Consumer Insights of COVID-19 Vaccines from Four Cities with Higher Percentages of African Americans to Inform Local Health Campaigns: Topic, Sentiment, and Textual Analyses

Ye Wang, Erin Willis, Vijaya Kumari Yeruva, Yugyung Lee

Table of Contents

Consumer Insights of COVID-19 Vaccines from Four Cities with Higher Percentages of African Americans to Inform Local Health Campaigns: Topic, Sentiment, and Textual Analyses

Ye Wang[1*] PhD; Erin Willis[2*] PhD; Vijaya Kumari Yeruva[1] PhD Candidate; Yugyung Lee[1] PhD

[1]University of Missouri-Kansas City Kansas City US
[2]University of Colorado Boulder Boulder US
[*]these authors contributed equally

Corresponding Author:
Ye Wang PhD
University of Missouri-Kansas City
5120 Rockhill Road
Kansas City
US

Abstract

Background: COVID-19 vaccination rates have waned across the country since the rollout in early 2021, especially among African American neighborhoods. Vaccine hesitancy is a recurring theme challenging the world's public health. Yet, months after efforts to vaccinate the world's population, we still do not have a good understanding of consumer insights about those who choose to be vaccinated and those who refuse. This also suggests that many vaccination campaigns are running on assumptions, not evidence-informed by consumer insights.

Objective: The purpose of this study is to understand consumer insight of COVID-19 vaccines in Kansas City, a city with higher percentages of African Americans, to contextualize the insight and further compare data from Kansas City with insights from three other similar-sized towns (Long Beach, California; Omaha, Nebraska; Raleigh, North Carolina) that also have higher percentages of people of color.

Methods: The researchers collected and analyzed 180,128 tweets from four cities. Triangulated methods were used to look at the breadth and depth of data to provide validity to the findings. In addition, health communication experts, informed by machine learning/deep learning topic and emotion models, conducted a textual analysis of the tweets. The strength of this study is the compilation of methods and the ways in which the data was analyzed and visualized.

Results: Four major themes about COVID were discovered from the mass of tweets: "COVID Vaccines," "Politics," "Mitigation Measures," and "Community/Local Issues." The tweets per topic and emotion category were visualized to show regional differences and longitudinal changes. Critical moments of emotional changes were detected. Textual analysis based upon data partitioned by the models identified national and local themes. Insights into strategies of appealing to residents are discussed.

Conclusions: This project's data reveal wavering relationships of trust among residents and the government and its entities. While long-term initiatives should be used to rebuild and strengthen relationships among residents in cities with higher percentages of people of color, additional attention should be given to the health messaging directed at this audience. Practical implications are offered to inform local vaccination campaigns.

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
  Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
  Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Title: Consumer Insights of COVID-19 Vaccines from Four Cities with Higher Percentages of African Americans to Inform Local Health Campaigns: Topic, Sentiment, and Textual Analyses

Abstract

Background: COVID-19 vaccination rates have waned across the country since the rollout in early 2021, especially among African American neighborhoods. Vaccine hesitancy is a recurring theme challenging the world's public health. Yet, months after efforts to vaccinate the world's population, we still do not have a good understanding of consumer insights about those who choose to be vaccinated and those who refuse. This also suggests that many vaccination campaigns are running on assumptions, not evidence-informed by consumer insights.

Objective: The purpose of this study is to understand consumer insight of COVID-19 vaccines in Kansas City, a city with higher percentages of African Americans, to contextualize the insight and further compare data from Kansas City with insights from three other similar-sized towns (Long Beach, California; Omaha, Nebraska; Raleigh, North Carolina) that also have higher percentages of people of color.

Methods: The researchers collected and analyzed 180,128 tweets from four cities. Triangulated methods were used to look at the breadth and depth of data to provide validity to the findings. In addition, health communication experts, informed by machine learning/deep learning topic and emotion models, conducted a textual analysis of the tweets. The strength of this study is the compilation of methods and the ways in which the data was analyzed and visualized.

Results: Four major themes about COVID were discovered from the mass of tweets: "COVID Vaccines," "Politics," "Mitigation Measures," and "Community/Local Issues." The tweets per topic and emotion category were visualized to show regional differences and longitudinal changes. Critical moments of emotional changes were detected. Textual analysis based upon data partitioned by the models identified national and local themes. Insights into strategies of appealing to residents are discussed.

Conclusions: This project's data reveal wavering relationships of trust among residents and the government and its entities. While long-term initiatives should be used to rebuild and strengthen relationships among residents in cities with higher percentages of people of color, additional attention should be given to the health messaging directed at this audience. Practical implications are offered to inform local vaccination campaigns.

Keywords: COVID 19, vaccine hesitancy, Tweets, deep learning, textual analysis

Introduction

At the beginning of August, the COVID-19 vaccination (two shots) rate in Missouri was 41.5%. However, due to the Delta variant, Kansas City was identified as having a lower vaccination rate and a rising number of new cases. Therefore, moving forward, it is important to increase vaccination rates in areas of Kansas City. Accordingly, local healthcare leaders have begun to pursue a more targeted approach to vaccination campaigns and address health disparities that impede Kansas City's vaccination rate progress. In particular, a few zip code areas and surrounding communities on the east side of Kansas City (64106, 64109, 64127, 64128, 64129, and 64130) are prioritized as the focal point of the Our Healthy KC Eastside Campaign, an initiative funded by Jackson County. This is because these areas have more African Americans [1] and are experiencing disproportionate health outcomes related to COVID-19 due to low vaccination rates [2]. For example, a report by KCUR [3] noted that as of May 2021, the population of zip code 64130, 88% African American, had a 15% partial vaccination (one-shot) rate.

Table 1. Demographics of the Targeted Zip Codes and Their Vaccination Rates in May 2021

Zip Code in KC	Percentage of African Americans	Partially Vaccinated (at least one dose)		
		White	Black or African American	Hispanic or Latino
64106	44%	22%	6%	17%
64109	48%	23%	18%	15%
64127	51%	11%	15%	9%
64128	81%	9%	14%	19%
64129	45%	12%	13%	9%
64130	88%	13%	15%	28%

Vaccine hesitancy is a recurring theme challenging the world's public health [4]; often, the solution lies in public health communication [5] and subsequent policymaking [6]. "Effective communication strategies are important to engage policy-makers and communities in prevention and control efforts and to increase buy-in and confidence" (p. 10) [7]. In a recent letter to the editor of *Psychological Medicine*, Barello et al. [5] call for more research to "look inside the 'black box' of vaccine hesitancy" (p. 701). This call is timely but also concerning as the U.S.'s death toll has been climbing toward 750,000 since the beginning of the pandemic last year [8]. Moreover, months after global efforts to vaccinate the world's population, we still do not have a good understanding of consumer insights about those who choose to be vaccinated and those who refuse [5]. This also suggests that many vaccination campaigns are running on assumptions, not evidence-informed by consumer insights.

To obtain consumer insights, studies have been done at national and global levels to analyze the sources of vaccine hesitancy in Australia, the United Kingdom, South Korea, etc., thanks to the copious amounts of data available on social media. Shim, Ryu, Lee, et al. [9] applied Latent Dirichlet Allocation (LDA) topic modeling to 3,509 tweets from Korea and discovered that vaccine hesitancy, including the safety of the vaccine and degree of symptoms experienced, was the top topic of discussion about COVID-19. Kwok, Vadde, and Wang [10] found from 31,100 tweets that Australian users largely supported infection control measures and refuted misinformation, but some were influenced by conspiracy theories. Cotfas, Delcea, Roxin, et al. [11] discovered a largely neutral stance of people in the United Kingdom toward lockdown measures in November 2020 based upon over 2 million tweets. Lyn, Le Han, and Luli [12] analyzed 1,499,421 tweets and revealed global trends of opinions on vaccination. Large data analysis allows researchers to make broad generalizations about specific populations.

In addition to machine learning and deep learning methods, surveys, content analysis, and qualitative analysis have been applied to provide more context and depth to understanding vaccine hesitancy. For example, a nationwide survey of 1,005 Italians conducted between November 27 and December 3, 2020, showed that confidence in the safety of the vaccine played a major role in affecting vaccine hesitancy, and collective responsibility had only marginal importance. The researchers argued that vaccination campaigns should aim at increasing individuals' trust in the effectiveness and safety of vaccines; trust is also key to discrediting conspiracy theories [5]. Conducting content analysis on a national sample of tweets, Griffith, Marani, and Monkman [4] identified significant vaccine themes hesitancy in Canada, including political skepticism, lack of knowledge, distrust toward authorities and institutions, distrust in the legal system, and the legacy of harm caused by health care institutions on people of color. Such contextual information suggests that no one can take public trust for granted, especially not health authorities [13].

Trust is monumental to vaccination campaigns targeting African Americans yet challenging to create among this population. As pointed out by Salmon, Opel, Dudley, et al. [14] the pandemic has affected African American communities more so than other communities, compounded by memory of the historical and cultural experience with experimental medical research for people of color. Additionally, trust may be put on shaky ground simply because health institutions are perceived as powers that individuals have to face up to [13]. This feeling may sometimes resonate with the suppression that African Americans have experienced historically. Additionally, the public is increasingly aware of the uncertainty of scientific knowledge and the lack of scientific consensus [15, 6]. Thus, simply depicting the vaccine refusers as uneducated may misguide communication efforts against disinformation [13]. Moreover, this one-way communication of being told "what to do" and "what to believe" may unintentionally fuel conspiracy theories: health agents are just part of the suppressive institution.

Studies on social-technical decision-making argue that two-way communication is key to increasing community engagement, or "buy-in," mainly when science is not definitive and/or experimental [15]. Two-way communication is grounded in empathy. Empathy is the ability to suspend judgment and to walk in the other's shoes. Cognitive empathy includes three stages: resonated empathy expressed empathy, and perceived empathy [16]. Resonated empathy is a vaccination campaign that listens to its target audience. Expressed empathy requires a communication campaign to recognize the unique situations and feelings of the target audience. Perceived empathy measures whether the target audience feels they are heard and understood. Empathy is associated with better therapeutic outcomes [17]. Therefore, listening could be one of the most effective tools to combat vaccine hesitancy and increase vaccination rates among historically suppressed populations.

One such channel to listen to local audiences regarding COVID-19 is social media, specifically Twitter. Kotliar [18] described social media as "sandboxes" for people to form narratives about health. In addition, people use social media to share their personal experiences and feelings about health [19], which is an influential source of psychological support and social persuasion [20]. Thus, social media becomes a unique source of "consumer insights," featuring its own "agenda" on health issues, independent of that of the news media [21].

The purpose of this study is to understand consumer insight of COVID-19 vaccines in Kansas City, a city with higher percentages of African Americans, to contextualize the insight; and further, compare data from Kansas City with insights from three other medium-sized cities (Long Beach, California; Omaha, Nebraska; Raleigh, North Carolina) that also have higher percentages of people of color. The city data comparison provides both reliability and a check on the validity of our findings. Our ultimate goal is to provide practical implications to public health communicators seeking to reach people of color.

Methods

To understand the essence of the data, triangulated methods were used in this study. Figure 1 shows the workflow of the data analysis. First, LDA topic modeling discovered underlying themes and extracted topic terms. Next, LDA+BERT assigned a tweet to a topic category, and accordingly, topics were visualized longitudinally for each city. Next, a human textual analysis was conducted to triangulate the machine analysis. Then, collected tweets were categorized into an emotion category, and results were longitudinally visualized for each city. Additionally, the sentiment trends were mapped onto the four national waves of the pandemic and key news events. The four COVID 19

waves were identified using CDC data, and the key news events were selected by human experts.

