller *topics* such as *disinformation* applying to 2 percent of the tweets, to *health,* which applied to 60 percent of tweets.

Table 3-1: Topic Labels, Descriptions, Examples, and Action/Subject Type

Topic	Description	Example
borders	Travel, migration, or relations between different regions	"We try to stay hopeful, because that's all we've got ... But each day, that becomes a little bit more difficult when country after country rejects..."
economy	Companies, employment, money, and markets	"If this isn't the retail apocalypse I don't know what would be." Thousands of stores will close as the coronavirus turbocharges a shift to ecommerce.
cancellation	Altered, delayed, closed, or cancelled events or business	McDonalds is shutting down restaurants in five cities in China amidst the Wuhan coronavirus outbreak
community	Groups of people living in the same place or having a particular characteristic in common	As residents of the city of Seattle, we've been hit hard and have witnessed firsthand how quickly these disastrous situations can escalate. Our kids' schools have closed along with universities and businesses.
conspiracy	Hidden organization or idea that is responsible for circumstances	"The virus is influenza, and it poses a far greater threat..."
criticism	Disapproval of someone or something based on perceived faults	"We're gonna lose over 100,000 perhaps" -- Trump just moved the goalposts *again* about the projected US coronavirus death toll
death	Death caused by the virus	"thank you, you gave your lives for others" Dr Shafi will be hugely missed by all
demographics	Scientific or statistical findings about human populations and the spread of the virus	BREAKING NEWS: This is not a scene from some apocalyptic horror movie, this is a #coronavirus outbreak in China. The SARS like virus has already spread to four countries and infected more than 1700 people. US airports are monitored.
disinformation	Narrative that is misinformed or detrimental to public health	"I am opposed to vaccination and I wouldn't want to be forced by someone to take a vaccine in order to be able to travel..."
entertainment	Sports, movies, cruises, and recreation	"In light of the COVID-19 outbreak, the UAAP, after a thorough deliberation by the Board of Trustees and the Board of Managing Directors, have come to a decision to postpone all sporting events..."
health	Conditions of health and healthcare	"Early data from some of these studies suggest that a relatively small percentage of the population may have been infected, even in heavily affected areas"-@DrTedros #COVID19
humor	Sentiment attempting to be amusing and comic	"Coronavirus cases has risen" Me still catching flights this year
media	Journalists and news organizations	"Under no circumstance should these briefings be carried live. Doing so is a mistake bordering on journalistic malpractice. Everything a president does or says should be documented but airing all of it, unfiltered is irresponsible."

Table 3-1: Continued

Topic	Description	Example
news	Text written in an informative third person tone	#China Travel Advisory Update: The Travel Advisory remains Level 2: Exercise Increased Caution. Some areas have increased risk. Hubei province currently Level 4: Do Not Travel due to novel coronavirus first identified in Wuhan, China. Read Advisory here:
overwhelmed	Infrastructural inability to respond to the virus	"Hundreds of patients in #Wuhan who have yet to be confirmed as carrying the new strain of #coronavirus are becoming increasingly desperate as the city struggles to cope..."
politics	Political figures and actions, often regarding leadership as much as health issues	The Tories have announced a 2 year public sector pay freeze. So instead of the billionaires paying for the Coronavirus bill, the Tories want the doctors, nurses and social workers to pay. The very people who have risked their own lives and saved others.
reassurance	Sentiment attempting to remove fear	We've won a battle, we've done well, but #COVID19 is still out there and most Ohioans are still susceptible to it. The spread concern is still as strong today as it was a month ago.
rumor	Circulating story or report of uncertain or doubtful truth	#China is welding people into their homes because of the threat of #coronavirus 🧐 😷 ☠️
uplifting	Positive sentiment attempting to inspire hope	"A good Samaritan named Monica sent me 6,000 shillings." Kenya's "Adopt a Family" program is connecting well-off families..."
warning	Impending danger, problem, or other unpleasant situation	"A single highly contagious person in a crowded space can start a chain of disease that quickly encompasses dozens if not hundreds of people."

3.1.2 Tweet Share Type Coding Scheme

The coding scheme we developed for *share_type* contains far fewer categories than the scheme for *topic*. While the *topic* scheme identifies what each tweet was about, *share_type* describes the types of relations between the quote-tweet author and the author of the original tweet. The first outcome we discovered through analysis of quote-tweets was agreement. Quoting users often displayed either agreement, neutrality, or disagreement with the original tweet. This meant that context mattered with *share_type* in a way that it didn't with *topic.* Agreement was the most basic descriptor of the relationship between quoting and original tweet.

Sometimes it was difficult to evaluate agreement because the quoting tweet was too short or lacked context. This behavior has been identified by previous research (Garimella et al. 2015). As an example, sometimes users simply tagged other users in the text of their quoting tweet as a way to forward the original tweet to other users. Other users included emojis or a single word. We used the *simple share_type* to categorize single word and forwarding tweets. We used the *discussion share_type* to categorize tweets that attempted to further explain or opine. All tweets were labeled as either the *simple* or *discussion share_type.*

*S*ome tweets displayed an emotional response reflecting *disbelief* while others attempted *humor*. Enough tweets followed a pattern found in these two categories that we created a *share_type* for each of them. Some tweets reflected both *simple* and *disbelief* qualities, but not all *disbelief* tweets fit into the *simple* subset.

One unexpected finding was that sometimes the original tweet and quote-tweet were written in different languages. We labeled these cases as the *mixed share_type* and performed a deeper analysis on this behavior that can be found in section 4.5. We derived six possible categories that applied to *share_type*, and these could be divided into 2 couplings and 3 other outcomes which can be seen in Table 3-2.

Our final dataset was gathered by querying 500 random unique quote-tweets from each of our 5 data sets. The resulting 2500 tweets were coded by two people. To ensure consistent output, the two coders both analyzed 800 common tweets. We calculated inter-reliability for

Table 3-2: Share Type Labels, Descriptions, and Examples

share_type	Description	Example
simple	Any response that is less than a complete idea that doesn't elaborate on the topic at hand	***Quoting:*** sigh lol 😊
discussion	A response that adds additional talking points to a quote	***Original:*** U.S. senators propose $500 billion rescue for state, local governments ***Quoting:*** Blue State Governors cannot retain the current lock-down status without funding to enforce it....
	Agreement Type	
agree	A response of any length that holds the topic of the quote in a positive or affirmative light	***Original:*** I gave convalescent plasma! If you've had COVID-19 symptoms and recovered, you can too - sign up here: ***Quoting:*** Note excellent taste in books
disagree	A response of any length that holds the topic of the quote in a negative or disagreeing light	***Original:*** Rumors spread on social media that snorting cocaine and drinking bleach can cure coronavirus – they can't ***Quoting:*** Liberals are stupid.... Media? even worse.
	Other Types	
humor	A lighthearted response attempting to be amusing and comic	***Original:*** #BREAKING City of Las Vegas allowing sidewalk dining, sales during reopening phases ***Quoting:*** Sidewalks are in!
disbelief	An emotional response to the quoted message.	***Original:*** Trump is total meltdown. He told aides he thinks journalists want to get coronavirus on purpose to spread it to him on Air Force One. My latest: ***Quoting:*** We. Are. In. Danger.
mixed	An indicator that the quoted text is written in a different language than the quoting user's text	***Original:*** #BREAKING Two cases of coronavirus confirmed in France, first in Europe, says minister ***Quoting:*** ฝรั่งเศส ยืนยันผู้ติดเชื้อ #ไวรัสโคโรนา 2ราย และถือเป็นที่แรกในยุโรปด้วย

these tweets across all coding categories using an online calculator called Recal2 (Freelon D, 2008). The results were mixed between the categories, with some results being more consistent than others. This helped us to identify some ambiguity that existed between our topic definitions. Our primary measurement of success was the Cohen's Kappa unit, but we also needed a benchmark for a minimum acceptable reading of Cohen's Kappa. The general table of measurements was gathered from an article (Schnell, 2020), which provided a range of Kappa values that could be considered acceptable for our coding schema.

- 0.01-0.20 no agreement
- 0.21-0.40 as none to slight
- 0.41-0.60 as moderate
- 0.61-0.80 as substantial
- 0.81-1.00 as almost perfect agreement

We set a goal of having each field classification from both *topic* and *share_type* having a moderate (0.41) Kappa score or better. The categories that didn't originally satisfy this benchmark were *borders, community, conspiracy, disinformation, rumor,* and *news.*

To better understand the disjoint between these coded fields, we read though a filtered list of tweets that had been coded with one of these weaker topic categories. Most of the fields needed to be more narrowly defined. For example, we adjusted the definition of the *borders* category to be more about travel between countries and less about international relations. We also rescoped the *community* category to focus on those tweets that relate to a real, physical community, instead of also including more abstract concepts such as tweets that fostered a sense of community. *News* was applied to most tweets. Of all the topics we defined, *news* was the broadest. The reason this was originally scoring low was because it was easy to forget to mark it when other topics were more prominent.

Table 3-3: Interrater Reliability Metrics for Each Topic

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha	N Agreements	N Disagreements
borders	95.05300353	0.390836408	0.400049481	0.391195162	807	42
cancellation	96.81978799	0.735301759	0.735559785	0.735457647	822	27
community	91.401649	0.520303714	0.524369748	0.520586221	776	73
conspiracy	88.57479388	0.334405573	0.345578945	0.334797561	752	97
criticism	89.39929329	0.702555199	0.702888294	0.702730373	759	90
death	98.58657244	0.834706989	0.834792761	0.834804335	837	12
demographics	95.28857479	0.830866386	0.831002737	0.830965994	809	40
disinformation	98.93992933	0.684262722	0.684269245	0.684448668	840	9
economy	96.1130742	0.768394251	0.768555922	0.768530651	816	33
entertainment	98.11542992	0.812271973	0.81231866	0.812382532	833	16
health	82.09658422	0.628916494	0.630042028	0.629135036	697	152
humor	97.40871614	0.563557342	0.563883441	0.563814375	827	22
uplifting	97.52650177	0.65382263	0.653906629	0.654026504	828	21
media	93.28621908	0.521810663	0.522492476	0.522092282	792	57
news	76.32508834	0.478326661	0.484024781	0.47863389	648	201
overwhelmed	95.17078916	0.428071473	0.428188912	0.428408297	808	41
politics	86.45465253	0.690441201	0.692500779	0.690623509	734	115
rumor	92.5795053	0.252980077	0.266306361	0.253420018	786	63
tip	96.34864547	0.567566235	0.56859049	0.567820908	818	31
warning	96.4664311	0.426196269	0.429307641	0.426534198	819	30
combined rumor/consp/disinform	84.92343934	0.403543437	0.404073306	0.403894707	721	128
Average	93.09776207	0.591279701	0.5939314558	0.5915204081	790.4	58.6

Conspiracy, disinformation, and *rumor* were the most difficult to improve. Reading through these tweets made us realize that each of these topics had similar and often overlapping meaning. The result was a mix of these categories being used to code tweets that contained unverifiable and/or misleading statements. When we combined each of these fields shown on the last row in the following table, the Kappa value met our desired outcome.

Share_type required less rework because the definitions were clearer. Though one point of confusion needed to be addressed early in the coding process. Agreement was harder to evaluate when the original tweet contains news, about which that the quoting user opines. For example, regarding a statement made by a federal judge, a news organization tweeted:

> JUST IN: A federal judge in California has ordered ICE to make "custody determinations" — consider releasing — *all* immigrants over 55, those who are pregnant and detainees who suffer from chronic health conditions, for as long as coronavirus poses a "substantial threat of harm"

To which the author of the quote-tweet responded:

> Heaven forbid we fast track deportations in order to "protect them".

The user disagrees with the sentiment expressed by the federal judge, who is not the author of the tweet, but is the subject of the tweet. From examples like this, we determined that agreement should be defined by the attitude displayed toward the subject of the tweets that don't directly express opinion.

We had a multilingual coder assign values to the majority of our final study. This was invaluable to us because the majority of non-English tweets were Spanish in this set. Of the final study's 2500 tweets, ~1800 were English, 280 Spanish, 50 Portuguese, and the remaining 370 were one of about 30 other languages. We used Google translate to convert these 370 other language tweets into English.

Table 3-4: Interrater Reliability Metrics for Share Type

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha	N Agreements	N Disagreements
agree	81.38987044	0.559161058	0.560317023	0.55942068	691	158
discussion	86.92579505	0.724729481	0.724762041	0.724891595	738	111
disagree	87.04358068	0.477087953	0.483016319	0.47739591	739	110
humor	92.34393404	0.653844344	0.654157815	0.654048205	784	65
simple	82.44994111	0.648366585	0.651862452	0.648573672	700	149
disbelief	90.69493522	0.506651318	0.510969661	0.506941865	770	79
Average	87.89163722	0.602135936	0.604953658	0.60237025	746.2	102.8

The final product of executing this methodology was a set of 2500 quote-tweets linked to original tweets ranging from January to May of 2020. Within this set of tweets, we coded each to contain one or more topic mapped to the original tweet and one or more share type mapped to the quote-tweet. Later we analyzed and produced chronological graphs for each of these topics and share types and compared them.

4 ANALYSIS OF DATA GATHERED

In this section we discuss the outcome of executing our coding methodology upon the set of 2500 tweets. These tweets came from a variety of different languages, locations, and were spread from January to May of 2020. A summary and frequency section is provided for both topic and share type. The purpose of each summary section is to provide an overview of the outcome of the coding schema itself, with graphs displaying the makeup of topic and share type throughout the entire set. Frequency on the other hand engages in a deeper discussion for each individual category within topic and share type. In addition to the coded fields, we also comment on the languages found within our data reserves from which the set of 2500 tweets was drawn. Concluding this section, we compare our results to those found in previous research dealing with quote-tweets.

4.1 Topic Summary

The frequency of each topic in the tweet data set for each month varied over time. We generated graphs displaying the percentage of tweets coded with each *topic* for the complete dataset (Figure 4-1 Left) and for each monthly dataset (Figures 4-1 through 4-3). Across all datasets (Figure 4-1 Left), *health* applied to about 50 percent of the tweets, but this ranged between over 70 percent (Figure 4-1 Right) in January to almost 30 percent in April (Figure 4-3 Left). The average number of topics that each tweet had was almost 3.

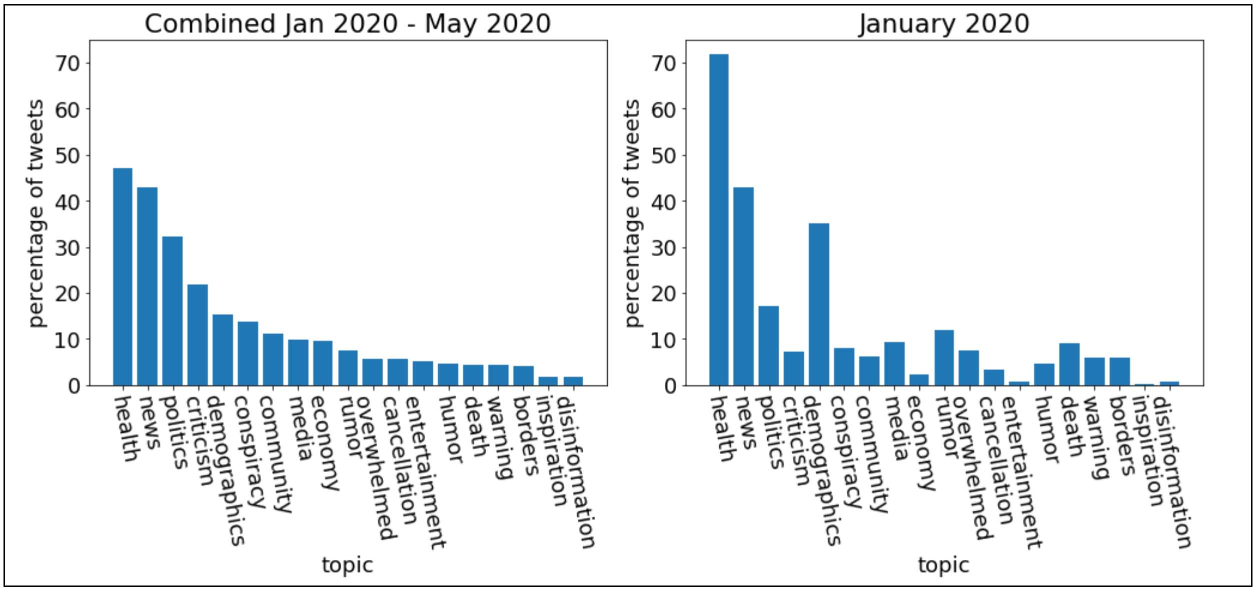

Figure 4-1: Topic Distribution for All Sets, January-May 2020 (Left) and Jan 2020 (Right)

In January (Figure 4-1 Left), *health* and *demographics* appeared much more frequently than in the complete five-month dataset, while the topics of *criticism, politics, and economy* saw lower than average numbers in January. If we combine all the sets and calculate the average frequency for each topic, *health* tweets were almost twice as likely to occur as *political* tweets. In January, *health* tweets were almost 4 times as likely to occur as *political* tweets, but in March *health* tweets were occurring at just about the same frequency as *political* tweets, *political* tweets being slightly more likely than *health*. In other words, political tweets more than doubled over the course of 2 months while health tweets halved in the same space of time. Based on this data, it seems that the discourse about Coronavirus and its related societal impact grew more political and less health-centered over time.

Politicization may have been at least partially due to the social impact COVID-19 had on people's lives. Lockdown and quarantine seemed to affect people almost as heavily as the virus itself in some cases. We saw the term 'quarantine' shift from being coupled with demographics and news, to more often applying to political criticisms such as when a public figure refused to quarantine, or acknowledge mask mandates:

Tweet from January: China quarantines 35 million people amid the coronavirus outbreak ***Tweet from March:*** Rep. Louis Gohmert (R-TX) was told he was in proximity of individual at CPAC who tested positive for coronavirus, according to a Gohmert aide. Gohmert is choosing not to self-quarantine. @FoxReports reporting

Figure 4-2: Political Tweets Containing the Word 'Quarantine' Increased Over Time

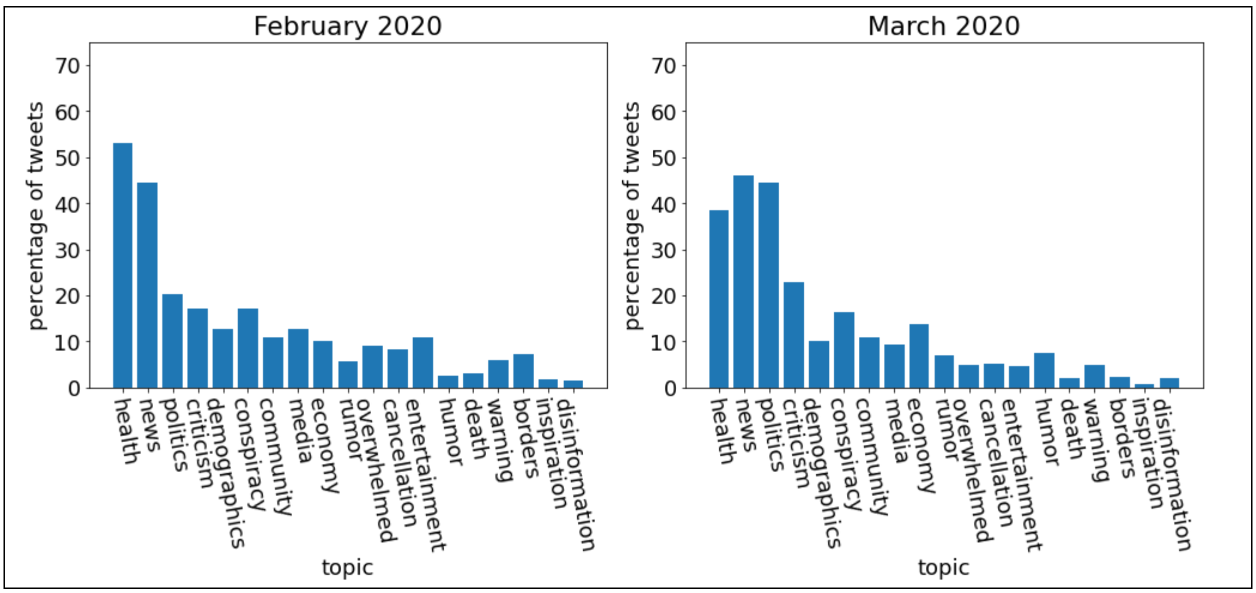

Figure 4-3: Topic Distribution for February (Left) and March (Right) of 2020

February (Figure 4-3 Left) was the closest to the average of all the sets. January was an outlier to the other months, and it was February that resembled January the closest. By March (Figure 4-3 Right), *health* was no longer the highest occurring topic. We saw a higher influx of tweets reflecting *politics* and *criticism. Politics* and *criticism* probably increased due to negative social impact coming from both policy and the virus itself, but why was there such a drastic decrease in *health*-related tweets? One reason might be that in January the status of Wuhan, China (the origin of the COVID-19 virus) dominated the news cycle, and most tweets contained dialogue attempting to explain scientifically what the virus was, and where it came from. In later months fewer tweets offered this kind of exposition. The conversation seemed to shift from ‘What is Coronavirus?’ to ‘How does Coronavirus affect me?’.

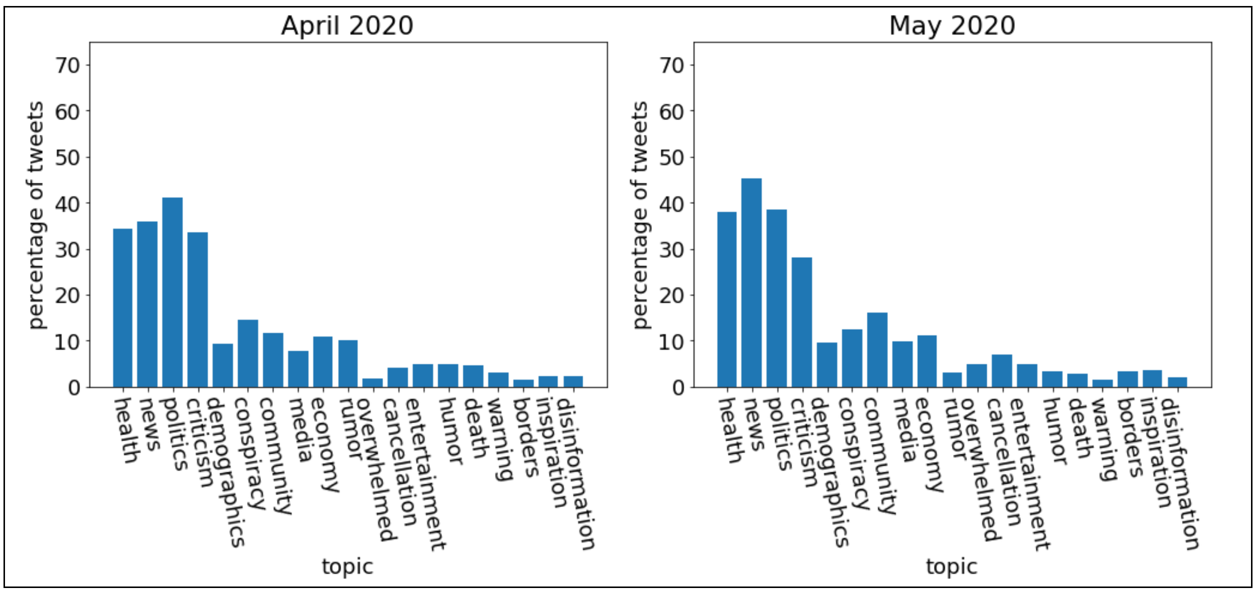

Figure 4-4: Topic Distribution for April (Left) and May (Right) of 2020

April and May followed a similar trend to March, with higher levels of *political* tweets than *health* tweets. Figures 4-3 and 4-4 show that only about 10 percent of tweets were marked with the *topic* of *demographics* between February and May, but in January (Figure 4-1 Right) almost 40 percent of tweets were marked with this *topic*. This meant that the average seen in Figure 4-1 Left skewed high. Overall, these graphs illustrate that Coronavirus was a dynamic Twitter topic that changed over time.

4.2 Topic Frequency Over Time

After coding each topic, we plotted the frequency with which each category appeared in our dataset across the 5 months we collected the data. Together these graphs paint a broad picture of how the pandemic was talked about on Twitter at the time. The frequency of some topics increased over time such as *community*, *politics*, *disinformation*, and *economy*, while other topics decreased such as *death*, *demographics*, *health*, and *warnings*.

The first two graphs (Figure 4-5) describe the frequency of *borders* and *cancellation* tweets in our data. Each point on these graphs represents the number of tweets out of 500 that contained a particular topic. The highest point on the *borders* and *cancellation* graphs is about 35 and 40 respectively, meaning that at their highest usage these topics each applied to about 8 percent of that month's tweets. We could see that their shapes are similar with a frequency count that is similar on the y axis. This means that tweets about *borders* and *cancellations* appeared in the same months with the same frequency, but we were unsure if the two were linked.

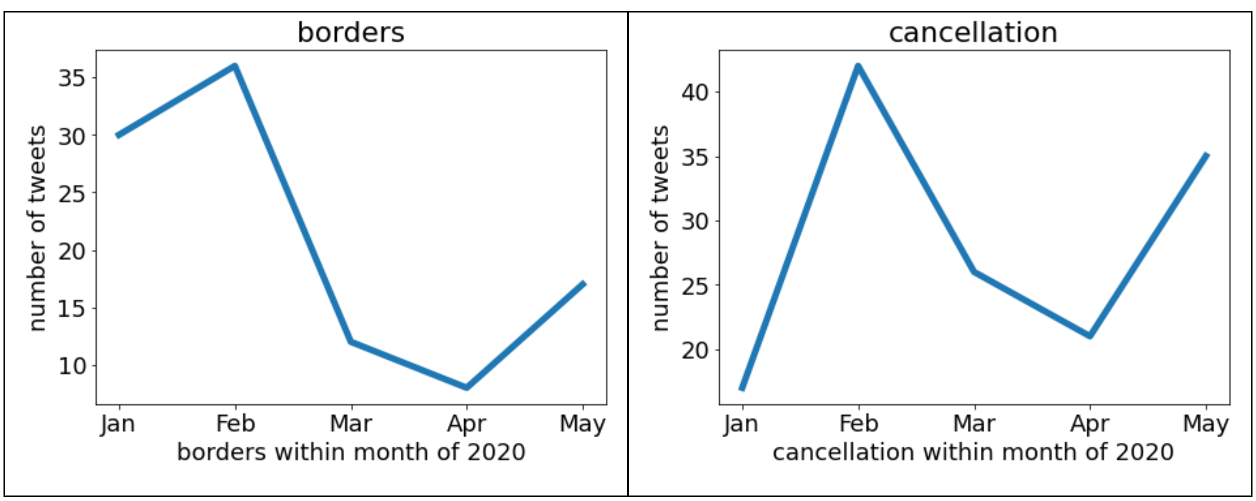

Figure 4-5: Borders (Left) and Cancellation (Right) Topic Count Within Each Set

Tweets tended to discuss the topic of *borders* more in the early stages of the pandemic. In January and February, many political posts discussed whether travel relations should continue with China and other hard-hit countries.

> ***Tweet:*** Hey, we have a better idea. Let's suspend ALL flights from China carrying mainland Chinese. Who's with me?! 🙋🏻‍♀️ #coronavirus

Figure 4-6: Tweet Calling for More Travel Restrictions with China

Cancellation followed a similar trend, the biggest difference being that January contained no mention of cancellations. The initial large-scale, international responses to COVID-19 cases began in February, which issued the closure and cancellation of many non-essential businesses and events. The *cancellation topic* label was used to indicate business and event status changes. So, when some businesses began to open their doors again, if it was announced via a tweet then we would likewise apply the *cancellation* label. The upward trend in May indicates cancellations as well as reopening events.

> ***Tweet:*** Boy Scouts banned from planting #American flags on veterans' graves for #MemorialDay due to coronavirus.
>
> ***Tweet:*** Gyms and fitness centers can reopen on May 26 if they can meet safety protocols. Protocols will be on later today.

Figure 4-7: Tweets Announcing Cancellations and Reopening Events in May

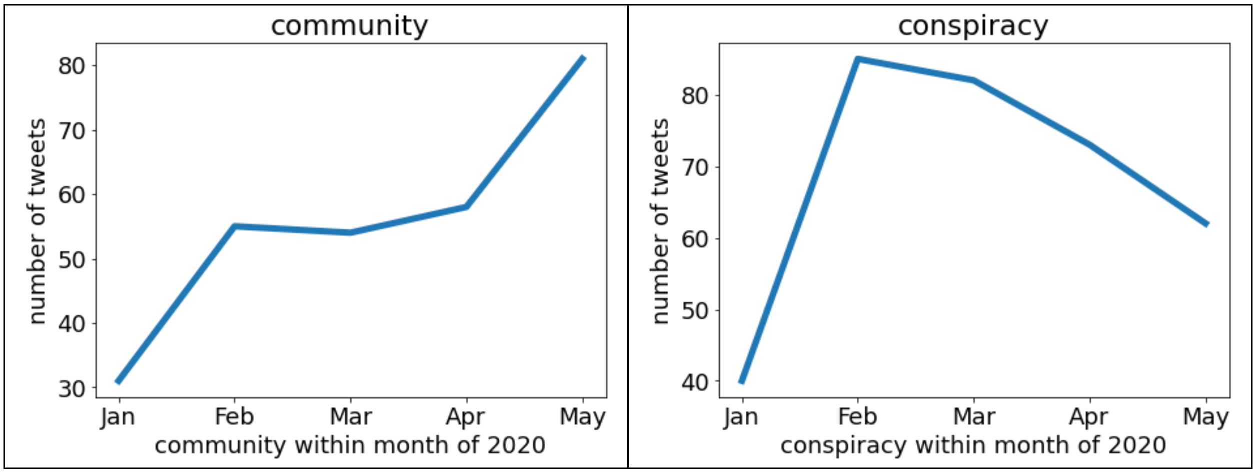

Figure 4-8: Community (Left) and Conspiracy (Right) Topic Count Within Each Set

The *community* topic (Figure 4-8 Left) followed a consistent upward trend. This topic centered around the impacts COVID-19 had on a community level. Many tweets are addressed to a local rather than a global audience. We sought to categorize these local audience tweets by labeling them with the *community topic*. As we moved forward in the pandemic, its effect on society increased, and it became more and more relevant to discuss local events in relation to COVID-19. The most common theme of *community* tweets was mentions of local events or news stories. Often these tweets contained political commentary or criticism toward people associated with the event. This led to the *community topic* commonly being paired with other *topics* such as *politics, cancellations,* and *criticism.*

These tweets demonstrate the depth to which COVID-19 permeated community news. Neither tweet (see Figure 49) would have made sense had they happened in January: Lockdown measures had not been legislated, and frontline workers had not yet been exposed to increased risk. However, by May tweets like this had become commonplace.

Tweet: In East Texas, armed protesters patrol a tattoo shop that defied the lockdown and reopened Friday

Tweet: #LIVE A B-52 bomber is flying over the Antelope Valley to honor front line workers during the coronavirus pandemic.

Figure 4-9: Tweets Discussing Community News in Texas and Antelope Valley California

Conspiracy tweets doubled between January and February 2020 (Figure 4-8 Right), then gradually fell in the three months that followed. In February, the virus began to take world stage, but politicians had not started speaking about it. Much news of demographics and death combined with a lack of political coverage may have encouraged Twitter users to discuss conspiracy theories freely. The gradual decline may have been due to debunking and the slow tiring out of the same conspiracy focus topics like COVID-19 patents, 5G spread, or Bill Gates' involvement (Reuters, 2020).

Tweet: Lawyer and citizen journalist Chen Qiushi vanished while documenting the coronavirus lockdown in Wuhan. He spent weeks filming patients and overrun hospitals

Figure 4-10: Tweet Discussing the Disappearance of a Media Figure

In February, users were particularly vocal about government coverup and silencing of media figures such as Chen Quishi.

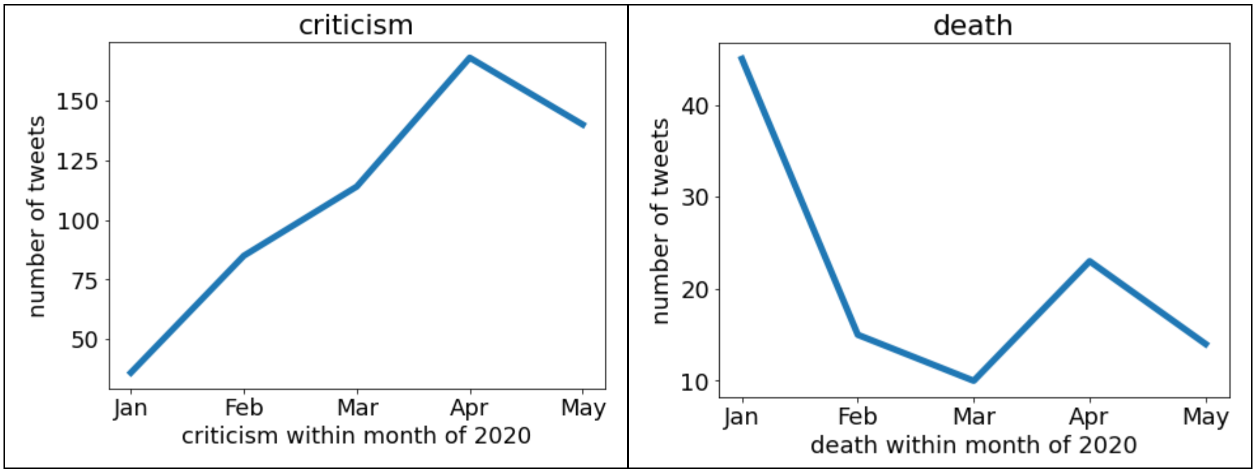

Figure 4-11: Death (Left) and Criticism (Right) Topic Count Within Each Set

Criticism followed the same consistent upward trend as *community* (Figure 4-11 Left), probably for the same reasons. As the burden of quarantine, social distancing, and general economic downturn increased, so did criticism, especially directed at political figures. Criticism invited more criticism.