Figure 1. The workflow of the data analysis.
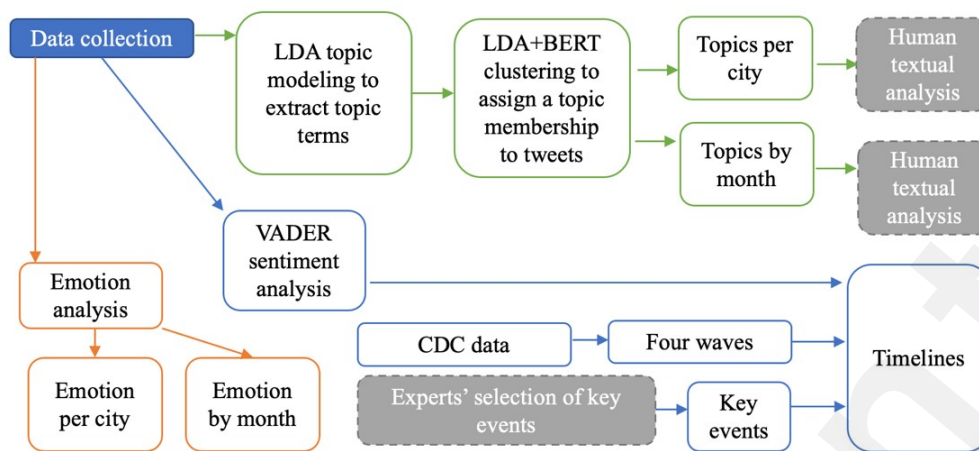

Data Collection
The data collection period is January 2020 to June 2021 since it covers the beginning, the peaks, and the subsiding of the spread of the coronavirus (see Figure 2). The cases are growing and declining captured from the graphic trend site [22].

Figure 2. The confirmed cases of coronavirus in the U.S. by month, March 2020 to June 2021.
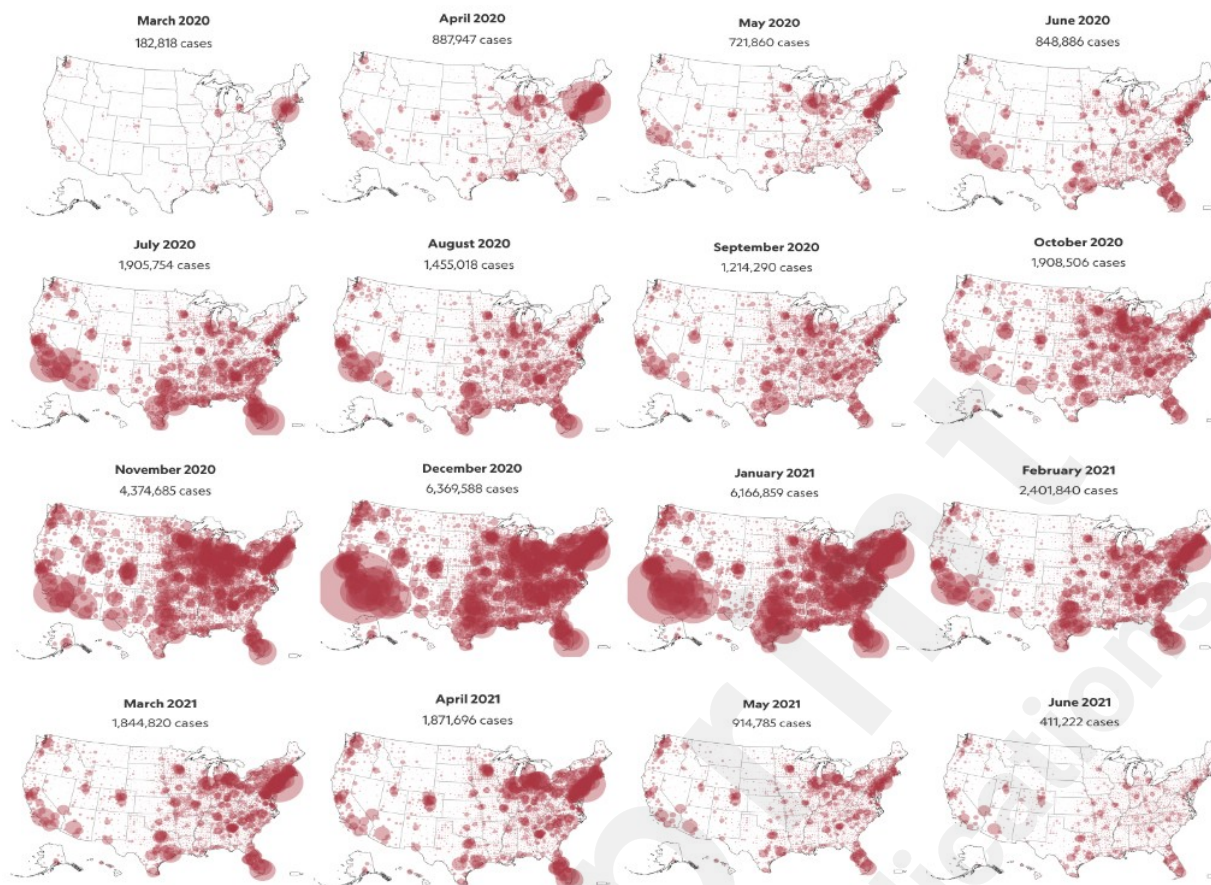
We collected data from Twitter since it is widely used in the US [23]. According to recent statistics, 21% of US adults are Twitter users [24]. Representation of Black and Hispanic users was similar or slightly higher than the general population [25]. We used Twitter API and collected tweets from four cities in the US during the period between January 2020 and June 2021: 133,844 from Long Beach, California, 8,291 from Omaha, Nebraska, 18,332 from Kansas City, and 19,661 from Raleigh, North Carolina. Our search criteria included: terms "covid" or "coronavirus" in the tweet and geo-location from the United States. Additionally, we filtered only English tweets 20 miles around the point of interest (latitude and longitude of the cities).

Figure 3. Geographic distribution of the sampled tweets.
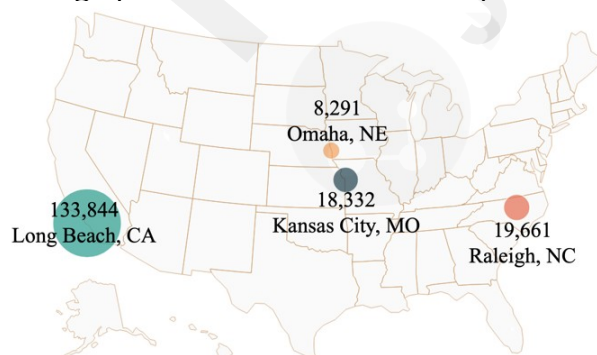
To analyze the major themes, topic modeling and human expert textual analysis were employed. Sentiment analysis used machine learning methods to gauge anger, sadness, optimism, and joy over time.

Data Preprocessing

We preprocessed the tweets to find insights. Initially, we converted the tweets to lower case, then removed special characters and URLs from the tweet so that the original text would process further. We also eliminated tweets that had single or double words that are "stop words." For example, tweets like "No" or "I did." After preprocessing, there was 0.1% of data loss from raw tweets.

Topic Modeling

A two-step approach was applied: 1) LDA topic modeling was used to identify latent semantic spaces, or in other words, underlying themes; 2) LDA+BERT clustering model was used to assign exclusive memberships for tweets.

Extracting Themes and Topic Terms

Latent Dirichlet Allocation (LDA) is a generative probabilistic model based on words, topics, and documents [26]. A document is a tweet, and a tweet consists of words. Topics are latent semantic spaces characterized by a distribution of words, and a tweet has a mixture of latent topics. Given an input corpus D ($d \in D$) with V unique words from M documents, each document d contains a sequence of n words $d = \{W1, W2, …, W_N\}$, $n \in \{1, 2, …, N\}$. Given a topic number k, $k \in \{1, 2, …, K\}$, the generative process will create documents based upon per-document topic distribution and per-topic word distribution. We used coherent analysis to determine k, and the final model had four topics.

α is the per-document topic distribution; it indicates the likelihood that a document contains topic Z_k, $k \in \{1, 2, …, K\}$. β is the per-topic word distribution. θ_d is a multinomial distribution of documents drawn from a Dirichlet distribution with the parameter α. φ_k, $k \in \{1, 2, …, K\}$ is a multinomial distribution of words in a topic drawn from a Dirichlet distribution with the parameter β. For each word position $n \in \{1, 2, …, K\}$ in a document, select a hidden topic Z_n from the multinomial distribution parameterized by θ_d. And then select W_n from φ_{Zn}.

Clustering Tweets

LDA topic models are sufficient to extract keywords of underlying themes from a collection of documents. However, the probability-based approach of LDA ignores the dependence between words, or in other words, the *context*. Thus, sometimes unigrams, bigrams, and/or trigrams are fed into an LDA model to increase the model's comprehension of context. Meanwhile, the cutoff probabilities to determine membership are arbitrary. It is hard to apply uniformly to all topics. Topics of higher coherent scores may have a different probability threshold than topics of lower coherent scores. Lastly, as LDA topic modeling assumes that topics are distributed in all documents, memberships are not exclusive, making it difficult to determine the dominant types of tweets. One approach is to extract the top probable tweets as representatives of the topic.

To address these shortcomings of LDA topic modeling, this study applied LDA+BERT clustering [27] to assign exclusive membership to tweets. First, the uncased pre-trained sentence BERT from the Hugging Face library [28] was used to generate a contextual representation of each tweet. Then, the sentence BERT embedding [29] concatenated with LDA sentence vectors were entered into a simple autoencoder with a dense layer. Next, each tweet was encoded into a latent vector space representation. Finally, K-Means clustering was applied to the latent representation and exclusively put a tweet into a topic category. We used the same number of topics k in LDA to do K-Mean clustering.

Figure 4. The LDA+BERT clustering model.
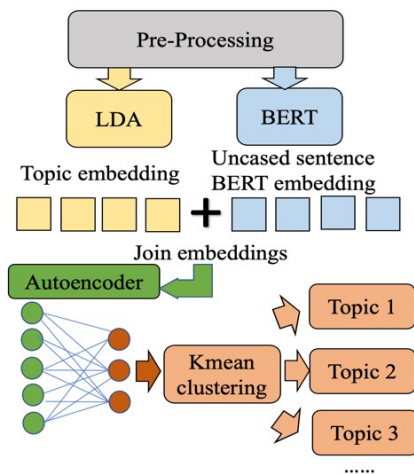
Textual Analysis of Global, Local, and Key Event Topics

Textual analysis is necessary to triangulate and contextualize topic terms from LDA topic modeling. Thus, we conducted a textual analysis of the tweets. Berger [30] describes textual analysis as a way for researchers to gather information about how others make sense of the world. The textual analysis allowed for a local perspective to emerge from the tweets above general quantification. A human expert (one of the authors) conducted a textual analysis of 4,000 randomly selected tweets from January 1, 2020, through June 30, 2021. Each tweet had a date and a city of publishing, and a topic membership assigned by the LDA+BERT model. Tweets were read and analyzed longitudinally. Tweets were examined several times: themes emerged during initial readings, and additional readings were done to explore those themes [31]. Miles and Huberman [32] instruct researchers to isolate themes "(a) that happen a number of times and (b) that consistently happen in a specific way" (p. 215). Based upon the topic category, the textual analysis contextualized topics discovered by machine analysis. Based upon location information, the authors summarized common themes of four cities and local themes of each city.