Original Tweet: Passenger blasts United Airlines for yet another packed flight amid coronavirus

Quote-Tweet: So people on planes are getting mad at the airlines for there being *checks notes* other people on the planes?

Figure 4-12: Critical Quote-Tweet Countering Another Critical Tweet Discussing Social Distancing Within Air-Travel

We found a rise in the *share type* of *disagreement* that was linked with this rise in *criticism.* As criticism rose within topic tweets, a higher percentage of responses contained disagreement (Figure 4-13 below).

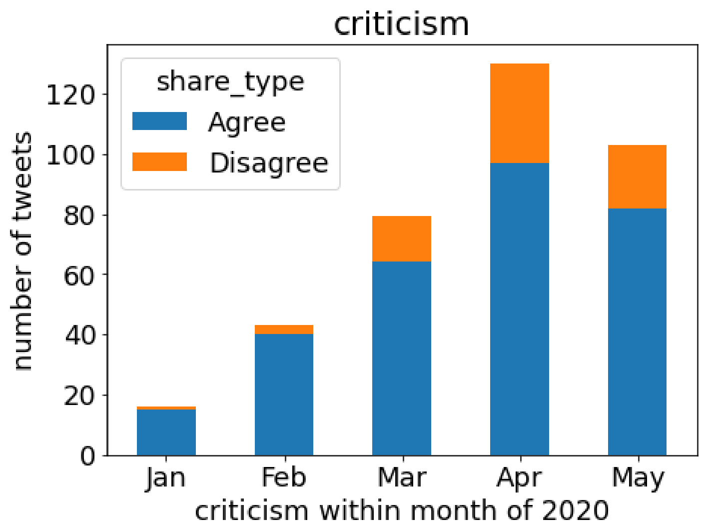

Figure 4-13 Disagreement in Quote-Tweets Increased as Criticism Increased

Death as a topic followed an unexpected trend (Figure 4-11 Right). Mentions of death were highest in January when most related news coverage followed the spread and destruction of the virus in Wuhan. In February and onward, other topics may have overshadowed *death*, even though deaths were increasing as time went on.

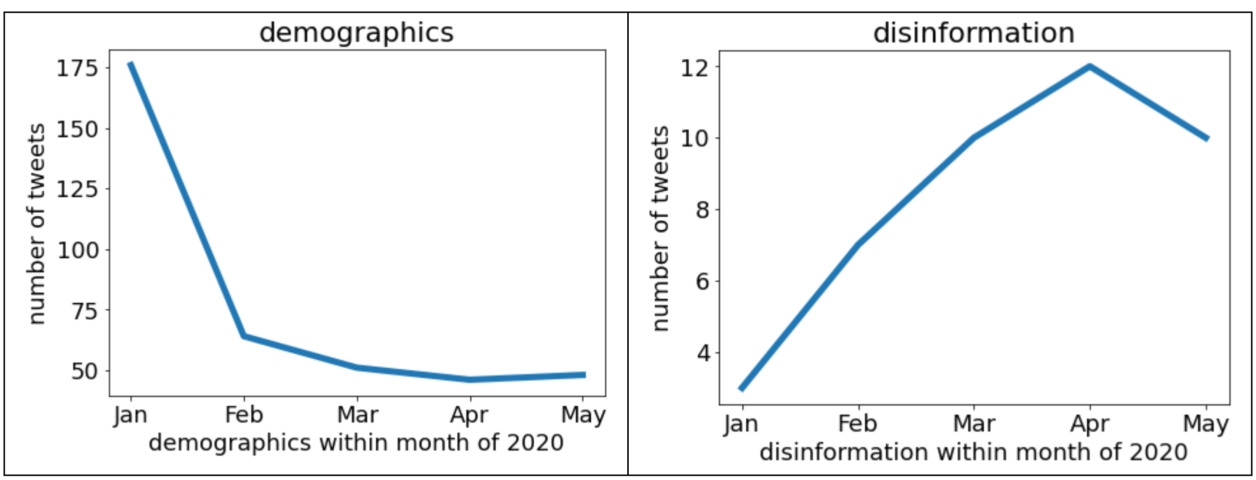

Figure 4-14: Demographics (Left) and Disinformation (Right) Topics Within Each Set

Demographics (Figure 4-14 Left) was closely tied with *death*. Almost all tweets marked with this topic were reporting number of cases, deaths, or recoveries. As with *death,* this topic was likely overshadowed by others in later months. *Demographics* was a topic that displayed consistent characteristics such as naming a geographic area and citing numbers and percentages of people affected within that area. This contrasted with other topics like *conspiracy* or *economy* which had new stories and conversations develop over the course of the study. *Demographics* tweets repeated the same messages over and over again and we likely saw people grow tired of the unchanging, albeit concerning narrative that cases were increasing in nearly every geographic region. This may have contributed to the declining numbers over the months.

Tweet from January: BREAKING: A case of coronavirus has been confirmed in Victoria.

Tweet from May: #BREAKING: Georgia verifies 1K new COVID-19 cases in 24 hours

Figure 4-15: Tweets from January and May Reporting Demographic Information Using a Similar Pattern

Disinformation (Figure 4-14 Right) followed the same trend as *conspiracy, b*ut we didn't apply this topic label very much. This was probably because we used this topic to cover fringe disinformation cases when most disinformation was either a conspiracy theory, or just a rumor.

Economy was barely present in January and peaked in March at about 15 percent of the set. The March dataset was gathered just a week after the U.S. issued non-essential businesses a stay-at-home order. Economic tweets tended to be tied to the topics of *community, politics,* and *criticism.* The focus of the tweets of this topic changed over time. The ~10 *economy* tweets from January discussed travel, and business status within Wuhan. The tweet seen in Figure 4-17 below makes reference to the wet markets of Wuhan which were rumored to have had a role in the origin of COVID-19.

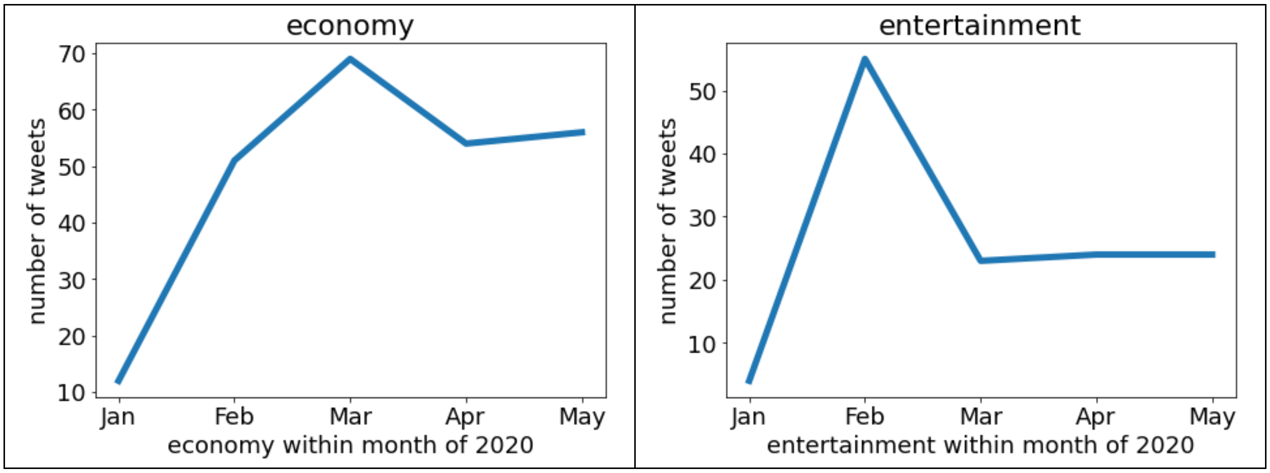

Figure 4-16: Economy (Left) and Entertainment (Right) Topic Count Within Each Set

Tweet from January: This market reportedly sold 112 different types of live animals for human consumption.

Figure 4-17: Tweet Discussing the Markets with Potential Role in the Origin of COVID-19

By February, the pandemic carried much more economic weight. Tweets sampled just 3 weeks after those in January focused more heavily on the world economy. This increased focus on the economic aspects of the pandemic remained a factor for the remainder of the study. January held only a quarter of the economy focused tweets of average found throughout the whole study.

February Tweet Translated from Spanish: China is printing money galore. Q1 and Q2 are already yielding negative results. China's GDP will fall to less than half in less than 66 days. Will the west enter the crisis of the #coronavirus and #vivaespaña?

February Tweet Translated from French: This health crisis in the #coronavirus once again stresses that, in the absence of an industrial relocation policy led by a strategic state, the market is driven by health insecurity and threatens the French." #DirectAN #QAG My question to @agnesbuzyn.

Figure 4-18: Translated Tweets Discussing Economic Impacts of the Virus in February

In some ways *entertainment* and *cancellation* were subsets of the *economy topic. Entertainment* was one of the most heavily affected industries in the world economy and *cancellation* was one of the most prominent economic actions taken. February saw the most tweets for both *entertainment* and *cancellation.* Almost 10 percent of tweets in February had both of these topics.

Tweet from February: JUST IN: UAAP will postpone all sporting events due to the COVID-19 outbreak.

Figure 4-19: Tweet Announcing the Cancellation of UAAP Sporting Events

Most *entertainment* tweets were tied to *cancellation.* It's important to remember that this does not reflect Twitter as a whole. These tweets were gathered using 'coronavirus', and 'COVID-19' as search topics. Most *entertainment* tweets probably were not cancellation announcements, but the tweets that mentioned entertainment and coronavirus were.

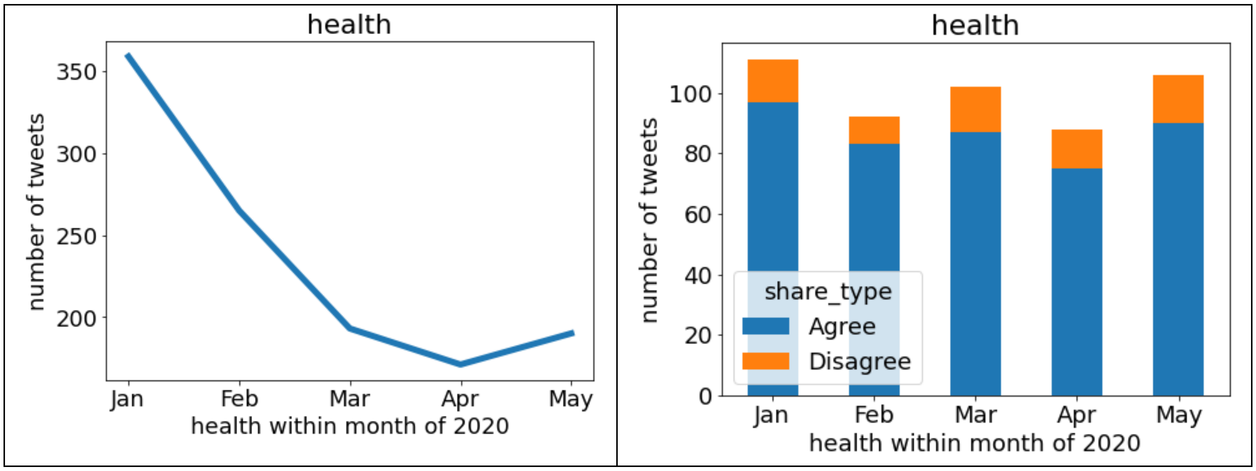

Figure 4-20: Health Topic Count Within Each Set (Left) and Health Agreement (Right)

Health was a high frequency topic throughout the entire study. The *health* label indicated that the tweet conveyed information about healthcare, symptoms, community wellbeing, or virus transmission—even if one of these was only mentioned in passing. If a tweet *didn't* get marked with the *health topic,* it usually meant that the tweet was entirely focused on politics, the economy, or humor. Even at this graph's lowest point in April, *health* accounted for 35 percent of the dataset. In January 70 percent of tweets were marked with *health*. About half of the *health* tweets were also marked with *demographics.* The other half covered a variety of news such as hospitals being built, potentially new symptoms of the disease being discovered, etc. *Health* was paired with every other topic.

We saw *agreement and disagreement* remained at roughly consistent levels across the months when applied to *health* tweets (Figure 4-20 Right). One area of future research could explore whether this spread of *agreement* applied equally to each geographic region or language group of users.

Original Tweet from May: Family opens coffin at wake and five are contaminated by COVID-19 in Bahia

Quote-Tweet from May: About fake news being a crime

Figure 4-21: Tweet Announcing Casualties in Bahia and Quote-Tweet Calling This Fake

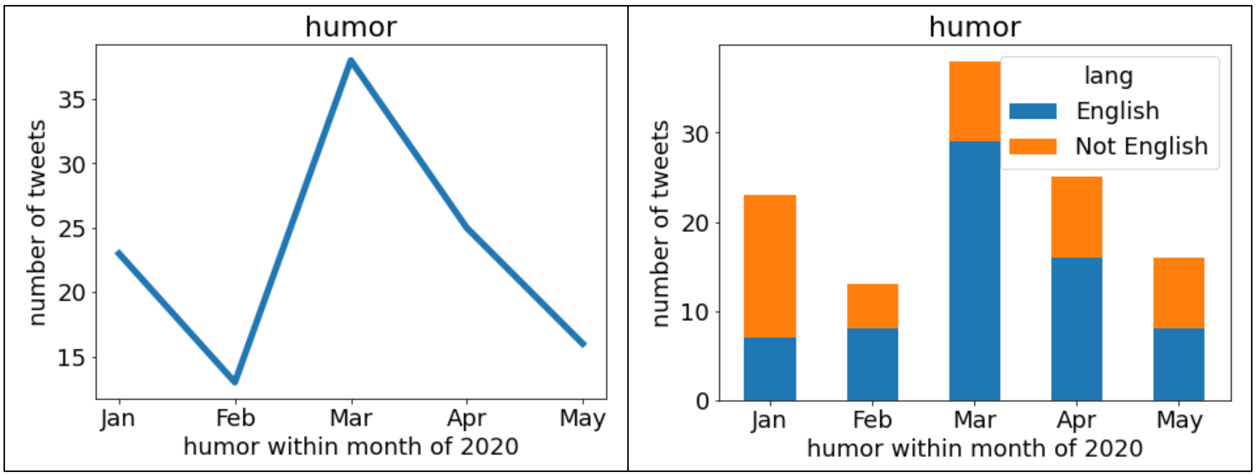

Figure 4-22 Humor Within Each Set (Left) and Language Breakdown for Humor (Right)

Humor did not occur as frequently as we initially thought it would. It was sometimes difficult to know whether the author was trying to be funny.

Original Tweet from January: Congratulations you've found a milk bucket to cure your Coronavirus

Quote-Tweet: steve has it

Figure 4-23: Tweet and Quote-Tweet Where Humor Was Difficult to Detect

Humor was the only label that applied both to *topic* and *share_type. Topic* and *share_type* share similarly shaped graphs but *share_type* consistently had about 3 times as many tweets marked with *humor*. This tells us that the twitter users in our set favored humor as a response much more often than as a topic. Quote-tweets give users something to point and laugh at.

A disproportionate number of tweets labelled with *humor* were written in a language other than English. Non-English tweets accounted for 27 percent of our sample overall. For the tweets labelled with *humor* they accounted for over 40 percent of the sample. This might indicate broader differences between how Twitter is used in the U.S, and other regions and cultures.

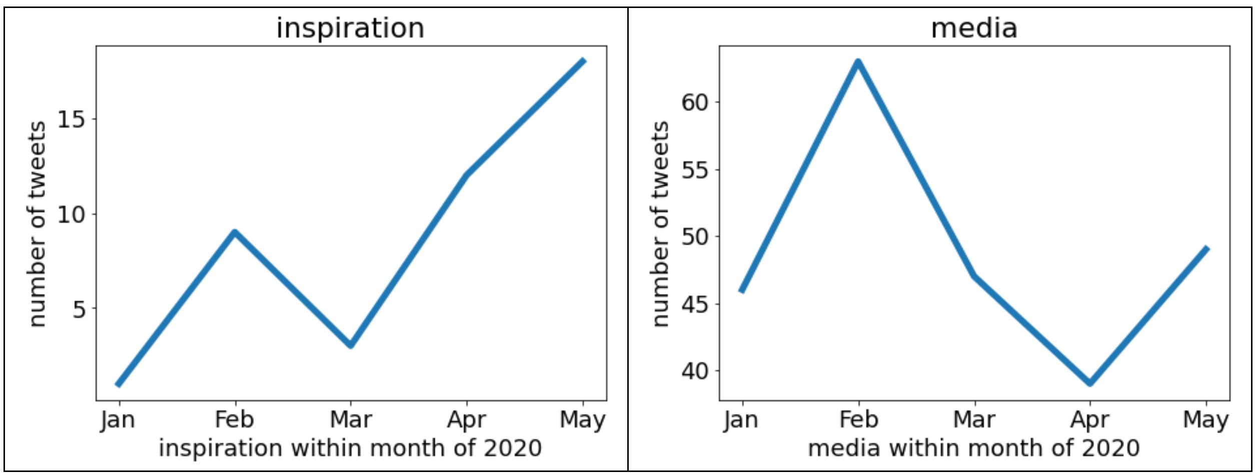

Figure 4-24: Inspiration (Left) and Media (Right) Topic Count Within Each Set

Inspiration was one of the smaller topics, peaking in May with just 17 tweets. These tweets usually contained pictures or stories about healthcare workers. In the case of the example tweet below, inspiration was found in the return of sea turtles to the now locked-down beaches. *Inspiration* was most commonly tied to the *community* topic.

About 10 percent of tweets consistently addressed the media itself, usually it was the media announcing a segment that they would run later in the day. Some tweets were criticisms of news organizations. In January and February, most of these tweets talked about news outlets lack of coverage of the virus' spread in China, or the missing Chen Qiushi and other Chinese journalists. In the later sets we saw a stronger link between *media* and *politics.* Shown below in Figure 4-26 is a tweet from a French media organization.

Original Tweet Translated from French: TH #coronavirus pandemic: In #Phuket (# Thailand) 11 turtle nests were found on beaches. It hadn't happened for almost 20 years. The turtles returned to the completely empty beaches following the #confinement. (BBC) # COVID19 ***Quote-Tweet Translated from French:*** Nature takes back its rights

Figure 4-25: Uplifting Tweet About Nature Thriving During Quarantine

February Tweet Translated from French: Arab media accuse Israel and the United States of creating and spreading coronavirus. See all the news on # i24NEWS ***Quote-Tweet Translated from French:*** Should know, was it not Allah who had created the virus to punish the Chinese?

Figure 4-26: Tweet About Arab Media Turned Religious Criticism via Quote-Tweet

This tweet and quote-tweet response are also a good example of how quote-tweets have the ability to change the topic of the original tweet. The original tweet addressed media in a foreign nation, while the quote-tweet addressed the dominant religion of that nation in the form of an irreverent joke. Many exchanges followed a similar pattern to this one where the original tweet is objective and inclusive, and the quote-tweet will be subjective and exclusive, and even offensive to some.

The broader implication of this pattern could be that this is how different political groups adopt tweets from trending topics and utilize quote-tweets to present that trending topic through their political lens. Many original tweets labeled with the topic of *community,* were presented through a quote-tweet as stories primarily about racism rather than community.

Original Tweet: Men less likely to wear face masks because they're 'not cool' and 'a sign of weakness'

Quote-Tweet: I would be interested to see this particular reason given broken down by race bc the part about racist profiling of Black men in masks is also a thing.

Figure 4-27: Original Tweet with Quote-Tweet Bringing up the Topic of Race

Other topics like immigration, states' or individual rights, international relations, and foreign aid were frequently brought up in a quote-tweeted response while not directly appearing in the original tweets themselves.

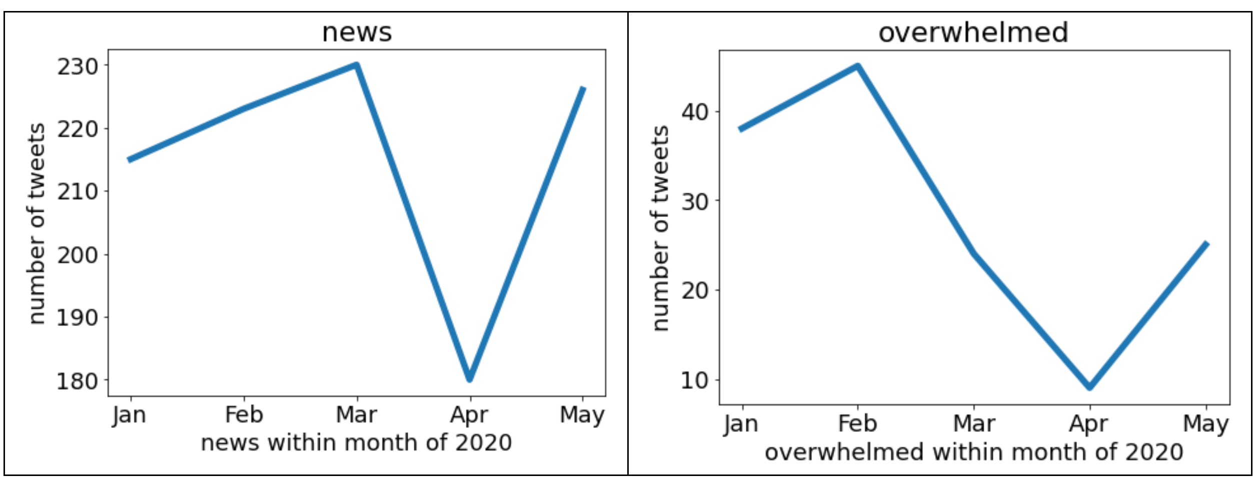

Figure 4-28: News (Left) and Overwhelmed (Right) Topic Count Within Each Set

News was a consistent topic applying to ~45 percent of the tweets in each set. April saw a large dip, dropping this down to about 35 percent. This dip may have been due to the increase in *criticism* tweets, which peaked in April at about 32 percent of the set. Sometimes these tweets contained headline paragraphs about the news stories, other times they contained no information directly and the story was only accessible through an embedded link.

Original Tweet: 14.05.2020 – COVID-19 Coronavirus Amakuru Mashya | Update | Mise à Jour ***Quote-Tweet:*** Intsinzi ndi kuyibona😊😊😊New Case:0 Recovered Today:4 #Tuzatsinda

Figure 4-29: Original Tweet with a News Story and Quote-Tweet Discussing Content

In our research, we never followed the links to read the news stories embedded in tweet, often quote-tweets gave clues as to what the articles contained. In the Kinyarwanda example above, the quote-tweet tells us that the cases within their region are declining.

In the early stages of the pandemic, one of the main goals each nation had was to flatten the curve so that healthcare within each region wouldn't become overwhelmed with cases. We saw places like Wuhan building new hospitals in a matter of days because case numbers were rising so dramatically (Lu, 2020).

Overwhelmed was a frequently appearing topic in the early months—like *health* or *demographics,* which were the most common *topics* associated with *overwhelmed*. The occurrences of this topic noticeably decreased as the political discussion increased. Political discussion shifted the mood of similarly worded tweets from *overwhelmed* to *criticism.* Tweets addressing the state of healthcare workers in Wuhan were often *overwhelmed.* Tweets addressing the conditions of American healthcare workers were often *criticisms.*

Overwhelmed Tweet: Photo from #Wuhan hospital. The sign says that all ER medical staff have been infected with #Coronavirus and the entire area is under #quarantine

Criticism, Warning, and Overwhelmed Tweet: Lombardy went from 0 cases to its health system being on the "brinks of collapse" in 3 WEEKS. Any large US city hospital not working right now to triple its ICU beds (and having a contingency plan for 10x) is irresponsible.

Figure 4-30: Overwhelmed Tweets Comparison

There was a sense of "we should know better" contained in tweets discussing American healthcare that wasn't present with the January situation in Wuhan. In January, many tweets seemed to be caught off guard by how quickly the situation grew so severe. In later months, the underlying question emerged of "why haven't they adapted yet?" that seemed to tamper off the *topic* of *overwhelmed.*

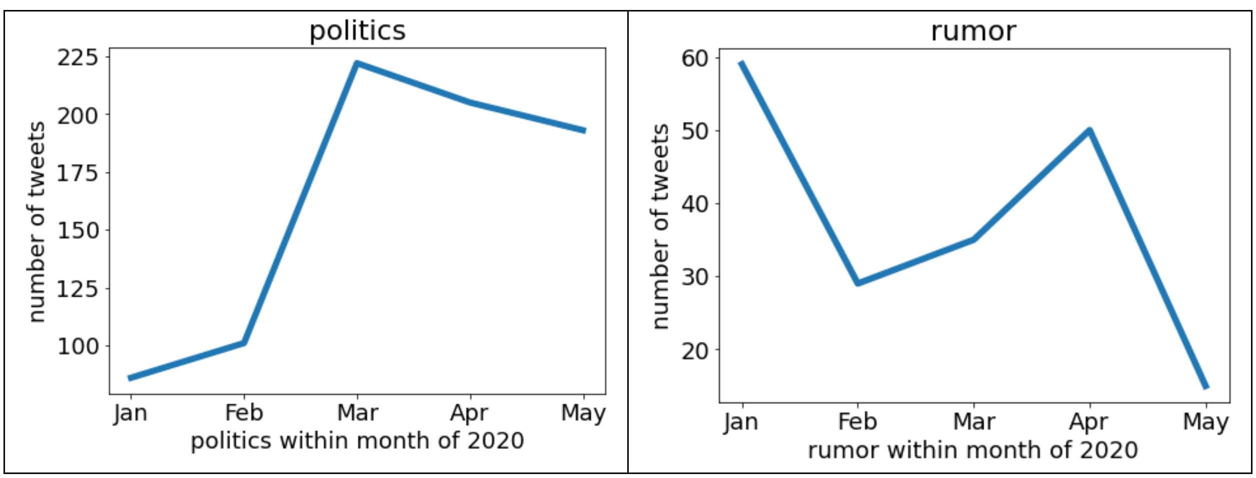

Figure 4-31: Politics (Left) and Rumor (Right) Topic Count Within Each Set

Political discussion about the virus doubled sometime between February and March, when widely followed political figures, especially the U.S president, @RealDonaldTrump, began to tweet more about the virus. *Politics* was one of the larger topics for all the sets.

Tweet from May: "We're gonna lose over 100,000 perhaps" -- Trump just moved the goalposts *again* about the projected US coronavirus death toll

Quote-Tweet from May: Was it worth it for the judges and the tax cuts?

Figure 4-32: Original Tweet and Quote-Tweet Using COVID-19 for Political Criticism

In January, most political tweets were tied to the *borders topic,* discussing how many east Asian countries were reacting to the first cases showing up. American politics, as seen in Figure 4-32 were almost entirely absent in January.

Rumor surprisingly didn't follow the same trend seen in *disinformation* and *conspiracy.* It peaked in the month of January—maybe because during this time fewer official news outlets were reporting about the virus and its origin. One way that *rumor* differed from *conspiracy* was scale. Rumors don't have to be as big as conspiracy theories and they don't have to explain as much. From the examples below in Figure 4-33, we see that rumors can relate almost anything to COVID-19. There was a lack of news presence within China and other east Asian countries that presented to a western audience. This might be one reason why *rumor* peaked in January. When a story broke in Australia or Europe, there was enough coverage to stop unqualified sources from gaining traction.

January Tweet Translated from Filipino: The edible snake or bat of the people of Wuhan, China is said to be the source of the new strain of #coronavirus. #nCoV

February Tweet: Because of confirmed #Coronavirus infection inside the apartment, residents doors are being welded shut.

March Tweet: Rumors spread on social media that snorting cocaine and drinking bleach can cure coronavirus – they can't

April Tweet: GOP's #Gohmert busted for HYPING NONEXISTENT 'MAGIC POWDER' that purportedly kills COVID-19 instantly in #Germany Dr. Wegner, the head of the German Hospital Association, [says] that literally NOTHING ABOUT Gohmert's CLAIM IS TRUE #Texas? #MOG

Figure 4-33: Examples of Rumor Spreading Between January and April

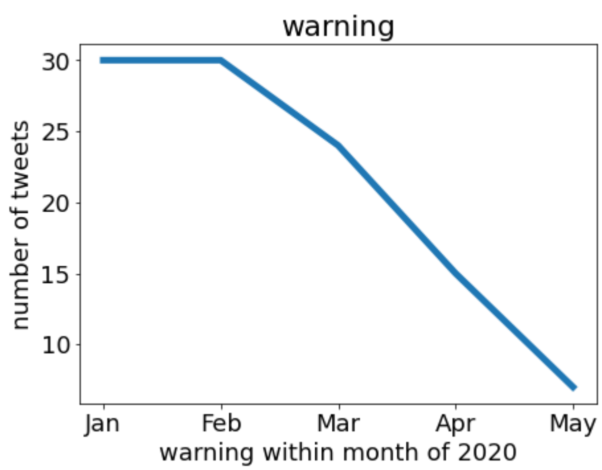

Figure 4-34: Warning Topic Count Within Each Set

Warning saw its peak usage in the early months while there was a lack of media and political discussion. This might give a hint as to why tweets grew so critical later on because from as early as January warnings were being sounded by Twitter users.

Tweet from January: #coronarovirus CDC "Centers for Disease Control and Prevention"recommends travelers avoid all nonessential travel to #Hubei Province, #China, including #Wuhan. 😟 Be Careful 🙏

Figure 4-35: Tweet Referencing an Early Warning from the CDC

Previously mentioned were the Cohen's Kappa results, which indicated better performance when the *topics* of *rumor, disinformation,* and *conspiracy* were combined. The combined graph of their outcomes was as follows:

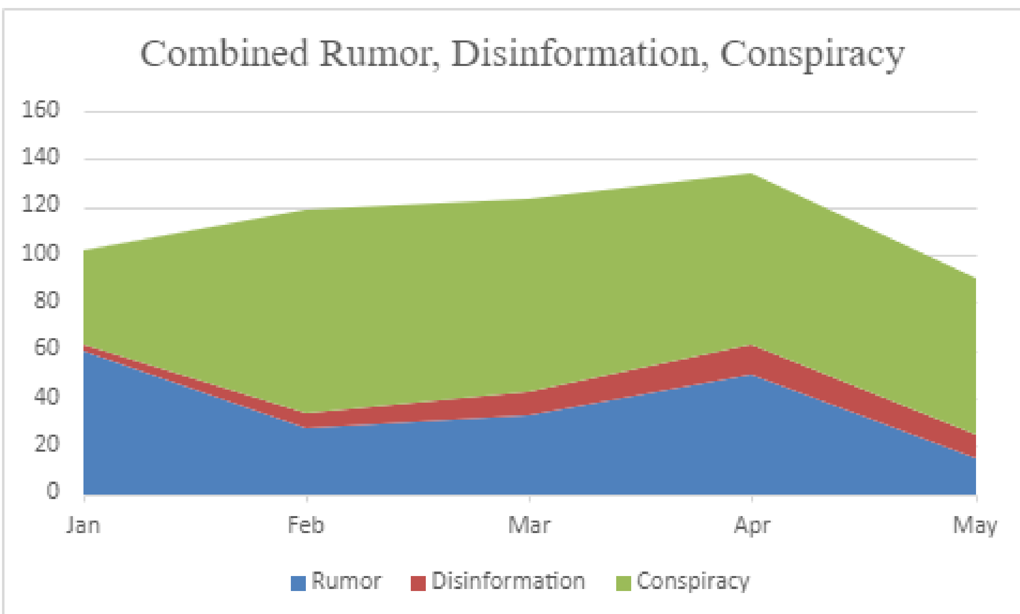

Figure 4-36: Combined Rumor, Disinformation, and Conspiracy Topics for Each Set

Combining these results presents an output like that of the *conspiracy topic*—the peak months are the same, but the overall graph shows a smaller decline in May and smaller increase in January than the original graph for *conspiracy* shown in Figure 4-8.

We should note that throughout the pandemic, Twitter itself presented an altered user interface to help point users toward expert opinions on health matters:

Figure 4-37: Twitter Interface After Searching 'COVID-19' Directing Users to Official Health Information (Captured May 2021)

However, the tweets we gathered did not contain any additional information about health guidelines, or whether tweets had been flagged for misinformation, which in many cases they could have been from a normal user's perspective. This is likely because these policies were not in place early in the pandemic when our datasets were collected. Twitter does have a medical misinformation policy that states that admins have the right to alter labeling on tweets with potentially misleading information (Twitter):

Labeling
In circumstances where we do not remove content which violates this policy, we may provide additional context on Tweets sharing the content where they appear on Twitter. This means we may:

- Apply a label and/or warning message to the Tweet
- Show a warning to people before they share or like the Tweet;
- Reduce the visibility of the Tweet on Twitter and/or prevent it from being recommended;
- Turn off likes, replies, and Retweets; and/or
- Provide a link to additional explanations or clarifications, such as in a curated landing page or relevant Twitter policies.