Emotion Analysis

Emotion Analysis is an extension of sentiment analysis [33]. As an extension of sentiment analysis, the emotion of the text can be predicted, such as feeling like optimism, joy, sadness, anger. Emotion recognition helps us dig deeper into users' opinions on products, services, and benefits and understand conversations more precisely using deep learning models [34]. Regarding BERT's implementation, we have used the pre-trained BERT model from the Hugging Face library. This model has different text classification tasks, including sentiment, toxicity, and emotion. The emotion task has four labels "joy," "optimism," "sadness," "anger." To label the Twitter data, we have parsed each tweet according to the model. The output from the model is the single label from the list of emotions.

Sentiment and Key Events Analysis

The purpose of the analysis of the key events is to investigate the timing of messages. First, VADER sentiment analysis was used to generate sentiment scores for each tweet. The compound sentiment scores range from –1 to 1, with (0, 1] indicating positive sentiment, [–1, 0) negative sentiment, 0 neutral. This timeline represents people's emotional responses from the four cities. The daily average sentiment scores were plotted along the timeline. Second, human experts selected 35 key events during the examined period, for example, the death of Breonna Taylor and the presidential inauguration. These key events represented the media agenda. The counts of COVID-related tweets of these key events were plotted along the timeline [35]. Third, the last timeline, depicting the infection trends, was created using the CDC's (20210 "Daily Trends in Number of Covid-19 Cases in the United States" [36].

Figure 5. Four waves identified from the CDC's daily trends' data.

Daily Trends in Number of COVID-19 Cases in the United States Reported to CDC

Third 'Wave'
roughly Nov 2020 - Jan 2021

Second 'Wave'
roughly July 2020

First 'Wave'
roughly April 2020

cases rising
at end of
Mar./start of
Apr. 2021

U.S. CENTERS FOR DISEASE CONTROL
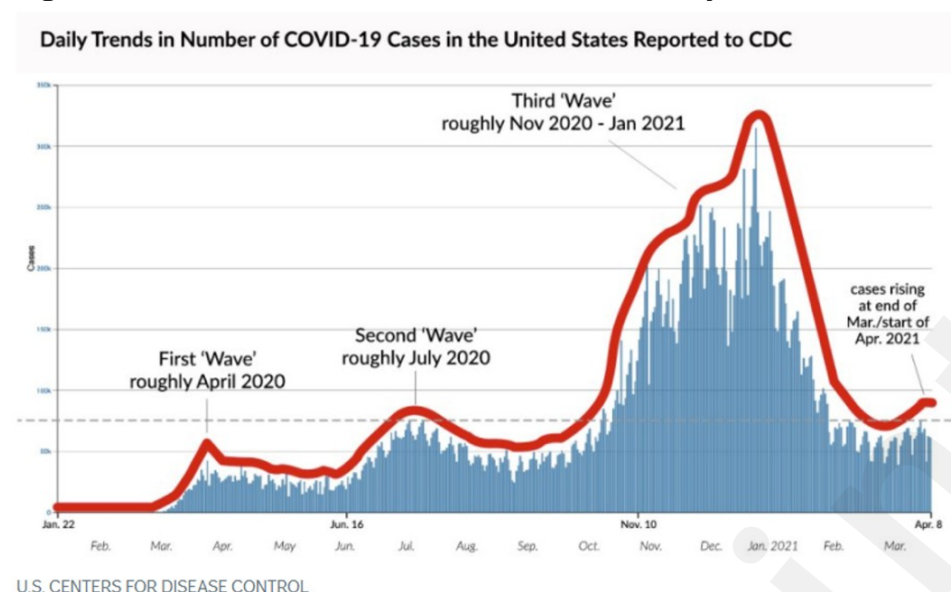
Results

This study examined consumer insight of COVID-19 vaccines from Twitter across four cities with higher percentages of African Americans for the purpose of understanding vaccine hesitancy and informing public health campaigns.

Themes and Topics

Four main topics were identified by machine analysis. Based upon the top words, we named and summarized each main topic (see Figure 4). The top words from Topic 1 show that this topic was about getting the COVID vaccine and/or tests. Topic 2 was about politics and the Trump administration, the Presidential election, COVID relief bills, and China. Topic 3 was about COVID mitigation measures, such as face masks, social distancing, etc. Topic 4 featured a focus on COVID-related community/local issues, such as students, schools, and outbreaks. Accordingly, we named Topic 1 "COVID Vaccines," Topic 2 "Politics," Topic 3 "Mitigation Measures," and Topic 4 "Community/Local Issues."

Figure 5. Top 20 Words in the descending order of topics.

Topic 1 "COVID Vaccines"	Topic 2 "Politics"	Topic 3 "Mitigation Measures"	Topic 4 "Community/Local Issues"

covid	trump	mask	case
get	coronavirus	covid	new
vaccine	american	home	coronavirus
like	china	flu	nc
im	death	year	vaccine
corona	people	stay	school
don't	country	day	student
know	president	wear	testing
test	response	hand	north
virus	business	time	county
people	pandemic	corona	state
going	via	game	carolina
getting	amp	season	health
think	america	family	amp
really	white	amp	community
thing	million	old	update
tested	say	last	read
would	worker	one	today
take care	vote	wearing	patient
cant	government	party	rate

We visualized the topic distribution in each city in 2020 and 2021(see Figure 6). The counts are the numbers of tweets from each topic. Figure 6 shows that Long Beach had the most tweets. The general pattern of mitigation measures was discussed the most across all four cities in 2020 and 2021, followed by COVID vaccines. Long Beach in 2020 had slightly more tweets about community/local issues than COVID vaccines.

The percentages reveal different patterns of topic distribution across the four cities in two years (see Figure 7). While people generally tweeted less about COVID in 2021 than in 2020, there was an increased interest in mitigation measures and the COVID vaccine and less focus on political issues in 2021 than in 2020 among the tweets.

Figure 6. Tweet counts of topics.
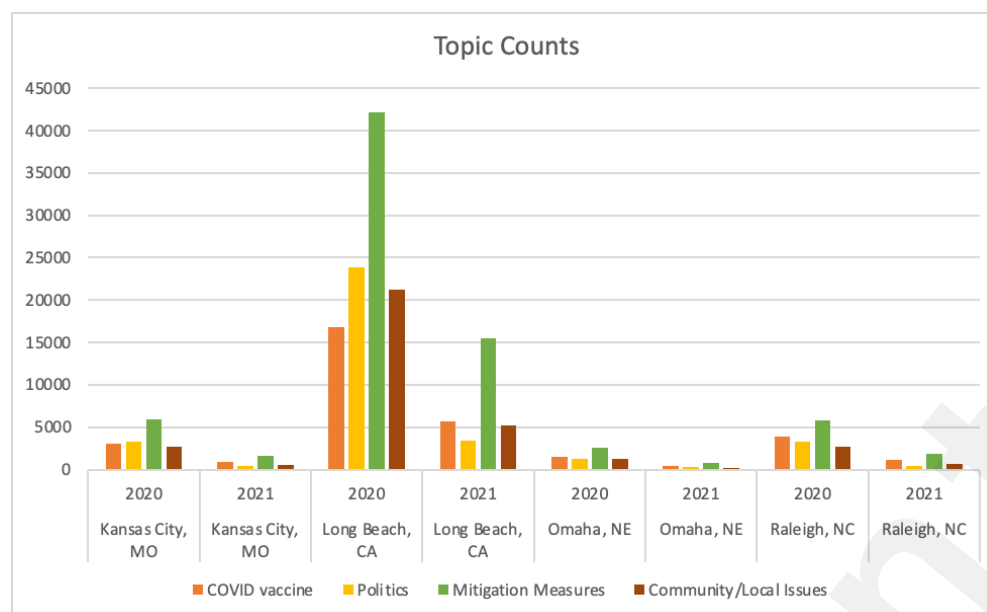
Figure 7. Tweet percentages of topics.
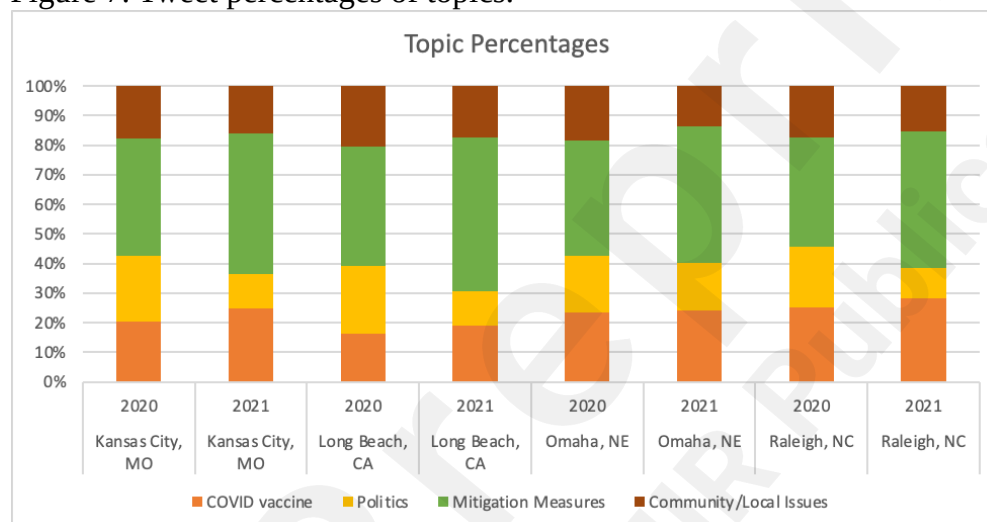
Human interpretation of local themes
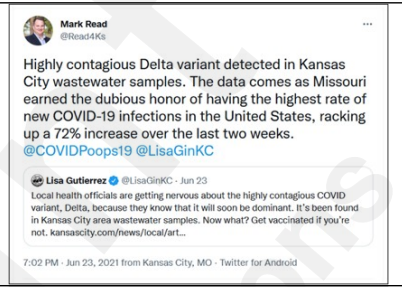
We wanted to gain insight into the local word-of-mouth so that public health campaigns might better connect with audiences and draw upon the themes that are relevant to the communities examined. The LDA+BERT clustering model partitioned the tweets into four topic clusters. These clusters became the starting points of human experts' identifying sub-themes that were unique to each location. Below is a human expert's interpretation of the sub-themes identified in each urban city and prescription for targeting local audiences in vaccination campaigns.

<u>*Kansas City, Missouri*</u>

Figure 8. Sub-themes and sample tweets in Kansas City, Missouri.