Figure 4-38: Twitter Medical Misinformation Labelling Policy

4.3 Share Type Summary

In addition to looking at the topic of tweets in our dataset, we also looked at how quote-tweets were related to the original tweet message that they quoted (share type). Most quote-tweets followed the *simple share type* (Figure 4-4 Left)*,* meaning they were too short to provide additional talking points to the *topic.* However, these *simple* tweets were a small majority. The *discussion share type* was usually only a few percentage points behind *simple. Agree* occurred two to three times more often than *disagree.* It seems that Twitter users are much more likely to follow and quote-tweet like-minded users, which falls into a user-behavior pattern observed on other social platforms (Del Vicario, M., Vivaldo, G., Bessi, A. et al, 2016).

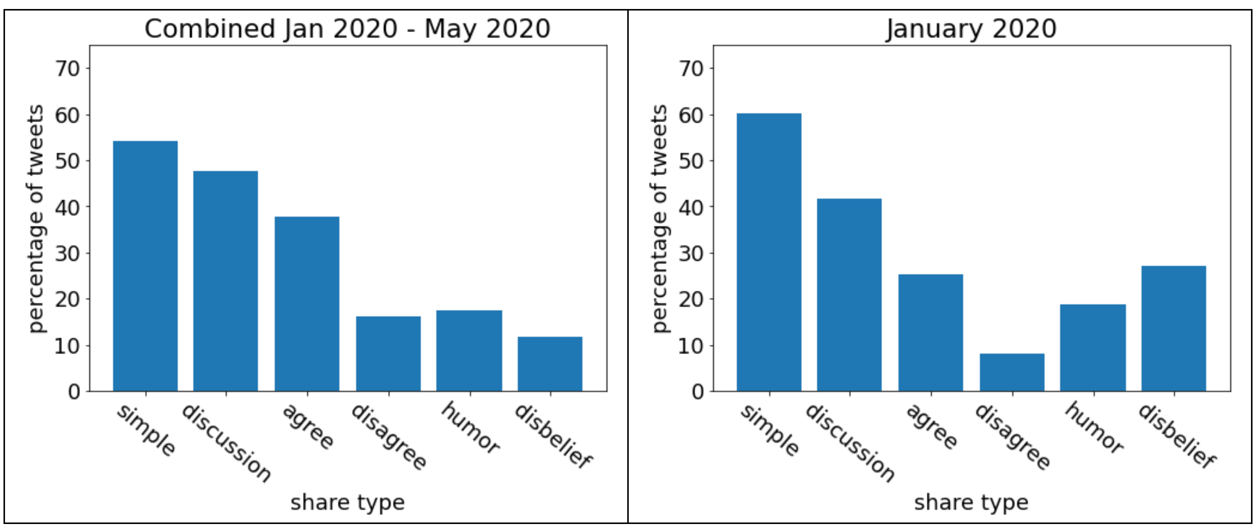

Figure 4-39: Share Type Distribution for All Sets, Jan-May 2020 (Left) and Jan (Right)

In January, the highest numbers of *simple* quote-tweets were recorded. This could have been because Coronavirus was such a new story at the time and fewer users felt adequately informed to add to the discussion about it.

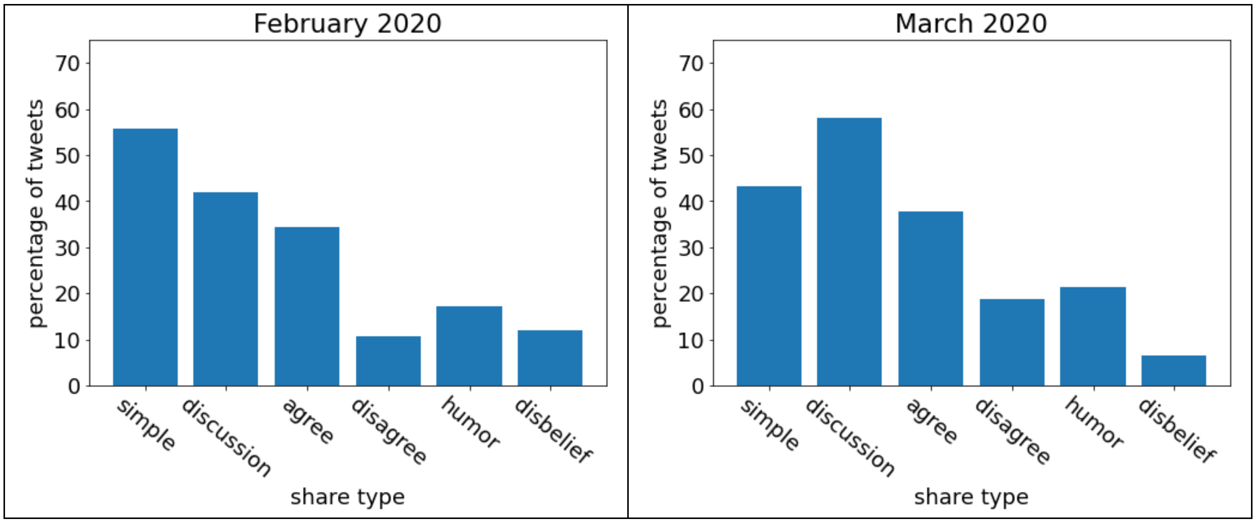

Figure 4-40: Share Type Distribution for February (Left) and March (Right) of 2020

Throughout February and March, we saw the number of *simple quote-tweets* fall, while the number of *discussion, agree,* and *disagree quote-tweets* steadily rose. This makes sense in the context of the most common *topics.* During this time the conversation about COVID-19 on Twitter shifted from "What is COVID-19?" to "Why is COVID-19 interfering with our lives?"; As *political* tweets increased, more *discussion share* type seemed to follow. Discussion was at its highest in March which was a time when Coronavirus was having a significant and growing impact on lives around the globe.

In April and May, the *agree share type* saw comparable numbers to that of the *discussion share type.* The main difference between April and May seen in Figure 4-41 was a small decline in the *humor* in May. It was interesting that *humor* and *discussion* peaked at the same time. March was a time when things like quarantine and social distancing were novel. Perhaps by April and May jokes relating to these things had worn out.

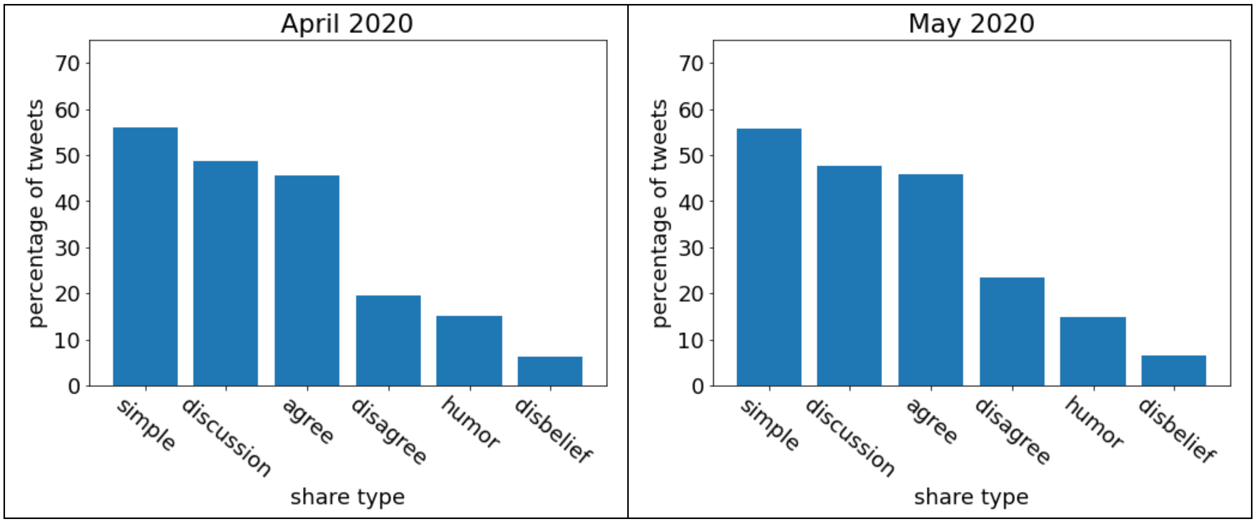

Figure 4-41: Share Type Distribution for April (Left) and May (Right) of 2020

4.4 Share Type Frequency

We expected some of the *share type* outcomes to be inverses of one another. We thought as agreement increased, disagreement would decrease, or vice versa. However, our intuition was incorrect. *Share types* actually seemed to be more closely tied to the *topics* in the tweets than they were to each other (Figure 4=42).

In Figure 4-43 *agreement* and *disagreement* both increased over time. In a word, we could describe the combined *topics of agreement* and *disagreement* as opinion. Twitter users didn't seem to voice opinion early on, especially when the primary topics of this time were related to death and demographics. In January, rather than expressing agreement or disagreement, users expressed *disbelief* as a primary reaction to virus-related tweets. As the months moved on and *criticism* became a more prominent *topic, agreement* and *disagreement* became more prominent *share types*. *Criticism* seemed to invite opinion. *Agreement* was more common than *disagreement.* About half of all tweets were marked with *simple agree* as the share type.

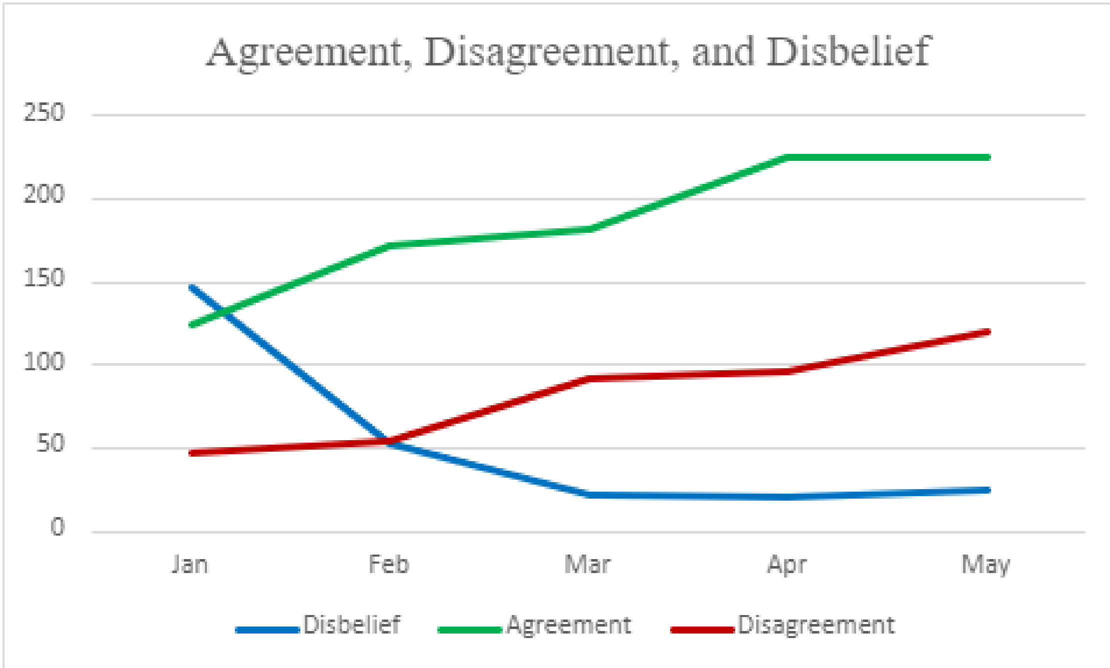

Figure 4-42: Agree and Disagree Share Type Count Within Each Set

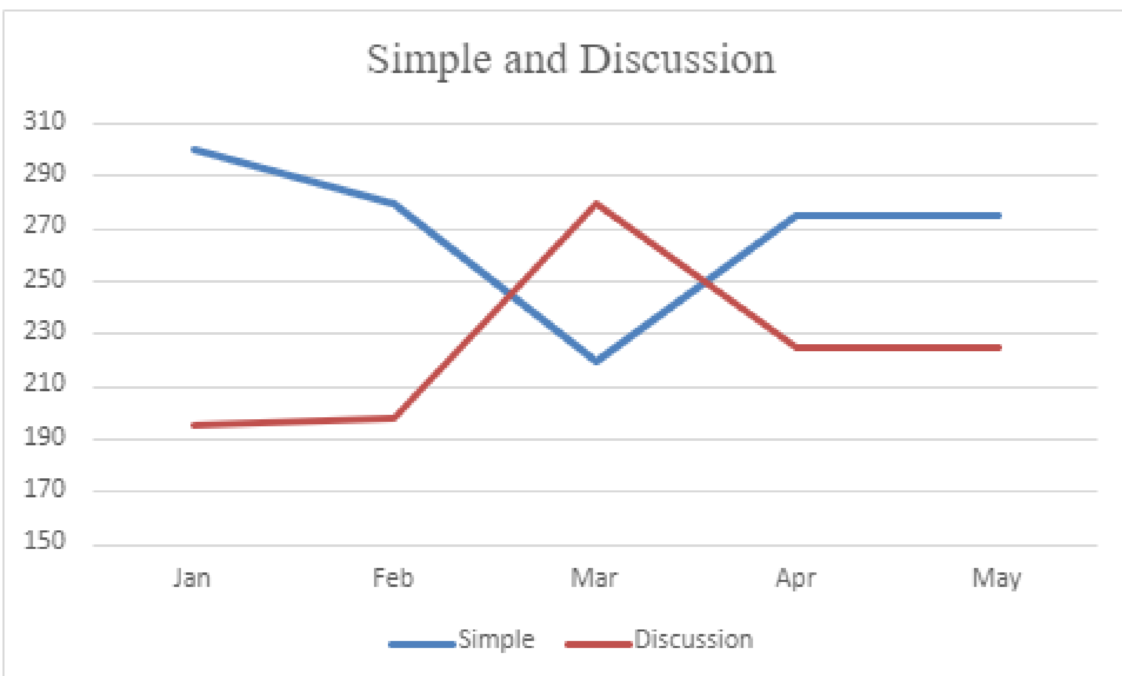

Figure 4-43: Simple (Left) and Discussion (Right) Share Types Count Within Each Set

The *simple* and *discussion* share type were inverses of one another. Our coding definitions for these types described every quote-tweet as falling into one or the other of these two categories depending on whether that quote-tweet adds additional context to the original tweet. Both *simple* and *discussion* saw relatively consistent outcomes—each maintaining approximately 45-55 percent of the tweets of each set. This is a smaller fluctuation than we observed with agreement and disbelief in Figure 4-42, but still worth noting the surge in *discussion* seen in March. Figure 4-43 shows *discussion* experiencing a single month outlier in March, while Figure 4-42 showed the shifts in *disbelief, agree,* and *disagree* as more of a trend taking place over the course of the whole study.

Discussion saw a short burst of activity that was not sustained, while *agreement* saw consistent month-by-month increase. Why was the burst in *discussion* short lived while other *share types* continued to increase? It may be because of all the *share types, discussion* required the most effort to maintain. March was the month when all the new health guidelines and mandates were implemented and weighed on peoples' minds. People saw large changes happening in their lives and went to Twitter to discuss these changes. People didn't necessarily have the strongest opinions during March compared with other months. March was the month when the novelty of COVID-19 was highest.

Disbelief was found less frequently in this data set as people grew accustomed to seeing headline news about the virus. *Disbelief* was a natural response to the shocking headlines of case numbers increasing. Figure 4-42 shows that January had more tweets marked with *disbelief* than all the other sets combined.