Kansas City, Missouri	Sample Tweet
Topic 1 "COVID Vaccine"	

Questioning the legitimacy and/or severity of COVID-19	**Hunter Taylor** @MrG_Taylor · · · This covid bullshit is way out of hand and way over played. 8:49 PM · Jun 26, 2021 from Overland Park, KS · Twitter for Android
Questioning how COVID-19 deaths are reported, why the vaccine is free	**TheMatt🇺🇸** @thematt81 · · · 78% of Covid patients hospitalized were obese. We don't have a covid problem, we have a weight problem. 10:50 AM · Jun 9, 2021 from Lee's Summit, MO · Twitter for iPhone
Upset that sporting events like March Madness were cancelled	**Bruce Johnson** @BhondoJohnson Replying to @cinbinbains I don't think any these post season tournaments should be played. It doesn't make sense to bring a whole bunch of teams together a week before the most important tournament during Covid. If a bunch of teams are eliminated from the NCAA tourney, it could jeopardize March Madness 11:32 AM · Mar 3, 2021 from Olathe, KS · Twitter for iPhone

Topic 2 "Politics"	
Discussing the strength of the body's immune system versus the COVID-19 vaccine	**TheMatt🇺🇸** @thematt81 · · · 78% of Covid patients hospitalized were obese. We don't have a covid problem, we have a weight problem. 10:50 AM · Jun 9, 2021 from Lee's Summit, MO · Twitter for iPhone
Debating the efficacy of the vaccine and the known/unknown side effects	**Amy Savage** @Amy45689600 · · · I'm not getting a vaccine for a virus I already got and recovered from. Why would I get igg antibodies that I already have, injected into me along with who knows what with the potential of infertility side effects? lol 9:53 AM · Jan 18, 2021 from Kansas City, MO · Twitter for Android
Attributing "getting back to normal," e.g., school, sporting events, to being vaccinated	**Spooky🧡** @bilbo813 · · · Wild that most people I know have received at least one dose of the vaccine because they either qualified or found places giving them away before tossing. I think we may have a normal late spring/summer in KC. 2:02 PM · Mar 24, 2021 from Kansas City, MO · Twitter for iPhone

Topic 3 "Mitigation Measures"	
Questioning the COVID-19 vaccine and the government's interest in increasing vaccination rates	**indigenous Xi●** @DocSmith1911 · · · Black people when you see white people coming with vaccines..RUN ‼🏃 They didn't rush to fix the water issue in Flint, Michigan did they?🤔 They can make a vaccine but can't cut a monthly check to help it's citizens impacted by this virus? #blackthoughts #tuskegeeDescendants 10:28 AM · Dec 22, 2020 from Kansas City, MO · Twitter for iPhone
Debating the politics of COVID-19 and the vaccine rollout	**Val June** @ValJuneEng · · · Replying to @cstampeen and @gates_mcfadden With respect to @gates_mcfadden , I'm glad she got her vaccine. My problem is with the government's poor rollout of it. I have family members older than her and with health issues who still don't have it. I'm allowed to be frustrated with the system. 4:51 PM · Jan 31, 2021 from Kansas City, MO · Twitter for Android
Comparing COVID-19 to the flu, incidence and severity of symptoms	**Brent Mertz** @Ult_HoneyBadger · · · Replying to @Zarcher12 and @ClayTravis A disease that infects 6-8% of the population… 99.5% of whom survive? 1% of US Pop lives in LTC facilities and they account for 30% of total deaths… in excess of 70,000 people living in nursing homes have died from COVID… flu is down 98%… only thing wrong is MSM news! 1:05 PM · Oct 28, 2020 from Overland Park, KS · Twitter for iPhone

Topic 4 "Community/Local Issues"	

Discussing the side effects of COVID-19 versus the unknown long-term effects of the vaccine	**M Robeson** @brooksidemb Replying to @WhiteHouse45 and @WhiteHouse When did they say that because just yesterday they reported that COVID infections among children are on a rapid rise. They also report they don't know what the long term health affects are for infected children. wfmz.com American Academy of Pediatrics: Number of kids testing positive for COVID-19 … While severe cases of COVID-19 in kids are rare, doctors say they carry a higher concentration of the virus and many are asymptomatic. 6:03 AM · Aug 13, 2020 from Kansas City, MO · Twitter for iPhone
Communicating the rise of the Delta variant locally and across the United States	**Mark Read** @Read4Ks Highly contagious Delta variant detected in Kansas City wastewater samples. The data comes as Missouri earned the dubious honor of having the highest rate of new COVID-19 infections in the United States, racking up a 72% increase over the last two weeks. @COVIDPoops19 @LisaGinKC Lisa Gutierrez ✔ @LisaGinKC · Jun 23 Local health officials are getting nervous about the highly contagious COVID variant, Delta, because they know that it will soon be dominant. It's been found in Kansas City area wastewater samples. Now what? Get vaccinated if you're not. kansascity.com/news/local/art… 7:02 PM · Jun 23, 2021 from Kansas City, MO · Twitter for Android
Debating different therapeutic treatments for COVID-19 and the financial interests of lawmakers and drug manufacturers	**KC Bankruptcy BSRT RRT JD - Humanity ALWAYS wins!** @kcbankruptcy Replying to @seanhannity Studies are underway to determine the efficacy in using chloroquine to treat COVID-19. If clinical data suggests this product may be promising in treating COVID-19, we know there will be increased demand for it. fda.gov/news-events/pr… 4:10 PM · Mar 19, 2020 from Independence, MO · Twitter for Android

Since the onset of the pandemic in the early months of 2020, users in Kansas City tweeted about their skepticism related to the existence and severity of COVID-19. Users questioned the legitimacy of rising COVID-19 numbers and the severity of symptoms, many comparing the virus to the seasonal flu. Users debated the necessity of the vaccine for those with healthy immune systems, especially since the long-term side effects of the vaccine are unknown. There was a misunderstanding about the purpose of the vaccine, many users believing that the vaccine could prevent contracting the virus altogether. While some users corrected this misnomer, many continued to stand by their choice not to be vaccinated and often used this point to rationalize their reasoning. Over time, users began to be vocal about the government's required lockdown and the negative effects on the economy and mental health. Users frequently debated the politics of COVID-19 and the vaccine rollout, acknowledging failures on the part of both the Trump and Biden Administrations, but also the varied messaging from government health agencies and pharmaceutical drug companies.
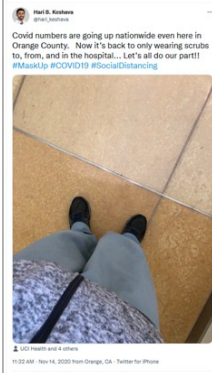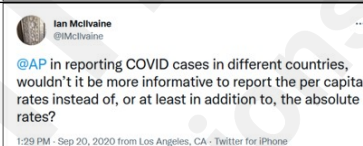
Despite rising COVID-19 numbers and increases in incidence in the Delta variant locally, many users longed for normalcy which meant a return to work and school and recreational activities, including professional sporting events. In addition, users in this area debated different therapeutic treatments for COVID-19 and the financial interests of both lawmakers and drug manufacturers. Ultimately, users expressed distrust of government and health organizations due to mixed messaging and inconsistent communication.

Long Beach, California

Figure 9. Sub-themes and sample tweets in Long Beach, California.

Long Beach, California	Sample Tweets

Topic 1 "COVID Vaccine"	
Debating the effectiveness of masks	**Britt Griffith** ✔ @britttgriffith Lockdown open up, don't wear mask must wear mask, there won't be a vaccine we must wait for a vaccine.. Wth is all I have to say. Last week this guy was saying we must stay locked down. Now he's saying open up. And ppl wonder why we question the experts. cnbc.com/2020/05/22/dr-… 9:57 AM · May 23, 2020 from Orange, CA · Twitter for Android
Discussing the severity of COVID-19 symptoms compared to the flu	**deshawn cabeza** @DeshawnCabezadk Replying to @cafe_health @cjgeorge32 and @KevinKileyCA Or is it because this is flu season, and since covid is similar to the flu, there are a lot of possible false positive test? Is it possible that if you have the flu, you can still test positive for Covid. @DrOz ?? Is it that the emergency rooms are filling up due to flu season? 12:09 AM · Dec 11, 2020 from Los Angeles, CA · Twitter for Android
Sharing the irrational fear of contracting COVID-19	**Lorraine Loiselle Cortez…proud resister** @L_Loiselle57 Replying to @DetroitSteel_ Stopped eating it months ago outta fear of catching covid which has just about taken over California. 11:28 AM · Jan 14, 2021 from Los Angeles, CA · Twitter for Android
Topic 2 "Politics"	
Discussing the origin of the virus, both in California and in the world	**james Landrigan** @California_LAX The coronavirus was made in a laboratory. To do exactly what it's doing. We are pretty sure we know who made it. Maybe a corporation. There is a lot more to this then they're telling us May God bless you and keep you alive @POTUS @realDonaldTrump @FLOTUS I am @california_lax 2:45 PM · Feb 25, 2020 from Anaheim, CA · Twitter for iPad
Debating President Trump's handling of the pandemic	**J'onn J'onzz** @erinupe was Trump right when he ran around the country in January saying it would be over after election day ? when we had deaths pile up to 600k after. was he right when he told people to take hydrochlorquine and then passed on it when he got covid ? He mishandled this thing badly **MetsJetsKnicks** @BrianA14585377 · Jun 22 Replying to @Retankerman @MittensRomney and @jimcramer Lol I am an independent. Have voted both sides. In this case, Trump was right. And that's sad that political association blinds you like that. 4:31 AM · Jun 23, 2021 from Irvine, CA · Twitter for iPhone
Discussing the potential for a COVID-19 vaccine	**JV** @DZ_49 Replying to @ABP2GoldCups and @MetcalfeDavid Like the regular flu those at higher risk are the elderly and those with compromised immune systems. Hate to sound like a conspiracy guy but that tidbit you gave(no vaccine) is why I think its being blown up. Once that vaccine hits….cha ching! 11:39 AM · Feb 26, 2020 from Bell, CA · Twitter for Android
Topic 3 "Mitigation Measures"	
Debating the politics of COVID-19 under the Trump Administration	**Pilar Marrero** @PilarMarrero Mike Pence promised a vaccine for the end of the year. Do they plan to use the Russian one (which is untested?) 9:54 AM · Aug 27, 2020 from Los Angeles, CA · Twitter for Android
Discussing the impact of COVID-19 on the stock market	**Jean-Pierre Dorléac** @spclsmthin COVID deaths are exploding, and the oh-so-precious stock market has been exposed as little more than a Rich People's Feelings Index rather than any credible metric of the health of a society or economy. 10:08 AM · Aug 7, 2020 from Los Angeles, CA · Twitter for iPhone
Connecting gun violence to COVID-19 deaths, asking for reform	**Christopher Reeve BLM** @NYCSuperMan We'll know the Coronavirus threat is largely over when the mass shootings start again. 🙂 10:00 PM · May 12, 2020 from Brea, CA · Twitter for Android
Topic 4 "Community/Local Issues"	

Sharing of local updates of COVID-19 outbreaks and safety protocols, e.g., travel limitations, closed borders, lockdowns, school closings.	
Debating how COVID-19 is transmitted	
Debating how COVID-19 numbers are reported, no consistency or standard for reporting	

In this urban area, many users began the pandemic by sharing fears of contracting COVID-19 and experiencing long-term effects or death from the virus. Some users had a fear of contracting COVID-19 and/or spreading it to their loved ones. Users discussed the severity of the virus symptoms, and some compared the virus to the flu. There was little discussion or debate related to the existence of the virus; instead, users debated the safety precautions instituted by the state government, like mask mandates or economic lockdowns. These types of discussions often turned to the politics of COVID-19 and the evaluation of administration efforts to contain the pandemic in the United States. Users also discussed where the virus originated, both in the world and in California. Users in Long Beach specifically debated President Trump's handling of the virus and the development of the COVID-19 vaccine.