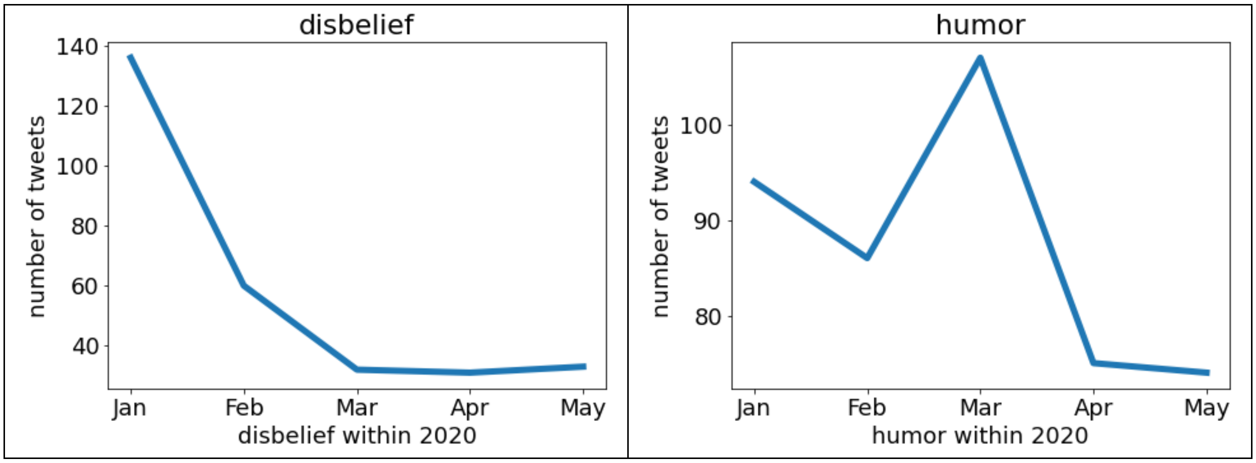

Figure 4-44: Disbelief (Left) and Humor (Right) Share Types Count Within Each Set

Humor was more consistent as a *share type* than as a *topic.* The range between highest and lowest number of occurrences was only about 30 tweets. *Humor* consistently applied to between 15 to 20 percent of the tweets in each set. It was much more common to see *humor* here in the quote-tweets than it was in the original tweets.

4.5 Language Observations

A large number of quote-tweets were written in a different language than their cited original tweets. This number was highest in January, which aligns with the time when language was most variable, as seen in Table 4-1. During this time English only made up about 41 percent of the original tweets, so it makes sense that more tweets would cross language barriers, bringing information from one language to another.

How did dialogue differ within mixed language tweets? Many of these quote-tweets offered translations of all or portions of their linked original tweets. Many used humor or commentary to respond to the original topic, almost as if the original tweet weren't written in a different language. Fewer of the mixed language tweets were marked with the *simple share type.*

This was partly because many simple tweets responding with emojis, or single words that were universally understood and didn't fall into a concrete language as easily as a full sentence would. Many mixed language tweets where the original tweet was English were news stories being translated to another language's audience. English carried the largest number of news stories by far, and if a non-English speaker wanted to link a news article often it would be written in English while the quote-tweet was translating highlights to another language.

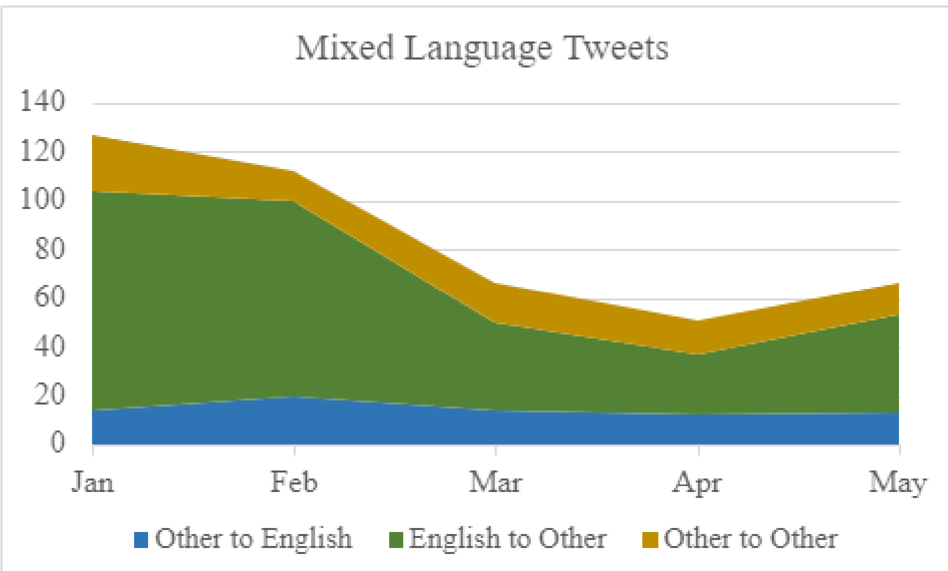

Figure 4-45: Mixed Language Tweets Within 2020

Langdetect, a simple python language prediction library, unlocked information about each collection of tweets. When applied to the original data collection of 100,000 tweets for each month (before we sampled the 500 tweets for content coding), it painted a fascinating picture where each number represents the number of 100,000 tweets estimated to be in a given language:

From the data shown in Table 4-1, we can see that English was consistently the most commonly utilized language for discussion about COVID-19. However, much of this trend is likely affected by Twitter user demographics: Twitter tends to be used more in English-speaking countries. One interesting outlier language was Thai (th), which gathered the highest number of tweets after English for the January set. Thailand, where the majority of Thai speakers live, does not share a border with China, but a survey found it holds more Twitter users than its less wealthy neighbors of Myanmar and Laos who do share a border with China (Tankovska, 2021). This survey found that many of the languages seen on this table have larger numbers of Twitter users than Thailand. Japanese (ja) users outnumber Thai users by a factor of 7:1, yet in January we saw 10 times as many Thai tweets as Japanese tweets. Similarly, Indonesian (id) and Tagalog (tl) also saw lower than expected numbers when compared with Thai. Note that Chinese is not a particularly prominent language on Twitter, nor is it found in Table 4-1, because it is banned throughout most parts of the country.

Table 4-1: Language Frequency Within Samples of 100k Tweets Produced by Langdetect

Month of 2020	January	February	March	April	May
Part of 100,000 tweets held by each language code	en 41197 th 17727 fr 6648 es 6600 id 5533 de 5266 na 2364 tl 2094 ja 1979 ar 1374 pt 1313	en 58847 es 7734 fr 5551 th 3950 ja 3432 cy 2874 it 2162 na 1591 id 1456 pt 1324 ca 1263 de 1079	en 73027 es 7495 ja 2944 th 2563 fr 1868 na 1772 tl 1622 pt 1096	en 65716 es 12564 fr 4079 pt 3546 ja 2695 na 1771 de 1122 id 1048	en 70779 es 9915 pt 3464 fr 3018 na 2053 de 1288 it 1189 et 1074
Time range of set	01/25/2020, 01:40:21 through 01/25/2020, 06:10:35	02/12/2020, 04:56:18 through 02/12/2020, 15:51:01	03/10/2020, 00:52:49 through 03/10/2020, 05:32:30	04/20/2020, 22:32:52 through 04/21/2020, 00:55:31	05/14/2020, 18:01:41 through 05/14/2020, 20:15:41

Thailand played the role of Twitter's first window into the pandemic as it was in mainland Asia. Tweets originating from Thailand seemed to respond initially with more discussion than the nearby island nations such as Japan, the Philippines, and Indonesia. Could it be that mainland Asia, countries with roads connecting to China, felt they had more cause to be worried about the spread of the disease than those countries separated by sea?

By February, the number of Thai tweets had dropped considerably, and the number of the Japanese tweets had doubled. By March English tweets made up nearly 75 percent of the set, almost double the proportion seen in January. For the last three samples taken, March, April, and May, the data remained relatively stable. Spanish, Portugese, French, and Japanese solidified into consistent percentages of each set and didn't fluctuate that much during this 3-month window displayed in the last three column of Table 4-1.

However, this data has some significant limitations. The time-of-day ranges for each of these samples are different, meaning these collections encompass tweets from differing months and hours. Still this table demonstrates the transition from many languages to primarily English. Also interesting is the amount of time it took for each of these sets to reach the mark of 100,000 tweets, which in the month of May took as little as 2 hours, but in February took nearly 11 hours. This could mean that the topic grew to become 5 times more prominent on Twitter within this timeframe, or it could also have something to do with the differing times of day these samples were collected.

4.6 Aligning with Previous Research

One of the first research teams to study quote-tweets was also a major influence on our research, Garimella et al. (2015). We wanted to expand on this work by comparing their findings to modern coronavirus examples. How has the quote-tweet feature changed in the 5 years since its release? For reference, Garimella et al. (2015) labeled a set of 500 quote-tweets all originating from @BarackObama's account where the quoting user was not @BarackObama. Simply put, users quote-tweeting the former president. The three labels they used to identify how a quote-tweet was used were *opinion*, *public reply*, and *forwarding*.

In our data sets we found that the categories of *opinion*, *public reply*, and *forwarding* were generally applicable. But the focus of our research was to learn more about the relationship between the original and quoting users. Garimella et al. laid a foundation for us to know that some quote-tweets contain opinion pieces, but we wanted to know what these opinions said about the relationship between the original and quoting users. Do they agree with one another? Does the quoting user take the original user seriously? The *opinion* field was the main area where our coding exercise would focus.

A public reply in Garimella et al.'s research was defined as a quoting user addressing the original user in their text. Forwarding was defined as a quoting user addressing other users in their text. These definitions are not as broad as the opinions we wanted to capture in our research. Forwarding behavior was localized to a smaller portion of tweets. We could detect it with a script rather than human analysis. Forwarding a tweet simply requires a '@' character followed directly by a username. By searching our sample texts for usage of the '@' character we got a rough estimate that 483 tweets utilized forwarding behavior. There may have been other usages for the '@' character other than forwarding, so we searched for the '@' character prefaced by either whitespace or the beginning of the tweet followed by alphanumeric characters. This brought the total down to 384, which was about 14 percent of our sample. This is an interesting number that helps to illustrate the intended audience of many quote-tweets. Of these 384 tweets, we counted how many contained the username of the quoted user, which behavior Garimella et al described as a public reply—because the audience is the author of the original tweet rather than a new user or group of users. 64 of the 384 tweets were public replies, where the tweet was directed to the username of the original tweet author.

Aside from public replies and forwarding behaviors, we sought to categorize a broader set of quote-tweeting behaviors relating to the opinions being shared by the authors of each quote-tweet. *Opinion* can't be categorized with a script as the connection to the original quote-is often unclear and open to interpretation. What quantitative research (i.e., sentiment analysis) might regard as negative sentiment, a closer look might reveal to be more nuanced. We see this below (Figure 4-43) where a quoted user announces an eviction due to coronavirus fears. The quoting user does not directly denounce any part of this but relates the topic to racial inequality and implies frustration. One might falsely assume that the user disapproves of impactful decisions being made due to coronavirus fears (the topic of the quoted user). However, with the context that the quoting user provides, we know that the quoting user has substituted the topic of the quoted user with the new topic of racial inequality. With this additional context we know less about what the quote-tweet is saying about coronavirus and more about this user's opinion about inequality and racial relations in Australia.

Original Tweet: 'You are no longer welcome': Malaysian student evicted by landlord over coronavirus fears

Quote-Tweet: I wonder if any English tenants in Australia would be evicted now because they flew to Bulgaria during the Mad Cow Disease outbreak. Nah. I didn't think so.

Figure 4-46: Quote-Tweet Redirecting Topic of Original Tweet

Another interesting quality about quote-tweets is their use of humor, which can indicate the quoting user's *opinion* on quoted users and topics. Conversely, sometimes a joke is just a joke and not a political stance. Some quote-tweets manage to remain completely separated from opinion, seemingly only using quote-tweets as a method of constructing a more engaging tweet. The example below shows an interesting case where the quoting user mimics the language used by North Korean nationalists. If this user were North Korean, then it might be assumed that they have a favorable opinion of the president. Certainly, the words alone would indicate that this tweet contains *positive* sentiment. But because the quoted user is talking about President Trump, we assume the quoting user is drawing a facetious comparison between President Trump and North Korean leaders.

Original Tweet: @PressSec: The President has not received COVID-19 testing because he has neither had prolonged close contact with any known confirmed COVID-19 patients, nor does he have any symptoms. Pres Trump remains in excellent health, and his physician will continue to closely monitor him

Quote-Tweet: Dear Leader is immune to any and all illness.

Figure 4-47: Sarcastic Tweet Example

4.7 Summary

In summary, this chapter describes trends within the topics and share types of COVID-19 quote-tweets that show interesting patterns. Some topics increased, others decreased, others remained stable. Upon deeper investigation we found that some topics and share types corelated with one another and tweets that were assigned a given topic were more likely to contain an associated share type. We found that language holds an interesting role within quote-tweets and COVID-19 tweets as a whole. Not all quote-tweets are written in the same language as the original. Quote-tweeting is a unique transaction that can be used to convey humor, respect, or distain in ways that aren't possible with other social media interactions. We continue to discuss the meaning of these outcomes in the following section.

5 CONCLUSION

Each of the trends discussed in the previous chapter points to a grander picture. Here, we discuss what we have learned about quote-tweets and what this online activity reveals about COVID-19. We also outline directions for future work.

5.1 As the Discussion Evolved, So Did Quote-Tweeting

COVID-19 was an interesting and ever-changing topic to study in the context of Twitter. Over the five months of data drawn from this study, we saw the perception of COVID-19 shift from a little-understood foreign health problem into a once-in-a-lifetime, global, socioeconomic crisis event. This shift in perception translated into shifts in the Twitter discourse that we saw reflected in our data.

We saw that as the discourse—the topics—became more political, the quote-tweeting share type changed. Users became more willing to voice opinions of agreement and disagreement as time went on, as evidenced by the increase of *agreement* and *disagreement share types,* as well as the increase of tweets labeled with the *criticism topic*.

Early Twitter coverage of COVID-19 carried many rumors and conspiracy theories about the virus itself. Quote-tweets during this time often used a tone of disbelief to convey neither *agreement* nor *disagreement,* but importance. Disbelief was coupled with tweets covering truthful stories as well as rumored stories. It seems that a lack of understanding about COVID-19 generated much of the discussion among Twitter users. During the early months this was manifested in conspiracy theories, some users turning to them as a source of truth, while other users turned to health officials or news organizations. Later, in April and May, this lack of understanding was manifested in the form of political arguments where many users included virus-response policy as a shortcoming of the opposing party and representatives. Would this politicization look different if the pandemic had occurred in a non-election year? Much of the arguments and discussion on Twitter seemed to be overflow from a larger argument about political ideals. Horwitz, S., & Stephenson, E. F. (2020) found much political criticism circulating in the US regarding COVID-19 relief distribution. And it's a well-known trend that disasters are often used for weight in political arguments, not just in the US (Billon, P. L., & Waizenegger, A, 2007).

5.2 What Health and Emergency Officials Can Take Away from This Research

Conspiracy theories and general disinformation were most prominent when political discussion and news coverage was low. It's important that platforms and health officials target misinformation, especially when it comes to health topics. Vraga, E. K., & Bode, L. (2020) found that correction as a rebut to misinformed posts on social media is one of the best ways to increase public confidence in health science. If health and government officials worked together early to generate a strong presence on Twitter, discussing health guidelines and current events, then we may have seen a less misinformed user base. However, in addition to conspiracy theories we recorded a wide range of reactions in the early months of this study. Many seemed to grasp the severity.

Original Tweet from* @SpeakerPelosi *January: Thank you, Chairman. #DefendOurDemocracy

Quote-Tweet from January: Do you think our Congress knows about Coronavirus in China? @SpeakerPelosi are you even concerned about the health of citizens or do you just want to ignore it until it becomes an epidemic or maybe just don't care? Do u know how many people have died from this disease???

Figure 5-1: Quote-Tweet Calling Out the Severity of COVID-19 in January of 2020

Most political figures, in the United States and elsewhere, tried to implement policy that would balance guidelines from health officials with popular opinion from constituents. This purpose of this research is not to critique the resulting health policies. However, we did pick up on a growing amount of criticism directed at political figures, health officials, and policies. Could worldwide perception of social policy have been better managed as the policies such as social distancing, work from home, and mask mandates were being implemented?

5.3 How This Applies to Future Disasters or Pandemics

It is not likely that we will see another world event such as the COVID-19 pandemic in the near future. COVID-19 is a prolonged health event continuing into present day, with economic impacts, social disruption, and localized outbreaks still greatly affecting many people across the globe. Many patterns seen in user-behavior during the pandemic are the same as those seen in other disaster events but on a longer scale of time. Compare the COVID-19 timescale with the research of Kim et al, (2016), who studied an Ebola outbreak over a period of two months comparing the relationship between news media and Twitter output. Although we didn't directly study news coverage in our research, we saw a depth to the relationship between Twitter and media coverage over a period of months that might have been hidden were it not for the length of the pandemic. Horwitz, S., & Stephenson, E. F. (2020), studied how disasters are often politicized, but the COVID-19 pandemic revealed more about this process through slow increases in criticism and rebuttal exchanges found in tweet quote-tweet exchanges. COVID-19 offers a lens to enable deeper understanding of the progression of user-behavior over the course of a disease outbreak, or disaster.

COVID-19's duration was only matched by its reach. In the age of information, no other health events have affected as many people as COVID-19. Few topics within Twitter gather interest from every demographic or nationality of user. Fewer of these topics center around an issue of public health. Within this international demographic we saw that much user behavior was universal. Every language that we translated could be categorized with one or more of our *topics*. COVID-19 helps to establish some user behavior across borders. While many nations might not have a significant number of Twitter users, COVID-19 gives us a close look into communication patterns that might work well with populations that don't have a significant online presence with social media platforms. Regardless of where a future disease outbreak occurs, COVID-19 offers a blueprint for health officials for how to communicate about disease with nearly any region.