As the pandemic progressed, users discussed the impact of COVID-19 on the economy and the stock market. California experienced a very strict statewide lockdown, especially in comparison to other states; due to the strict guidelines, many users raised concern about the future of California's economy and debated how the state would "bounce back." In response to rising COVID-19 numbers both in the state and the nation, users compared the virus to gun violence and the need for reform. Users reported updates of local COVID-19 outbreaks and county safety protocols that might affect others in the area. Many users debated how COVID-19 is transmitted, especially after being vaccinated. This led users to debate how COVID-19 numbers are reported, and many complained about the lack of consistency or that there was no standard for reporting.

Omaha, Nebraska
Figure 10. Sub-themes and sample tweets in Omaha, Nebraska.

Omaha, Nebraska	Sample Tweets
Topic 1 "COVID Vaccine"	

Discussing the severity of COVID-19 in comparison to the flu	**Omaha Tiger** @Omaha_Tiger Replying to @stltoday Daily reminder that influenza kills 40-50,000 people per year WITH a vaccine. 10:45 AM · Oct 10, 2020 from Omaha, NE · Twitter for iPhone
Debating how COVID-19 numbers are reported	**moknawiez12** @JamesSova17 Replying to @ZubyMusic and @erilby How the hell can all deaths in 2020 equal the same amount of deaths (10 years prior) but all from covid in 2020?..Did all other illnesses, accidents, heart attacks, old age and etc. Just not happen?..Wake Up People!!.. Or are you that Gullible!!..???..Fuck The Mask!! And Fauci!! 9:49 PM · Apr 13, 2021 from Bellevue, NE · Twitter for Android
Believing that COVID-19 and the vaccine is propaganda	**Tim Ell** @TimEll13 Just stop it please. No one believes your numbers. No one believes your fear porn. Shut up and go away.....and grab Dr Fauci on your way out will you? CDC director warns states against dropping mask mandates amid concern abo... As the spread of the coronavirus in the U.S. slows and the vaccine rollout out quickens, states have started loosening mask-wearing mandates. cnbc.com 10:57 PM · Feb 14, 2021 from Omaha, NE · Twitter for iPhone
Topic 2 "Politics"	
Upset about the interruption of various national sports conferences	**Tim H** @KnowitAle Well we won't be playing this weekend. Tournament canceled. COVID is too out of control. I am disappointed, but it's the right decision. I'm guessing they will probably be pulling youth sports (except HS) on Monday. People here can't stay out of the bar and just wear a mask 1:14 PM · Nov 13, 2020 from Omaha, NE · Twitter for iPhone
Attributing "getting back to normal," e.g., school, sporting events, to being vaccinated	**Eric Welte** @EricWelte Replying to @ooferdorfer Yes I can. People want to go to concerts and sporting events and don't want non-vaccinated people anywhere near them. If you aren't vaccinated then stay the hell home. GET YOUR VACCINE and we can all get back to normal. 1:13 PM · Apr 2, 2021 from Omaha, NE · Twitter for iPhone
Sharing vaccine hesitancy related to misinformation	**Traci Stoop** @T_Stoopendous Replying to @SteveNixon33 "I would be very skeptical of getting the vaccine for the virus in the next year"-Dr Fauci less than a year ago..... "Get the vaccine" Dr. Fauci today. 😂 what one is it, though? Ya, I'll pass. I trust my 99.9% effective immune system. I also eat right/ am active, though. 12:24 PM · Mar 1, 2021 from Council Bluffs, IA · Twitter for iPhone
Topic 3 "Mitigation Measures"	
Discussing the politics related to COVID-19 and who is to blame	**D. Erikson** @dlerikson1 Replying to @zoey_atlas and @ABC I agree he deserves credit for speeding up the vaccine process. Unfortunately, he bet everything on that alone. Downplayed the pandemic & provided misinformation. As a result, 585,000 Americans died thus far. Had no distribution plan to speak of. Biden finished the deal. 10:19 AM · May 18, 2021 from Omaha, NE · Twitter for iPhone
Discussing the inconsistency in public health messaging, e.g., masks, social distancing, vaccine, media	**Kyle Gordon** @boomkj Replying to @jayblackisfunny and @DonaldJTrumpJr I think the point here is that Joe isn't consistent in his actions. Also, the CDC says to have your mask on indoors regardless of vaccine status. PLUS... The vaccine doesn't stop the spread, it only lessens the symptoms. 12:02 PM · May 4, 2021 from Bellevue, NE · Twitter for Android

Debating COVID-19 financial relief and the economic implications	**James Sova** @JamesSova18 Biden pandemic plan Lie..$2,000 checks Lie..Unity in the government Lie..Hard on China Lie..Won't do executive orders Lie..60,000 covid deaths in the first three weeks of his presidency Truth.. The world is watching Truth..Pedophile Extortionist son Truth..America first Lie!!! 5:52 PM · Feb 9, 2021 from Bellevue, NE · Twitter for Android
Topic 4 "Community/Local Issues"	
Debating another wave of COVID-19 and mutated variants	**Brandon Adams** @BrandonAdams777 @GovRicketts please do something about this overbearing Omaha City Council and the continuous undermining of your rulings and decisions on Covid-19. They have once again extended a mandatory mask mandate city wide and it needs to stop and stop now. 7:05 PM · Oct 6, 2020 from Omaha, NE · Twitter for iPhone
Debating freedom of choice in being vaccinated in the midst of rising COVID-19 numbers	**Shirley Delair, MD, MPH** @ShirleyDelair "If as Black & Brown individuals w/access to the vaccine, we… decline…[We] are losing so many lives… What happens when COVID-19 slows down for the vaccinated community, and continues for the non-vaccinated…I made my decision. I leaned into hope. I got my shots"@JamesAbike Abike James @JamesAbike · Jan 15 As the #COVID—19 pandemic continues to ravage Black and Brown communities and as vaccine hesitancy in our communities has led to increasing disparities in vaccine uptake I felt the need to share my vaccine story. @PennMedicine #vaccinationCovid pennlisteninglab.org/stories/leanin... 4:04 PM · Feb 20, 2021 from Omaha, NE · Twitter for iPhone
Discussing the lack of confidence and trust in government agencies and pharmaceutical companies	**D. Erikson** @derikson1 Laughing only shows you being a fool. "Pfizer notably did not accept government money to develop, test or expand manufacturing capacity under Trump's Operation Warp Speed initiative to quickly find a vaccine and treatments for the disease sweeping the country." 10:38 AM · Nov 18, 2020 from Omaha, NE · Twitter for iPhone
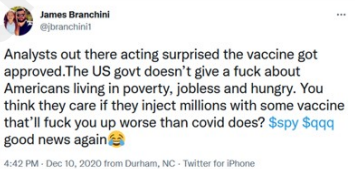
This urban area began the pandemic with a lot of doubts and hesitation related to the severity of COVID-19. Many users discussed the severity of the virus in relation to the flu, and some believed that the pandemic was political propaganda. Users in this region were especially upset that the pandemic interrupted national sporting events. Many attributed "getting back to normal" to being vaccinated and encouraged others to choose the vaccine when it was their turn. However, there were many users who shared vaccine hesitancy due to misinformation. Hesitancy then led to political debate and who is to blame for the pandemic. Interesting in this localized data is the attention on the inconsistent public health messaging related to masks, social distancing, and the vaccine. While many pointed to the media, social media was acknowledged as a major problem in spreading misinformation.
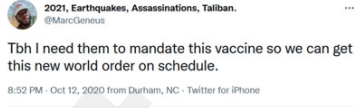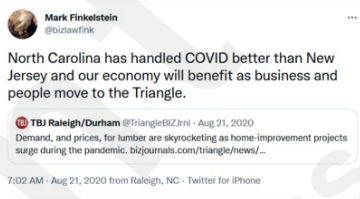
While the necessity of economic lockdown was debated among these users, many looked ahead and wondered what was next for their region. Users argued whether another wave of COVID-19 would hit and if there would be mutations of the virus. Fear of another wave led users to discuss vaccine mandates and whether Americans should have the freedom of choice during this public health crisis. In addition, vaccine mandates and emergency approval by the FDA fueled users' lack of confidence and trust in government agencies and pharmaceutical companies, questioning why they wouldn't have the right to choose.

Raleigh, North Carolina

Figure 11. Sub-themes and sample tweets in Omaha, Nebraska.

Raleigh, North Carolina	**Sample Tweets**
Topic 1 "COVID Vaccine"	

Understanding the effects of the national lockdown	
Discussing the necessity and effectiveness of safety precautions	
Discussing the severity of COVID-19 in comparison to the flu	
Topic 2 "Politics"	
Debating the effectiveness of the COVID-19 vaccine	
Discussion of "vaccine refusers" and freedom of choice	
Weighing the risks of contracting COVID-19 or the unknown long-term effects of the vaccine	
Topic 3 "Mitigation Measures"	
Discussing immigration and its contribution to the COVID-19 numbers/deaths	
Discussing violence against Asians and the dangerous rhetoric used by leaders	

Debating the handling of COVID-19 and the vaccine rollout by both Trump and Biden Administrations	
Topic 4 "Community/Local Issues"	
Debating the potential of the vaccine passport mandate	
Discussing the effects of COVID-19 on the economy	
Discussing COVID-19 treatments and precautions	

To begin the pandemic in this part of the country meant that many users had questions about the severity of COVID-19 and how it differed from the flu. As more state governors chose to close down local economies, users had plenty to say about it. Users debated the economic effects of the national lockdown on the local economy and the necessity of safety precautions when numbers were relatively low. As the potential for a vaccine became a reality, users debated whether or not the vaccine would be effective and in what capacity. Those who chose not to get the COVID-19 vaccine often pointed out that they had been vaccinated, e.g., MMR, tetanus, and were (usually) in favor of vaccines, but in this case were opting out due to the "warp speed" of the development and production of the vaccine and the lack of proper scientific trials to guarantee its safety.

Often, any discussion of COVID-19 turned to politics. In the case of Raleigh, users debated the handling of the virus and the vaccine rollout by both President Trump and President Biden. Users in this area also discussed immigration and whether immigrants contributed to the United States' pandemic numbers and deaths. Many users discussed the increasing violence against Asians and the dangerous rhetoric used by political leaders. Users seemed to be unhappy with their local government officials and the decisions being made related to COVID-19. The local economy was a popular topic of conversation; many users shared concern over local businesses and their ability to rebound after the lockdown. Users discussed different COVID-19 therapeutic treatments and the effectiveness of safety precautions in the midst of rising numbers nationally. Finally, many debated the potential of a vaccine mandate and whether or not vaccine passports should be used.

Emotions

Figures 12 & 13 show that there were more tweets with negative emotion than positive emotion. People were more joyful and optimistic in 2021 than in 2020. The most positive city was Raleigh, NC. The most negative city was Omaha, NE, in 2020 and 2021. The angriest city was Long Beach, CA in 2020, and Omaha, NE, in 2021.

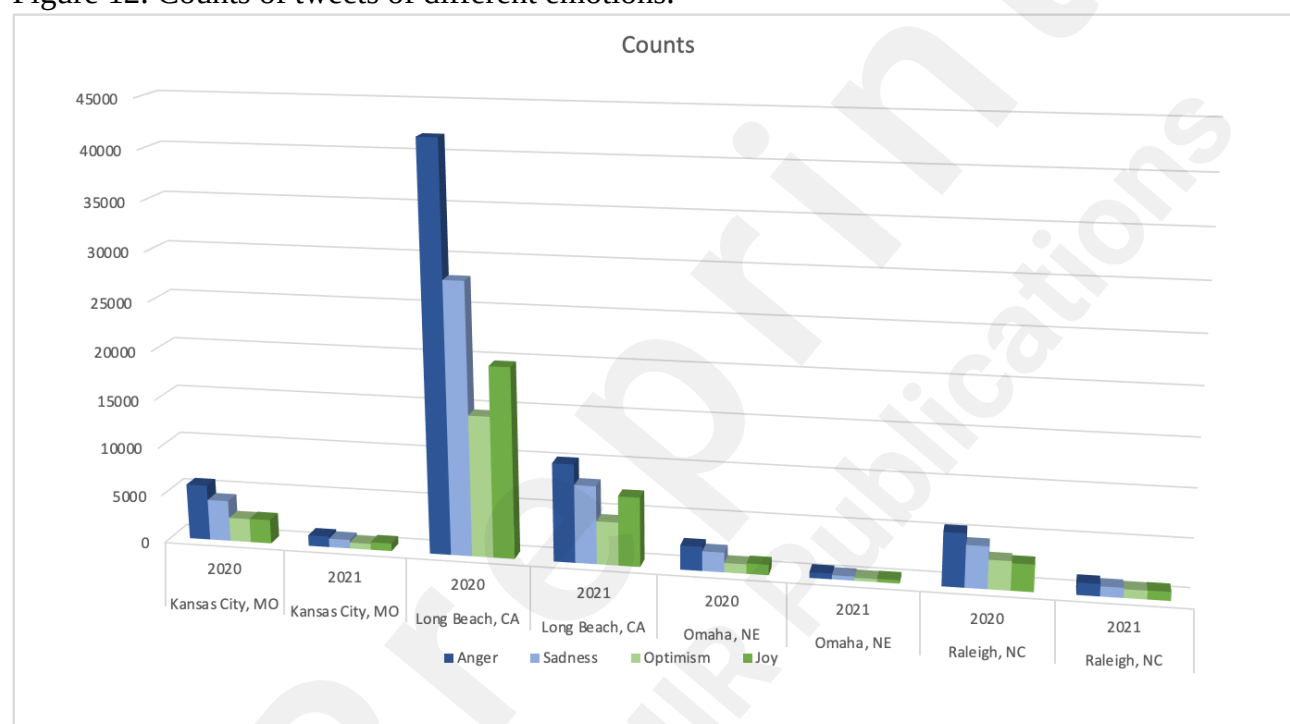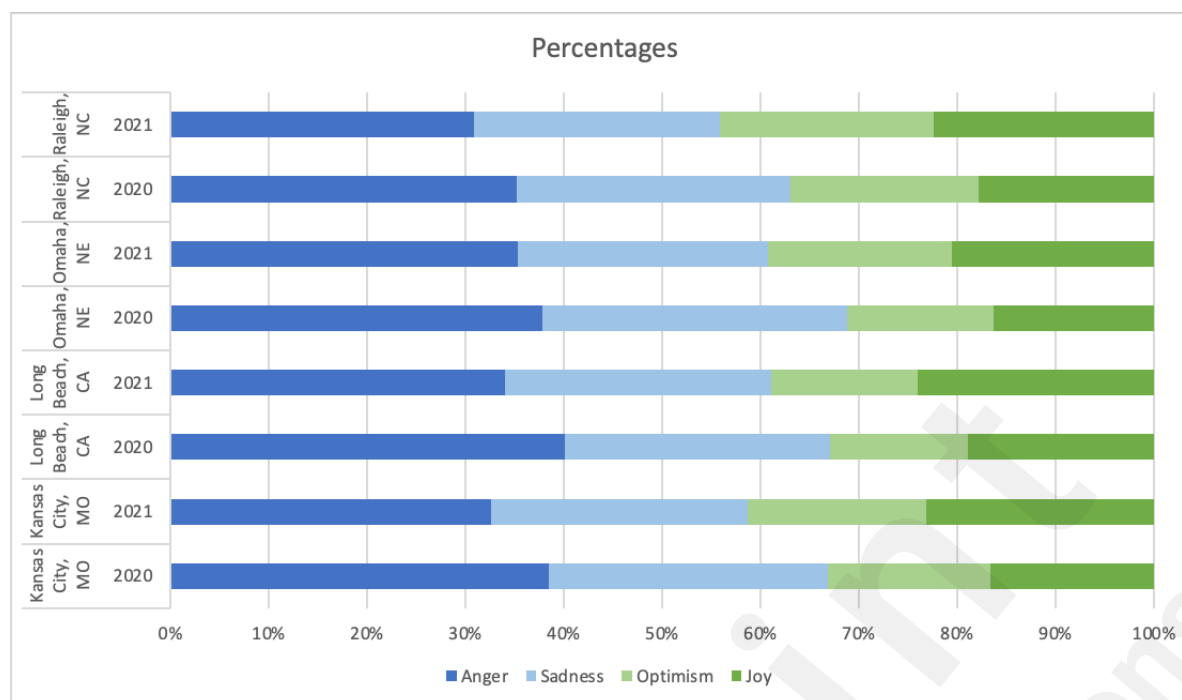Figure 12. Counts of tweets of different emotions.

Figure 13. Percentages of tweets of different emotions.

We also analyzed sentiment change over time in each city. Figures 14-17 show the change of tweet counts over time. February 2020, March 2020, September 2020, and May 2021 were the definitive moments in the sentiment distribution over time. All four cities had a sudden sharp increase in tweets about COVID in February 2020. This peak marked the beginning of the pandemic and featured a strong negative emotion. After a significant drop in COVID tweets in March 2020, there was an increase in April, which did not reach the peak amounts of tweets in March 2020. September 2020 marked a significantly larger drop in tweets in Long Beach, CA, than in other cities. May 2021 marked a significant increase in the tweets, coinciding with the rollout of the vaccine, and the number of tweets sharply fell in June 2021.

Figure 14. Month by month emotions counts, Kansas City, MO, 2020-2021.
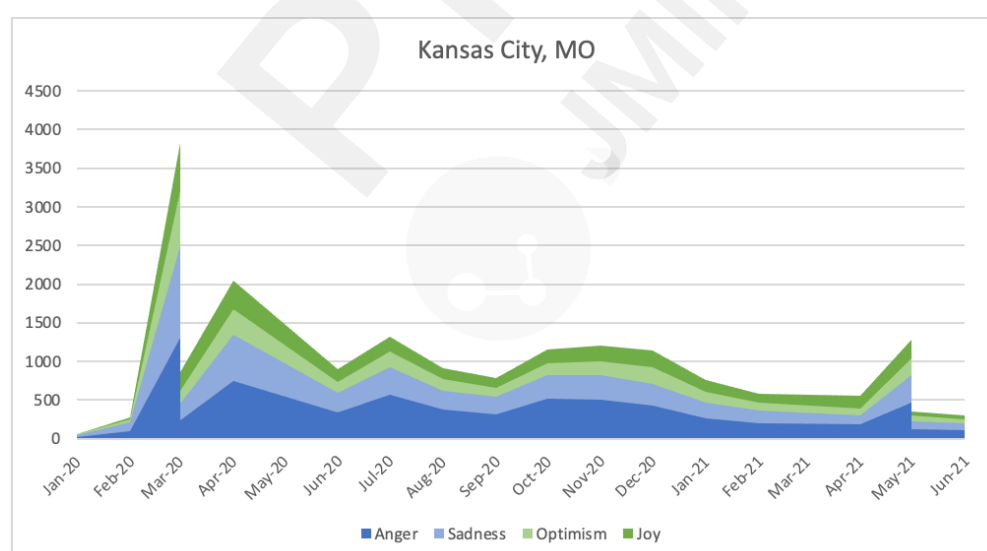
Figure 15. Month by month emotions counts, Long Beach, CA, 2020-2021.
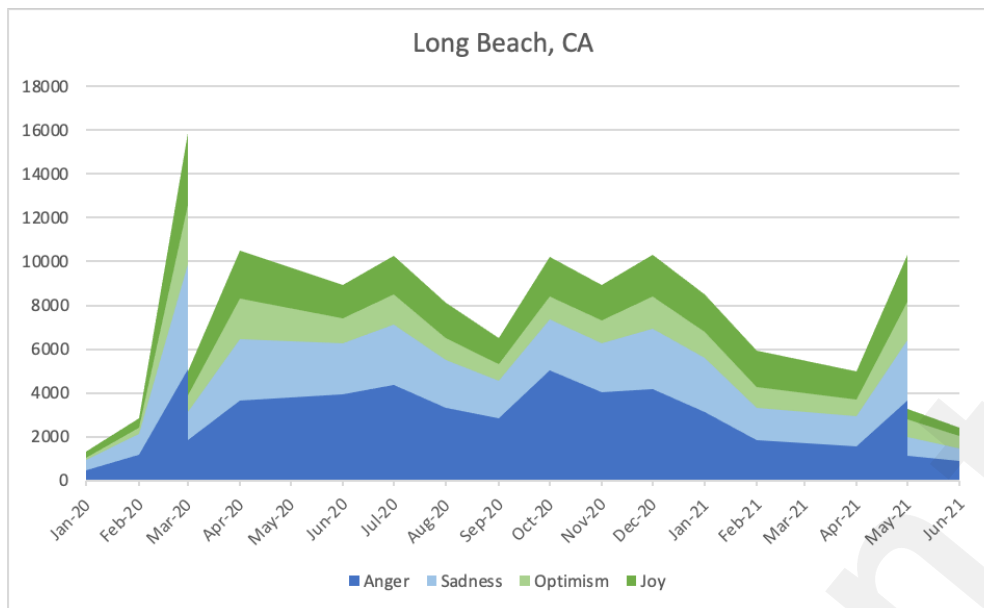
Figure 16. Month by month emotions counts, Omaha, NE, 2020-2021.
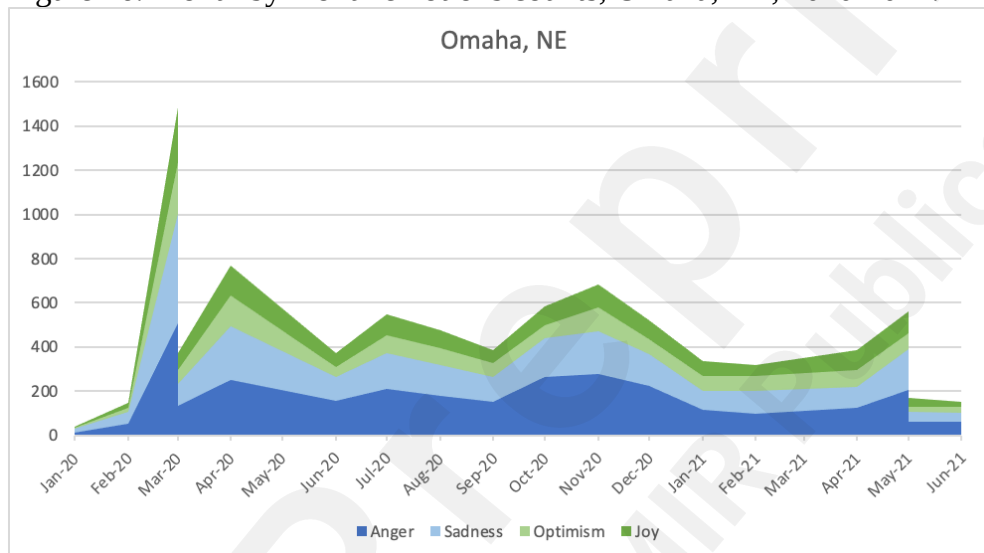
Figure 17. Month by month emotions counts, Raleigh, NC, 2020-2021.
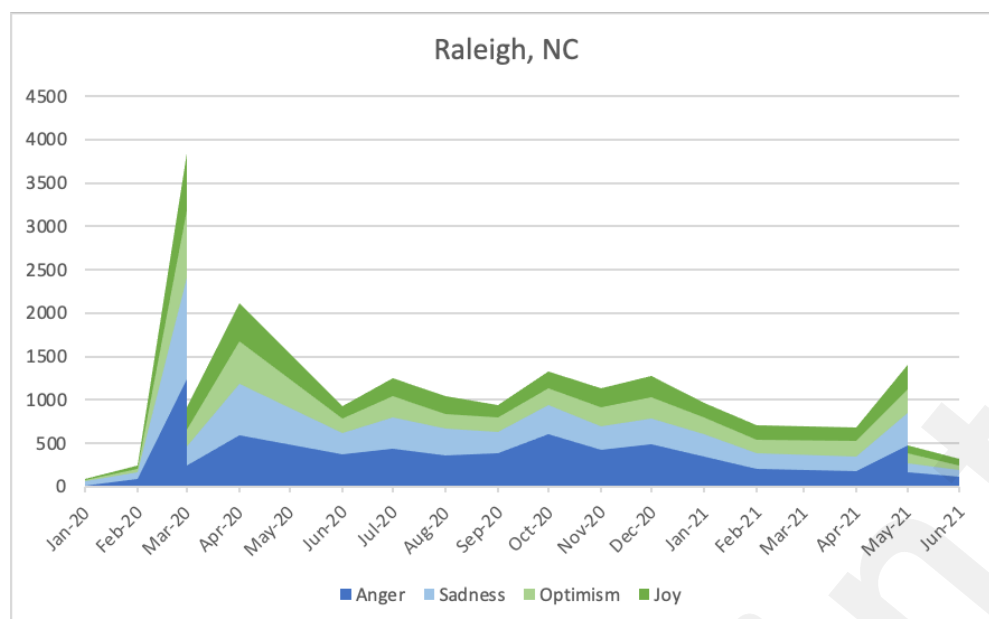
Figure 18 uses the percentages of emotion to indicate the change of moods among people who shared about COVID and vaccines on Twitter. In Kansas City, people suddenly became more joyful and optimistic in April 2021, when the COVID vaccine began to roll out. A similar surge of positivity in 2021 happened in February in Long Beach, CA, and April in Raleigh, NC, and Omaha, NE.