Online activity, such as tweeting and quote-tweeting, may have been exaggerated by the pandemic. We might not see the same level of activity with future health events because COVID-19 impacted person-to-person communication so heavily. Maybe a significant portion of the conversations that would have been held in a physical space, were transplanted to a virtual space (e.g., Twitter) once people began to worry about transmission. This is a limitation. We don't know if people will communicate the same way about future health problems if the related social atmosphere has a different makeup.

Besides our focus on COVID-19, we demonstrated the utility of quote-tweets as a tool for studying the relation between users. Many tweets on Twitter, regardless of connection to COVID-19, carry hidden information in their associated quote-tweets. By looking more closely at how the authors of quote-tweets react to information found in tweets posted by health officials, future research will be better equipped to measure public opinion and react in ways that prevent the spread of misinformation.

5.4 Limitations and Future Work

The pandemic had a drastic impact on culture and lifestyle across the globe. Many tweets probably made reference to aspects of these impacts that our search algorithm was not capable of picking up. For example, if a user mentioned behavior such as wearing a mask, going to multiple stores to look for product that was in short supply, or getting takeout instead of eating inside a restaurant, would a tweet that mentioned one of these behaviors get picked up using the Twitter Search API without a direct mention of COVID-19? In future research, we could analyze a more holistic snapshot of Twitter from the same timeframe as this study and compare user behaviors.

In a similar yet opposite vein, is it possible the search algorithm was too effective? If we had a holistic snapshot of Twitter during this timeframe, we could use it to reveal if many of the tweets we analyzed, especially the political ones, were actually just typical election-year banter having less to do with the pandemic and more to do with public opinion of an elected official. One way a follow-up study like this could be conducted would be by counting political tweets captured by our study during a particular window of time and comparing that number with the number of political tweets produced by Twitter as a whole during that same window.

A common way to search in Twitter is to capitalize on the use of hashtags. Often when someone writes a tweet about something, they will include a hashtag indicating the topic of the tweet. This is useful for searching for niche topics—a good example of this is with the wildfires of New South Wales that we intended to study early on. With a wildfire, information isn't always readily available. If a regular user were to tweet about them, a hashtag was a good way to link their audience to the topic. Our concern with hashtags and COVID-19 is that a potentially high number of tweets referenced COVID-19 without hashtags. Especially in the later months of our research, COVID-19 seemed to carry universal understanding and outgrow the need for users to use hashtags. At a certain point, months after the implementation of COVID-19 health precautions, people began to talk in such a way as if COVID-19 was a new normal part of life. Why use a hashtag to link your audience to a topic that they are already deeply familiar with?

When we first started this research, we had tweets referencing wildfires in Australia and tweets referencing COVID-19. We did not envision COVID-19 growing to a world-wide scale in a matter of months. At some point it would be useful to return to smaller health issues and disasters and compare the roles of quote-tweets and discussions on Twitter in that situation with the roles quote-tweets played in this study. With the wildfires in Australia, it would have been interesting to study the behaviors of a group of users linked to a specific location, rather than study behaviors of Twitter users as a whole.

Perhaps the biggest challenge of studying Twitter as a whole was the multicultural aspect of it. The coders who categorized each tweet of this study were multilingual but not deeply familiar with each and every region where large amounts of tweets originated from. For the topic of humor, non-English tweets were more likely to be labelled with humor than if they were written in English. Was this because we were more familiar with English and therefore more capable of looking past a shallow, sarcastic jibe at the underlying political ideological criticism? We may have lost some of the deeper meanings of what people were saying when we had to refer to Google translate for a translation of the tweet.

One of the findings of this study was that tweet topics changed over time in relation to COVID-19, moving away from a health-centered discussion to discussion rooted in political opinion. This finding has a limitation because the only tweets we gathered were quote-tweets linked with original tweets. How might the topics have differed if we had categorized all tweets tagged with COVID-19, not just the quote-tweets? When setting up the tweet collection script, we calibrated it to push each type of tweet to a different database table. The database had collections of every kind of tweet, but our study only drew from the quote-tweet table. Looking at the total number of records in each of these tables, we saw that between 30-40 percent of the tweets produced tagged with COVID-19 were quote-tweets. Most of the other 60-70 percent were regular tweets, and not represented in our research on tweet topics. Future research could look at all types of tweets (regular, reply-tweets, quotes-tweets) around a disaster event and seek to extract differences and similarities between the topics found in these tweets and how people use them.

REFERENCES

Acar, A., & Muraki, Y. (2011). Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, *7*(3), 392-402

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, N.J: Prentice-Hall

Billon, P. L., & Waizenegger, A. (2007). Peace in the wake of disaster? Secessionist conflicts and the 2004 Indian Ocean tsunami. *Transactions of the Institute of British Geographers*, *32*(3), 411-427

Del Vicario, M., Vivaldo, G., Bessi, A. et al. Echo Chambers: Emotional Contagion and Group Polarization on Facebook. Sci Rep 6, 37825 (2016). https://doi.org/10.1038/srep37825

Freelon, D. (2008). ReCal2: Reliability for 2 coders. Retrieved May 06, 2021, from http://dfreelon.org/utils/recalfront/recal2

Garimella, K., Weber, I., & De Choudhury, M. (2016, May). Quote RTs on Twitter: usage of the new feature for political discourse. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 200-204)

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Horwitz, S., & Stephenson, E. F. (2020). The Politicization of Disaster Relief. *Regulation*, *43*, 4.

Hossmann, T., Legendre, F., Carta, P., Gunningberg, P., & Rohner, C. (2011, September). Twitter in disaster mode: Opportunistic communication and distribution of sensor data in emergencies. In *Proceedings of the 3rd Extreme Conference on Communication: The Amazon Expedition* (pp. 1-6).

Hughes, A. L. (2019). Site-Seeing in Disaster: Revisiting Online Social Convergence a Decade Later. In K. K. Stephens (Ed.), *New Media in Times of Crisis* (pp. 193–208). Routledge.

Kim, E. H. J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, *42*(6), 763-781.

Lu, J. (2020, September 10). Whatever Happened to ... the Instant Hospitals Built for COVID-19 Patients in Wuhan? Retrieved May 06, 2021, from https://www.npr.org/sections/goatsandsoda/2020/09/10/909688913/whatever-happened-to-the-instant-hospitals-built-in-wuhan-for-COVID-19-patients

Palen, L., & Hughes, A. L. (2018). Social Media in Disaster Communication. In H. Rodriguez, W. Donner, & J. E. Trainor (Eds.), *Handbook of Disaster Research* (pp. 497–518). Springer International Publishing. https://doi.org/10.1007/978-3-319-63254-4_24

Reuter, C., Hughes, A. L., & Kaufhold, M.-A. (2018). Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research. *International Journal of Human–Computer Interaction*, *34*(4), 280–294. https://doi.org/10.1080/10447318.2018.1427832

Schnell, A. (2020, August 26). What is kappa and how does it measure inter-rater reliability? Retrieved May 06, 2021, from https://www.theanalysisfactor.com/kappa-measures-inter-rater-reliability/#:~:text=The%20Kappa%20Statistic%20or%20Cohen's,rater%20reliability%20for%20categorical%20variables.&text=If%20the%20two%2Draters%20can,evidence%20we%20have%20reliable%20ratings.

Takahashi, B., Tandoc Jr, E. C., & Carmichael, C. (2015). Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior*, *50*, 392-398.

Tankovska, H. (2021, February 09). Twitter: Most users by country. Retrieved May 07, 2021, from https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/#:~:text=Social%20network%20Twitter%20is%20particularly,and%2017.5%20million%20users%20respectively.

Twitter. (n.d.). *COVID-19 misleading information policy*. Twitter. https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy.

Vraga, E. K., & Bode, L. (2020). Correction as a Solution for Health Misinformation on Social Media

APPENDIX A. LIST OF TRANSLATED TWEETS

Below is a table referencing the figure number of the tweet being referenced in column A and the original text of the tweet before translation in column B

Table A-1: Translated Tweets Used in Figures

Column A: Figure Number	Tweet Original Text
4-18	💴 China está imprimiendo dinero a tutiplén 📉Q1 y Q2 ya arrojan resultados negativos 📊PIB de China caerá a menos de la mitad en menos de 66 días 💰 ¿Entrará en crisis occidente x el #coronavirus y #vivaespaña ? ⏯👉 📠 "Cette crise sanitaire du #coronavirus souligne à nouveau que, face à l'absence d'une politique de relocalisation industrielle pilotée par un Etat stratège, le marché commande et l'insécurité sanitaire menace les Français." #DirectAN #QAG Ma question à @agnesbuzyn ⤵️
4-21	Família abre caixão em velório e cinco são contaminados por COVID-19 na Bahia Sobre fake news ser um crime
4-25	TH Pandémie de #coronavirus : À #Phuket (#Thaïlande) 11 nids de tortues ont été retrouvés sur des plages. Cela n'était pas arrivé depuis près de 20 ans. Les tortues sont revenues sur les plages totalement vides suite au #confinement. (BBC) #COVID19 ***Quote-Tweet:*** La nature reprend ses droits
4-26	Des médias arabes accusent #Israël et les Etats-Unis d'avoir créé et répandu le #coronavirus 📺 Voir toute l'actualité sur #i24NEWS ➡ Faudrait savoir, c'était pas Allah qui avait créé le virus pour punir les chinois?
4-33	Ang kinakaing ahas o paniki ng mga taga-Wuhan, China ang sinasabing pinagmulan ng bagong strain ng #coronavirus. #nCoV

APPENDIX B. FULL LANGUAGE TABLE

Below is a table referencing the data found in Table 4-2, which gives much more detailed numbers for each language found in the collected sets.

Table B-1: Full Language Table from Section 4

coronavirus from dbList Language statistics:		cvq2 from dbList Language statistics:		cv3-q from dbList Language statistics:		cv4-q from dbList Language statistics:		cv5-q from dbList Language statistics:	
en	41197	en	58847	en	73027	en	65716	en	70779
th	17727	es	7734	es	7495	es	12564	es	9915
fr	6648	fr	5551	ja	2944	fr	4079	pt	3464
es	6600	th	3950	th	2563	pt	3546	fr	3018
id	5533	ja	3432	fr	1868	ja	2695	na	2053
de	5266	cy	2874	na	1772	na	1771	de	1288
na	2364	it	2162	tl	1622	de	1122	it	1189
tl	2094	na	1591	pt	1096	id	1048	et	1074
ja	1979	id	1456	id	858	it	878	ca	617
ar	1374	pt	1324	it	804	ca	625	nl	596
pt	1313	ca	1263	de	715	tl	592	tr	591
so	625	de	1079	nl	533	tr	582	id	588
it	541	tl	890	ro	406	af	493	so	502
cy	496	ko	714	so	393	nl	427	tl	390
ca	488	zh-cn	713	ca	356	ro	423	af	350
zh-cn	487	tr	596	cy	288	so	395	ro	311
sw	450	hi	529	af	268	cy	390	cy	305
et	389	ru	504	ko	268	lt	268	ar	290
ro	356	so	465	tr	254	sl	220	sw	259
af	356	nl	433	sl	213	et	206	no	215
nl	340	vi	418	pl	205	sw	185	ja	183
vi	303	ro	340	et	197	da	179	vi	173
ko	288	pl	331	fi	190	pl	155	hi	173
hu	279	et	309	sw	171	fi	153	da	172
sl	225	ar	291	da	157	no	151	pl	161
fi	222	sw	255	no	154	ko	139	sv	160
tr	221	af	215	zh-cn	137	sv	137	fi	159
pl	197	no	194	vi	137	ar	119	sl	127
no	181	fi	177	sv	128	vi	109	ru	95
ru	179	sl	174	ar	127	hr	103	hu	89
sq	171	fa	162	lt	98	th	69	hr	88
		sv	161	hr	87	sq	68	ta	83

Table B-1: Continued

sv 170	da 129	hu 81	hu 61	th 77
hr 156	lt 129	sq 67	zh-cn 50	lt 58
hi 151	hr 117	sk 66	ru 49	lv 55
da 128	el 68	fa 51	hi 46	sq 54
lt 123	hu 63	ru 45	fa 45	zh-cn 48
fa 88	sq 61	lv 42	sk 39	sk 44
sk 81	cs 50	cs 24	lv 38	fa 34
cs 54	zh-tw 47	el 18	cs 15	ur 32
zh-tw 40	sk 39	hi 17	el 13	cs 23
lv 37	ta 30	ta 11	zh-tw 9	ko 22
ur 19	lv 27	ur 10	ur 9	el 19
ta 18	he 27	zh-tw 8	ta 8	mr 12
ne 17	bg 17	uk 8	mk 5	uk 12
bg 5	ne 15	ne 6	bg 3	te 11
el 5	uk 15	mr 5	uk 1	he 10
he 5	ur 13	ml 4	ne 1	bg 9
uk 4	mr 9	he 3	he 1	gu 6
mk 3	mk 4	bg 2		ne 5
gu 2	kn 2	mk 1		ml 4
mr 2	ml 1			bn 4
ml 2	pa 1			mk 3
kn 1	te 1			zh-tw 1
	bn 1			
Set time range: 01/25/2020, 01:40:21 through 01/25/2020, 06:10:35 4 H 30 M	Set time range: 02/12/2020, 04:56:18 through 02/12/2020, 15:51:01 10 H 55 M	Set time range: 03/10/2020, 00:52:49 through 03/10/2020, 05:32:30 4 H 40 M	Set time range: 04/20/2020, 22:32:52 through 04/21/2020, 00:55:31 2 H 23 M	Set time range: 05/14/2020, 18:01:41 through 05/14/2020, 20:15:41 2 H 14 M

APPENDIX C. SEARCH API SCRIPT

The following script was used to stream tweets from Twitter's API and filter tweets unrelated to COVID-19. Tweets were then sorted according to whether they were quote-tweets or original tweets and stored in appropriate databases. The full library of scripts can be found here: https://github.com/david13ean/tweepyStudy