Figure 18. Month by month emotions percentages, 2020-2021.

Combining the above two groups of charts, we can see that April and May in 2021 were the two critical months for the vaccination in Kansas City, MO, Omaha, NE, and Raleigh, NC. February and May 2021 were the two critical months in Long Beach, CA. There was a clear sign of optimism in the first month when vaccines were largely available to the public. However, this momentum was quickly replaced by an uproar of anger in Ma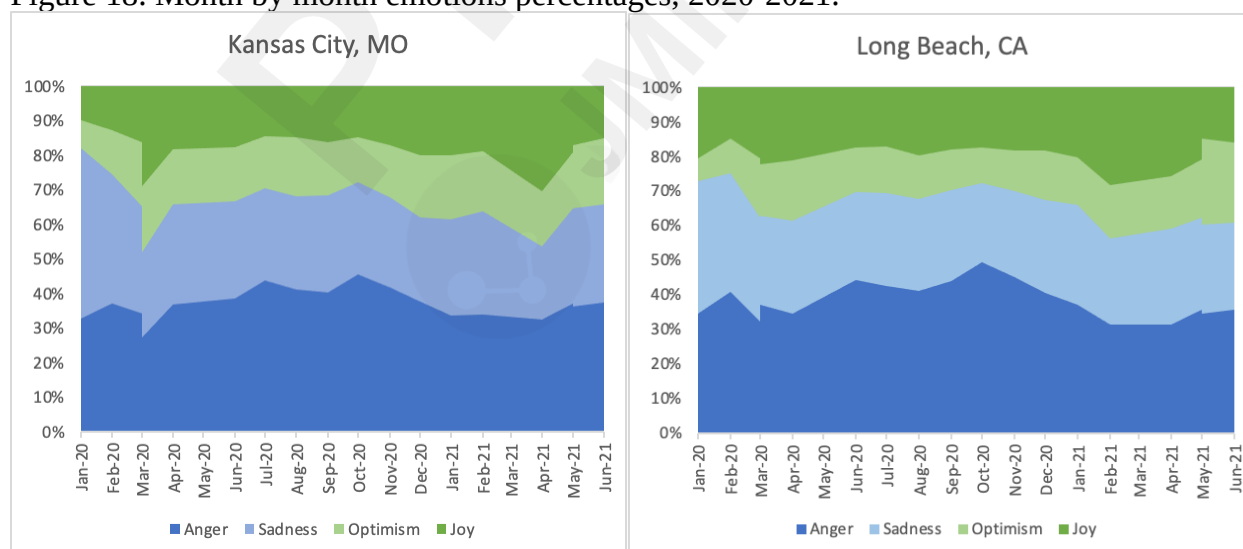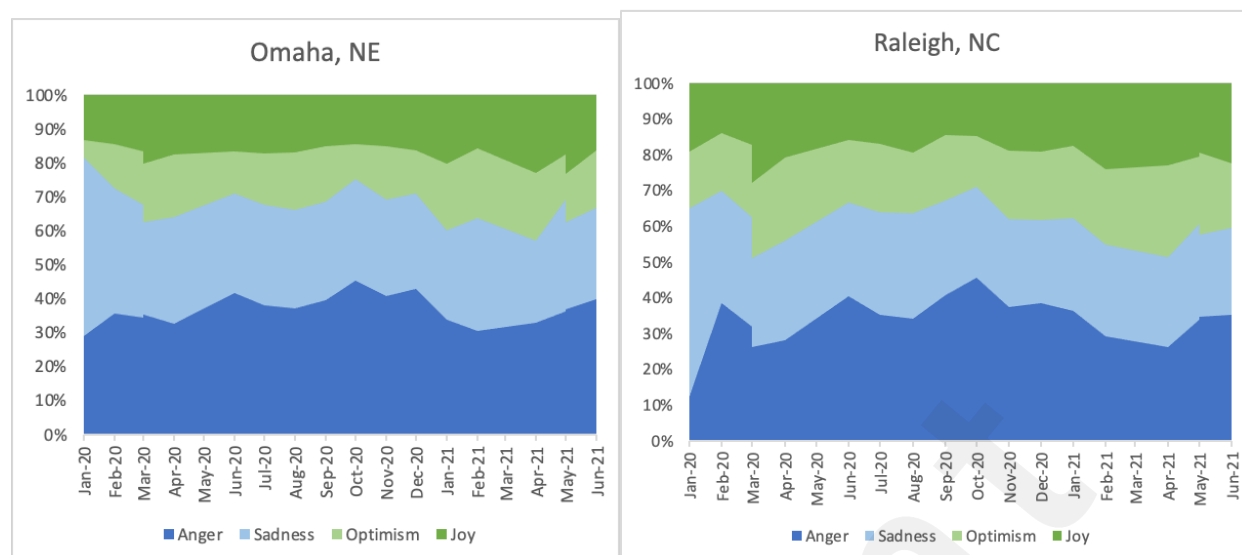y 2021. People's emotion quickly changed from optimism and joy to anger just within the first month when people were allowed to get vaccines in Kansas City, MO, Omaha, NE, and Raleigh, NC. This change was more gradual and spread over three months in Long Beach, CA.

Sentiment and Key Events

In Figure 19 below, the Twitter data's sentiment is overlaid with the number of COVID-19 infections during the course of the pandemic, including several key events that took place in society. Often, data is reported without the context of everyday happenings and a reminder of what else was going on during that time. Twitter is a known outlet for trending topics and for engaging with other users. Therefore, it is important to examine the data to see what else influenced these Twitter conversations. Due to states' stay-at-home public health orders, many people were a captive audience to events and took to social media to discuss. The CDC data (see the red timeline of the 7-day moving average), the four waves of infections during the measuring periods were (1) January 1 to May 31, 2020, (2) June 1 to September 11, 2020, (3) September 12, 2020, to March 12, 2021, and (4) March 13 to June 23, 2021, are displayed, and the green dotted vertical line represents the end of a wave.

Figure 19. Sentiment and daily COVID case numbers displayed over the pandemic.
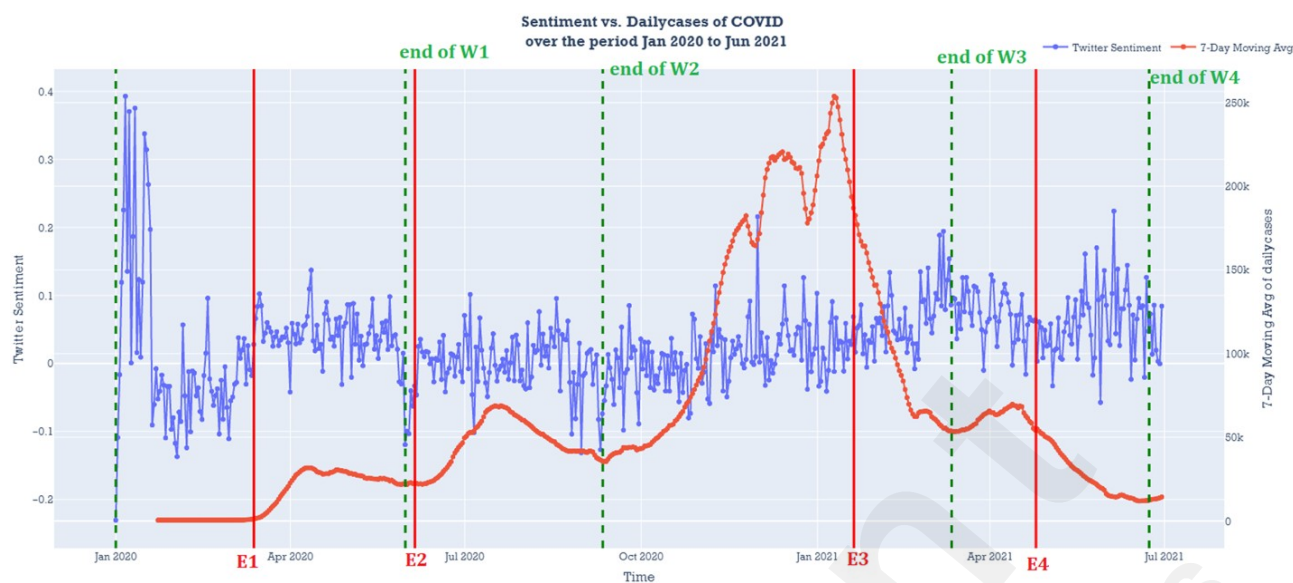
There were four key national events that had the most COVID-19 related tweets: the death of Breonna Taylor (Event 1), the nomination of President Joe Biden (Event 2), the presidential inauguration (Event 3), and the Academy Awards (Event 4). Therefore, they are marked as E1, E2, E3, and E4, accordingly by the vertical red lines in Figure 19.

The blue timeline is the sentiment. There was a sudden drop in sentiment at the beginning of COVID Wave 1. As more people across the United States were diagnosed with COVID and more stay-at-home orders were issued, Twitter users were showing fear and anxiety. During Wave 1, Event 1 (Breonna Taylor's death) increased the amount of COVID-related tweets but did not significantly influence the sentiment of the discussion. The end of Wave 1 almost coincided with Event 2 (Joe Biden's Democratic nomination for President), and there was a dip of sentiment right around the end of Wave 2 and then a quick bounce back around Event 2. During this time, many people were forced to stay at home, and the future was unclear. President Trump announced Operation Warp Speed in May, but a vaccine was not yet a reality when Joe Biden was nominated for President in August. After this, the sentiment scores bounced around 0 during Waves 3 to 4.