Search API Script:

```
import tweepy
import json
import csv
import sys
import dataset
import pymongo
from textblob import TextBlob
from sqlalchemy.exc import ProgrammingError

class MyStreamListener(tweepy.StreamListener):
    db = "twitter"
    col = "test"

    def getRetweeted(self, status):
        if hasattr(status, 'retweeted_status'): return True
        else: return False

    def getGeo(self, status):
        if status.geo is not None:
            return json.dumps(status.geo)
        else: return ""

    def getCoords(self, status):
        if status.coordinates is not None:
            return json.dumps(status.coordinates)
        else: return ""

    def getText(self, status):
        try:
            text = status.extended_tweet["full_text"]
        except AttributeError:
            text = status.text
        return text
```

```
    def getQuoteText(self, status):
        try:
            quoted_text = status.quoted_status.extended_tweet["full_text"]
        except AttributeError:
            quoted_text = status.quoted_status.text
        return quoted_text

    def logQuoteTweet(self, status):
        self.col = self.col+"q"
        text = getText(status)
        quoted_text = getQuoteText(status)

        description = status.user.description
        loc = status.user.location
        name = status.user.screen_name
        user_created = status.user.created_at
        followers = status.user.followers_count
        id_str = status.id_str
        created = status.created_at
        retweets = status.retweet_count
        quoted_description = status.quoted_status.user.description
        quoted_name = status.quoted_status.user.screen_name
        quoted_user_created = status.quoted_status.user.created_at
        quoted_followers = status.quoted_status.user.followers_count
        quoted_id_str = status.quoted_status.id_str
        quoted_created = status.quoted_status.created_at
        quoted_retweets = status.quoted_status.retweet_count

        return dict(
                    description = description,
                    loc = loc,
                    text = text,
                    coords = coords,
                    geo = geo,
                    name = name,
                    user_created = user_created,
                    followers = followers,
                    id_str = id_str,
                    created = created,
                    retweets = retweets,
                    retweeted = retweeted,
                    quoted_description = quoted_description,
                    quoted_text = quoted_text,
                    quoted_name = quoted_name,
                    quoted_user_created = quoted_user_created,
                    quoted_followers = quoted_followers,
                    quoted_id_str = quoted_id_str,
                    quoted_created = quoted_created,
                    quoted_retweets = quoted_retweets
                )

    def logReplyTweet(self, status):
        self.col = self.col+"r"
        try:
            reply_status = api.get_status(status.in_reply_to_status_id,
tweet_mode="extended")
```

```
            try:
                reply_text = reply_status.retweeted_status.full_text
            except AttributeError:
                reply_text = reply_status.full_text
            reply_description = reply_status.user.description

            reply_loc = reply_status.user.location
            reply_coords = reply_status.coordinates
            reply_geo = reply_status.geo
            reply_name = reply_status.user.screen_name
            reply_user_created = reply_status.user.created_at
            reply_followers = reply_status.user.followers_count
            reply_id_str = reply_status.id_str
            reply_created = reply_status.created_at
            reply_retweets = reply_status.retweet_count

            return dict(
                        description = status.user.description,
                        loc = status.user.location,
                        text = text,
                        coords = status.coordinates,
                        geo = status.geo,
                        name = status.user.screen_name,
                        user_created = status.user.created_at,
                        followers = status.user.followers_count,
                        id_str = status.id_str,
                        created = status.created_at,
                        retweets = status.retweet_count,

                        retweeted = retweeted,
                        reply_text = reply_text,
                        reply_description = reply_description,
                        reply_loc = reply_loc,
                        reply_coords = reply_coords,
                        reply_geo = reply_geo,
                        reply_name = reply_name,
                        reply_user_created = reply_user_created,
                        reply_followers = reply_followers,
                        reply_id_str = reply_id_str,
                        reply_created = reply_created,
                        reply_retweets = reply_retweets,
                    )
        except:
            return

    def on_status(self, status):
        with open('environment.json', 'r') as myfile:
            env=json.loads(myfile.read())

        auth = tweepy.OAuthHandler(env['consumer_key'],
env['consumer_secret'])
        auth.set_access_token(env['access_token'],
env['access_token_secret'])

        api = tweepy.API(auth)
        myclient = pymongo.MongoClient("mongodb://localhost:27017/")
        mydb = myclient[self.db]
```

```
        mycol = mydb[self.col+"g"]

        self.geo = getGeo(status)
        self.coords = getCoords(status)
        self.retweeted = getRetweeted(status)

        if hasattr(status, 'quoted_status'):
            # Tweet is a quote
            logQuoteTweet(status)

        if status.in_reply_to_status_id is not None:
            # Tweet is a reply
            logReplyTweet(status)

        else:
            mycol.insert_one(
                dict(
                    description = status.user.description,
                    loc = status.user.location,
                    text = text,
                    coords = status.coordinates,
                    geo = status.geo,
                    name = status.user.screen_name,
                    user_created = status.user.created_at,
                    followers = status.user.followers_count,
                    id_str = status.id_str,
                    created = status.created_at,
                    retweets = status.retweet_count,
                    retweeted = retweeted
                )

            )


    def on_error(self, status_code):
        if status_code == 420:
            #returning False in on_error disconnects the stream
            return False

def writeToFile(fileText):
    with open('data.json', 'w') as outfile:
        # outfile.truncate(0)
        json.dump(fileText, outfile)

def main():
    with open('environment.json', 'r') as myfile:
        env=json.loads(myfile.read())

    auth = tweepy.OAuthHandler(env['consumer_key'], env['consumer_secret'])
    auth.set_access_token(env['access_token'], env['access_token_secret'])

    api = tweepy.API(auth)
    myStreamListener = MyStreamListener()
    myStreamListener.db = sys.argv[1]
    myStreamListener.col = sys.argv[2]
    myStream = tweepy.Stream(auth = api.auth, listener=myStreamListener)
```

```
    #
myStream.filter(track=["wildfire","australia","bushfire","NSWfires","NSWfire"
,"pyrocumulonimbus"])
    myStream.filter(track=["coronavirus","COVID-19"])
    # writeToFile(data)

if __name__ == '__main__':
    main()
```

APPENDIX D. GRAPHING SCRIPT

The following scripts were used to gather data and output graphs using python matplotlib.

Graphing Script:

```
import pandas as pd
import matplotlib.pyplot as plt
from pandas.api.types import CategoricalDtype
plt.rcParams.update({'font.size': 18})
df = pd.read_excel(r'C:\Users\Lenovo flex3\Downloads\cv-sample-3-combined-david-lucia.xlsx', sheet_name='All', engine='openpyxl')
dfa = pd.read_excel(r'C:\Users\Lenovo flex3\Downloads\cv-sample-3-combined-david-lucia.xlsx', sheet_name='analysis', engine='openpyxl')
df.drop(df.tail(1).index,inplace=True) # drop last n rows
dfa.drop(df.head(1).index,inplace=True) # drop first n rows
dfa.drop(dfa.tail(2).index,inplace=True) # drop last n rows
# df['set'] = df['set'].replace('coronavirus', '1-January').replace('cvq2', '2-February').replace('cv3-q', '3-March').replace('cv4-q', '4-April').replace('cv5-q', '5-May')
# df['set'] = df['set'].replace('coronavirus', 'January 2020').replace('cvq2', 'February 2020').replace('cv3-q', 'March 2020').replace('cv4-q', 'April 2020').replace('cv5-q', 'May 2020')
# df['set'] = df['set'].replace('coronavirus', 'January').replace('cvq2', 'February').replace('cv3-q', 'March').replace('cv4-q', 'April').replace('cv5-q', 'May')
df['set'] = df['set'].replace('coronavirus', 'Jan').replace('cvq2', 'Feb').replace('cv3-q', 'Mar').replace('cv4-q', 'Apr').replace('cv5-q', 'May')
for index, row in dfa.iterrows():
    df[row['All Sets']] = df['quote_category_lucia'].str.contains(row['All Sets'][0:3]) | df['quote_category_david'].str.contains(row['All Sets'][0:3])
df['share_type_lucia'] = df['share_type_lucia'].str.replace('disa', 'negative')
df['is_english'] = df['text_lang'].str.contains('en:0.9')
df['qs_english'] = df['quoted_text_lang'].str.contains('en:0.9')


# topic line graphs
foo = pd.DataFrame(columns=['set', 'topic', 'count'])
for set in df.set.unique():
    dfb = df[df.set == set]
```

```
    for index, row in dfa.iterrows():
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'cnt': dfb
[row['All Sets']].value_counts().loc[True]}, ignore_index=True)

# months = {'coronavirus':'Jan', 'cvq2':'Feb', 'cv3-q':'Mar', 'cv4-q':'Apr
', 'cv5-q':'May'}
# foo['month'] = foo.set.map(months)
for title, group in foo.groupby('topic'):
    group.plot(x='set', y='cnt', title=title, legend=False, ylabel='number
 of tweets', xlabel=title+' within month of 2020', linewidth=5, figsize=(7
,5))

    plt.xticks(rotation = 0)



# share type line graphs
foo = pd.DataFrame(columns=['set', 'share_type', 'count'])
share_types = ['wow', 'com', 'agr', 'sim', 'dis', 'hum']
share_types_lucia = ['disb', 'disc', 'agr', 'sim', 'negative', 'hum']
share_types_names = ['disbelief', 'discussion', 'agree', 'simple', 'disagr
ee', 'humor']
for i in range(len(share_types)):
    df[share_types_names[i]] = df['share_type_lucia'].str.contains(share_t
ypes_lucia[i]) | df['share_type_david'].str.contains(share_types[i])
#     print(df[share_types_names[i]])
for set in df.set.unique():
    dfb = df[df.set == set]
    for share_type in share_types_names:
        foo = foo.append({'set': set, 'share_type': share_type, 'cnt': dfb
[share_type].value_counts().loc[True]}, ignore_index=True)

# months = {'coronavirus':'Jan', 'cvq2':'Feb', 'cv3-q':'Mar', 'cv4-q':'Apr
', 'cv5-q':'May'}
# foo['month'] = foo.set.map(months)
for title, group in foo.groupby('share_type'):
    print(group)
    group.plot(x='set', y='cnt', title=title, legend=False, ylabel='number
 of tweets', xlabel=title+' within 2020', linewidth=5, figsize=(7,5))
    plt.plot(x='Jan, Feb', y=(50, 100))
    plt.xticks(rotation = 0)



# topics bar chart percentage side by side
topics_df = pd.DataFrame(columns=['topic', 'count', 'og_count'])
for index, row in dfa.iterrows():
    if (row['All Sets'] != 'tip'):
        topics_df = topics_df.append({'topic': row['All Sets'], 'count': d
f[row['All Sets']].value_counts().loc[True] / 25, 'og_count': df[row['All
Sets']].value_counts().loc[True]}, ignore_index=True)

fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
```

```
ax.set_title("Combined Jan 2020 - May 2020")
topics_df = topics_df.sort_values('og_count', ascending = False)
ax.bar(topics_df['topic'].values,topics_df['count'].values)
plt.xticks(rotation = -80)
plt.ylim(0,75)
plt.xlabel('topic')
plt.ylabel('percentage of tweets')
plt.show()

for set in df.set.unique():
    dfb = df[df.set == set]
    topics_df = pd.DataFrame(columns=['topic', 'count', 'og_count'])
    for index, row in dfa.iterrows():
        if (row['All Sets'] != 'tip'):
            topics_df = topics_df.append({'topic': row['All Sets'], 'count
': dfb[row['All Sets']].value_counts().loc[True] / 5, 'og_count': df[row['
All Sets']].value_counts().loc[True]}, ignore_index=True)

    fig = plt.figure()
#     months = {'coronavirus':'Jan', 'cvq2':'Feb', 'cv3-q':'Mar', 'cv4-q':
'Apr', 'cv5-q':'May'}
    ax = fig.add_axes([0,0,1,1])
    ax.set_title(set)
    topics_df = topics_df.sort_values('og_count', ascending = False)
    ax.bar(topics_df['topic'].values,topics_df['count'].values)
    plt.ylim(0,75)
    plt.xticks(rotation = -80)
    plt.xlabel('topic')
    plt.ylabel('percentage of tweets')
    plt.show()



# share type bar chart side by side
share_types = ['sim', 'com', 'agr', 'dis', 'hum', 'wow']
share_types_lucia = ['sim', 'c', 'agr', 'negative', 'hum', 'disb']
share_types_names = ['simple', 'discussion', 'agree', 'disagree', 'humor',
 'disbelief']
plt.rcParams.update({'font.size': 18})
for i in range(len(share_types)):
    df[share_types_names[i]] = df['share_type_lucia'].str.contains(share_t
ypes_lucia[i]) | df['share_type_david'].str.contains(share_types[i])

share_df = pd.DataFrame(columns=['share_type', 'count', 'og_count'])
for share_type in share_types_names:
#     if share_type == 'agree':
#         share_df = share_df.append({'share_type': share_type, 'count': (
df[share_type].value_counts().loc[True] - df['disagree'].value_counts().lo
c[True]) / 25, 'og_count': (df[share_type].value_counts().loc[True] - df['
disagree'].value_counts().loc[True]) / 25}, ignore_index=True)
#     else:
    share_df = share_df.append({'share_type': share_type, 'count': df[shar
e_type].value_counts().loc[True] / 25, 'og_count': df[share_type].value_co
unts().loc[True] / 25}, ignore_index=True)
```

```
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.set_title("Combined Jan 2020 - May 2020")
# share_df = share_df.sort_values('og_count', ascending = False)
ax.bar(share_df['share_type'].values,share_df['count'].values)
plt.xticks(rotation = -40)
plt.ylim(0,75)
plt.xlabel('share type')
plt.ylabel('percentage of tweets')
plt.show()

for set in df.set.unique():
    dfb = df[df.set == set]
    share_df = pd.DataFrame(columns=['share_type', 'count', 'og_count'])
    for share_type in share_types_names:
#         if share_type == 'agree':
#             share_df = share_df.append({'share_type': share_type, 'count
': (dfb[share_type].value_counts().loc[True] - dfb['disagree'].value_count
s().loc[True]) / 5, 'og_count': (df[share_type].value_counts().loc[True] -
 df['disagree'].value_counts().loc[True]) / 25}, ignore_index=True)
#         else:
        share_df = share_df.append({'share_type': share_type, 'count': dfb
[share_type].value_counts().loc[True] / 5, 'og_count': df[share_type].valu
e_counts().loc[True] / 5}, ignore_index=True)

    fig = plt.figure()
#     months = {'coronavirus':'Jan', 'cvq2':'Feb', 'cv3-q':'Mar', 'cv4-q':
'Apr', 'cv5-q':'May'}
    ax = fig.add_axes([0,0,1,1])
    ax.set_title(set)
#     share_df = share_df.sort_values('og_count', ascending = False)
    ax.bar(share_df['share_type'].values,share_df['count'].values)
    plt.xticks(rotation = -40)
    plt.ylim(0,75)
    plt.xlabel('share type')
    plt.ylabel('percentage of tweets')
    plt.show()



share_types = ['wow', 'com', 'agr', 'sim', 'dis', 'hum']
share_types_lucia = ['disb', 'c', 'agr', 'sim', 'disa', 'hum']
share_types_names = ['disbelief', 'discussion', 'agree', 'simple', 'disagr
ee', 'humor']
df['share_type_lucia'] = df['share_type_lucia'].replace('disa', 'x')
for i in range(len(share_types)):
    df[share_types_names[i]] = df['share_type_lucia'].str.contains(share_t
ypes_lucia[i]) | df['share_type_david'].str.contains(share_types[i])

for index, row in dfa.iterrows():
    foo = pd.DataFrame(columns=['set', 'topic', 'share_type', 'count'])
    for set in df.set.unique():
        dfb = df[df.set == set]
```

```
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'share_typ
e': 'Simple', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['simple
']==True)])}, ignore_index=True)
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'share_typ
e': 'Discussion', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['di
scussion']==True)])}, ignore_index=True)

    cat = CategoricalDtype(categories=['Jan', 'Feb', 'Mar', 'Apr', 'May'],
 ordered=True)
    foo['set'] = foo['set'].astype(cat)
    pivot_df = foo.pivot(index='set', columns='share_type', values='count'
)
#     pivot_df['set'] = pivot_df['set'].replace('1', 'January').replace('c
vq2', '2').replace('cv3-q', '3').replace('cv4-q', '4').replace('cv5-q', '5
')
#     pivot_df.loc[:,share_types_names].plot.bar(stacked=True, figsize=(10
,7))
    pivot_df.loc[:,['Simple', 'Discussion']].plot.bar(stacked=True, figsiz
e=(7,5), title=row['All Sets'])
    plt.xticks(rotation = 0)
    plt.xlabel(row['All Sets'] +' within month of 2020')
    plt.ylabel('number of tweets')
    plt.show()



share_types = ['wow', 'com', 'agr', 'sim', 'dis', 'hum']
share_types_lucia = ['disb', 'c', 'agr', 'sim', 'disa', 'hum']
share_types_names = ['disbelief', 'discussion', 'agree', 'simple', 'disagr
ee', 'humor']
df['share_type_lucia'] = df['share_type_lucia'].replace('disa', 'x')
for i in range(len(share_types)):
    df[share_types_names[i]] = df['share_type_lucia'].str.contains(share_t
ypes_lucia[i]) | df['share_type_david'].str.contains(share_types[i])

for index, row in dfa.iterrows():
    foo = pd.DataFrame(columns=['set', 'topic', 'share_type', 'count'])
    for set in df.set.unique():
        dfb = df[df.set == set]
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'share_typ
e': 'Agree', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['agree']
==True)])}, ignore_index=True)
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'share_typ
e': 'Disagree', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['disa
gree']==True)])}, ignore_index=True)
    cat = CategoricalDtype(categories=['Jan', 'Feb', 'Mar', 'Apr', 'May'],
 ordered=True)
    foo['set'] = foo['set'].astype(cat)
    pivot_df = foo.pivot(index='set', columns='share_type', values='count'
)
#     pivot_df['set'] = pivot_df['set'].replace('1', 'January').replace('c
vq2', '2').replace('cv3-q', '3').replace('cv4-q', '4').replace('cv5-q', '5
')
```

```
#     pivot_df.loc[:,share_types_names].plot.bar(stacked=True, figsize=(10
,7))
    pivot_df.loc[:,['Agree', 'Disagree']].plot.bar(stacked=True, figsize=(
7,5), title=row['All Sets'])
    plt.xticks(rotation = 0)
    plt.xlabel(row['All Sets'] +' within month of 2020')
    plt.ylabel('number of tweets')
    plt.show()



for index, row in dfa.iterrows():
    foo = pd.DataFrame(columns=['set', 'topic', 'lang', 'count'])
    for set in df.set.unique():
        dfb = df[df.set == set]
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'lang': 'E
nglish', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['qs_english'
]==True)])}, ignore_index=True)
        foo = foo.append({'set': set, 'topic': row['All Sets'], 'lang': 'N
ot English', 'count': len(dfb[(dfb[row['All Sets']]==True) & (dfb['qs_engl
ish']==False)])}, ignore_index=True)

    cat = CategoricalDtype(categories=['Jan', 'Feb', 'Mar', 'Apr', 'May'],
 ordered=True)
    foo['set'] = foo['set'].astype(cat)

    pivot_df = foo.pivot(index='set', columns='lang', values='count')
    pivot_df.plot.bar(stacked=True, figsize=(7,5), title=row['All Sets'])

    plt.xticks(rotation = 0)
    plt.xlabel(row['All Sets'] +' within month of 2020')
    plt.ylabel('number of tweets')
    plt.show()
```