Generally, the sentiment has trended up since the end of Wave 2 (Sept. 11, 2020), with a few high positive days. However, the days around the end of the waves were comparatively more negative than the other days. The key events impacted the amounts of tweets but not the sentiment of the tweets. Figure 20 uses Google Trends to show the national search volumes of these four key events. The four key events coincided with the peeks of the national news agenda based upon Google trends.

Figure 20. Google trends data.
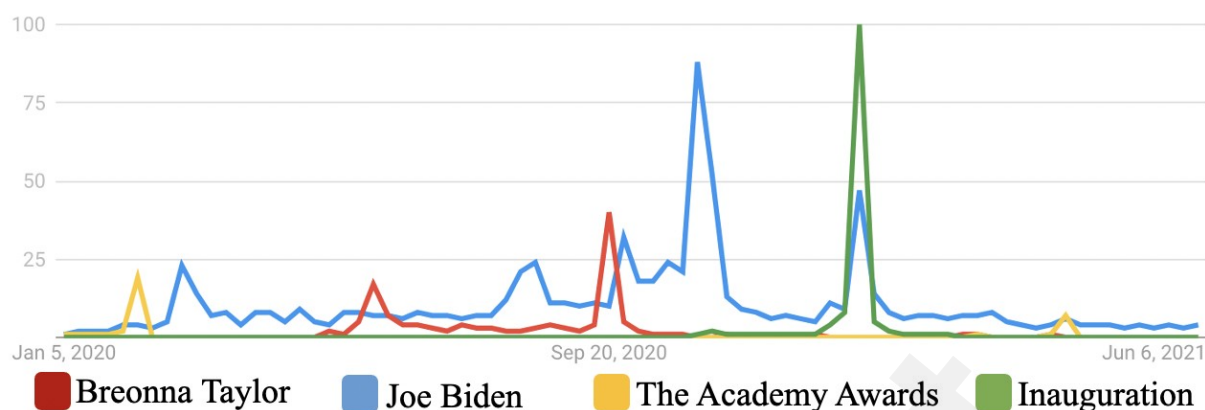
Discussion

Health communication is challenging because health is such a personal topic, often influenced by family and culture. As healthcare continues to move to patient-centered practices [37, 38], and patients increasingly utilize the Internet [39, 40], it stands to reason that public health communicators should engage in social listening and glean insights from target populations to better tailor health messaging. Engaging in social listening allows practitioners to track public opinion and predict trends in consumers' attitudes and behaviors. It is important to invest in such information so that campaign messaging can be designed to be the most effective.

This research project sought to examine Twitter posts from zip codes in mid-size cities across the United States with high percentages of African Americans to gather insight into the COVID-19 vaccine and related hesitancy so we can improve public health campaigns targeted at people of color. Triangulated methods were used to look at both the breadth and depth of data to provide validity to the findings. The strength of this study is the triangulated methods that combined the human expert and machine analyses and the Tweeter data from four cities with a larger population of people of color.

Findings from this study suggest public communication campaigns should seek to empower consumers by implementing three strategies. First, the objective of communicating science is to give people a sense of control and empower them to make good health decisions. The Twitter data from Kansas City revealed that this geographical area specifically does not understand the science related to COVID-19, including the severity of the virus, how the virus is contracted and/or spread. Public health messaging should focus on how COVID-19 differs from the flu and the potential of long-term health effects. In Raleigh, North Carolina, it should be said that many who are refusing the COVID-19 are not "anti-vaxxers" but instead are "refusers." This is an important difference to note as this population is targeted for persuasive messaging. Refusing the COVID-19 does not then make one against vaccines, some users argued. Instead, users weighed the risks of contracting the virus versus the unknown long-term side effects of the vaccine, which often determined their choice to be vaccinated. Thus, the campaigns need to communicate the science around vaccines and educate the public to become informed decision-makers.

Second, campaigns should seek to build trust locally and among local leaders on health issues that affect community residents and have a strong local theme and local relevance. One of the main topics discovered was the importance of community and user's investment in local issues. Specifically, the Twitter data from Omaha revealed that this geographical area values its independence and does not like government interference. To minimize local resentment to national pushes, appealing to "hometown values" might also be an effective persuasion strategy. For example,

the local Twitter data showed that users were very community-oriented and especially concerned about the future of the local community. Thus, appealing to rebuilding the local communities and putting in the sense of pride in the messaging may help strike a more positive tone and counter the negativity growing from economic and mental health concerns. Elected leaders and community members should also begin planning how to support initiatives that address these same concerns. When communicating with people of color, culture and tradition must be considered and integrated into the health messaging [41]. Messaging that is not authentic will only erode efforts toward building trust with people of color.

Third, timing is important to communication campaigns and vaccination operations. The sentiment analysis suggested that vaccination efforts in these cities missed the opportunity of capitalizing on people's excitement about the vaccines being available to them in early 2021. The earlier days of vaccination rollout often required people to navigate county websites and fill out lengthy forms online to make an appointment. This could be challenging for some and might discourage and limit communities of color to access vaccines due to a myriad of reasons: lack of Internet connection or electronic devices, not being able to find the link to sign up, navigate the county website, understanding the language about eligibility, understand questions about health insurance, provide documents of health insurance, fill out online forms, not having an email address, home address, or other required information to make an appointment, let alone difficulties of finding transportation and accommodating work schedules. Perhaps justifiably, users who faced challenges to getting the vaccine were angry. Arguably, this change of sentiment availed opportunities to vaccine hesitancy and disinformation, which disproportionally affect people of color. The data in June 2021 clearly shows a clip of people's interest in discussing COVID and the vaccine and persistent anger.

Limitation

Findings from this study may be limited by the sampling bias of the data. Twitter users are not necessarily representative of the general public or all local communities in a city. According to a Pew Research Center's report, U.S. adult Twitter users are younger, more educated, wealthier, and more likely to be Democrats than the general population [25]. On average, 10% of U.S. users generate 80% of the tweets [25]. Thus, any generalization of the results should be cautioned. However, the findings provide in-depth insight into word-of-mouth surrounding COVID-19 and the vaccine.

Final Thoughts

Public health communication competes for users' attention [42] and sometimes can be in conflict with news media reports [43]. Health literacy is associated with patients' engagement in disease management and treatment decision-making [44]. According to reports, approximately 80 million adults in the United States have limited or low health literacy skills [45]. Little is known about the state of health literacy in minority populations, including African Americans and Hispanics. Each local community has its own concerns and unique culture that must be included in public health communication so as to be perceived as credible and trustworthy. So often, organizations are ill-prepared to clearly and effectively communicate with target audiences. Science and medicine are complex topics that laypeople do not understand. It is not enough to communicate with audiences when there is an emergency like COVID-19 or the introduction of a vaccine. Instead,

public health entities must communicate regularly with target audiences to improve health literacy and work toward community goals. City governments must continue to build trust throughout the year by being transparent but also recognizing where their community is in terms of beliefs and education. Consumer insights are used in a variety of industries, including marketing and advertising [46, 47], real estate [48], travel [49, 51], and many others. As misinformation continues to threaten public health goals, it only makes sense that public health practitioners and city governments also use consumer insights to inform campaigns and reach specialized audiences most effectively.

References

1. U.S. Census. 2019. Jackson County, Missouri; Kansas City, Missouri.https://www.census.gov/quickfacts/fact/table/jacksoncountymissouri,kansascitycitymissouri/PST045219.

2. Belshe S. 2021. KC health officials work to increase COVID vaccination rates as Delta variant spreads, *Missouri Independent*. https://missouriindependent.com/2021/08/02/kc-health-officials-work-to-increase-covid-vaccination-rates-as-delta-variant-spreads/.

3. Fortino, J. 2021, May 10. Jackson County Oks $5M to ramp up COVID vaccinations on Kansas City's East side. KCUR.org. https://www.kcur.org/health/2021-05-10/jackson-county-oks-5-million-project-to-ramp-up-covid-19-vaccinations-on-kansas-citys-east-side.

4. Griffith J, Marani H, Monkman H. COVID-19 vaccine hesitancy in Canada: content analysis of tweets using the theoretical domains framework. *J Med Internet Res*. 2021;23(4):e26874. Published 2021 Apr 13. doi:10.2196/26874

5. Barello S, Palamenghi L, Graffigna G. Looking inside the 'black box' of vaccine hesitancy: unlocking the effect of psychological attitudes and beliefs on COVID-19 vaccine acceptance and implications for public health communication [published online ahead of print, 2021 Mar 8]. *Psychol Med*. 2021;1-2. doi:10.1017/S003329172100101X

6. Opel DJ, Marcuse EK. Rethinking vaccine policy making in an era of vaccine hesitancy: time to rebuild, not remodel?. *Hum Vaccin Immunother*. 2013;9(12):2672-2673. doi:10.4161/hv.26651

7. World Health Organization. 2019. Overview of the global influenza strategy 2019-2030. https://www.jstor.org/stable/resrep27891.5.

8. Center for Systems Science and Engineering (CSSE). 2021. COVID-19 Dashboard by CSSE at Johns Hopkins University (JHU). https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6.

9. Shim JG, Ryu KH, Lee SH, Cho EA, Lee YJ, Ahn JH. Text mining approaches to analyze public sentiment changes regarding covid-19 vaccines on social media in korea. *Int J Environ Res Public Health*. 2021;18(12):6549. Published 2021 Jun 18. doi:10.3390/ijerph18126549

10. Kwok SWH, Vadde SK, Wang G. Tweet topics and sentiments relating to covid-19 vaccination among australian twitter users: machine learning analysis. *J Med Internet Res*. 2021;23(5):e26953. Published 2021 May 19. doi:10.2196/26953

11. Cotfas LA, Delcea C, Roxin I, Ioanăş C, Gherai DS, Tajariol F. The longest month: analyzing

covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. IEEE Access. 2021 Feb 16;9: doi: 10.1109/ACCESS.2021.3059821.

12. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-related discussion on twitter: topic modeling and sentiment analysis. *J Med Internet Res*. 2021;23(6):e24435. Published 2021 Jun 29. doi:10.2196/24435

13. Ihlen Ø, Toledano M, Just SN. Using Rhetorical Situations to Examine and Improve Vaccination Communication. Frontiers in Communication. 2021 Jun 30;6:130. https://doi.org/10.3389/fcomm.2021.697383

14. Salmon D, Opel DJ, Dudley MZ, Brewer J, Breiman R. Reflections on governance, communication, and equity: challenges and opportunities in covid-19 vaccination. *Health Aff (Millwood)*. 2021;40(3):419-425. doi:10.1377/hlthaff.2020.02254

15. Collins HM, Evans R. The third wave of science studies: studies of expertise and experience. Social Studies of Science. 2002;32(2):235-296. doi:10.1177/0306312702032002003

16. Kunyk D, Olson JK. Clarification of conceptualizations of empathy. Journal of Advanced nursing. 2001 Aug 23;35(3):317-25. https://doi.org/10.1046/j.1365-2648.2001.01848.x

17. Teding van Berkhout E, Malouff JM. The efficacy of empathy training: A meta-analysis of randomized controlled trials. Journal of counseling psychology. 2016 Jan;63(1):32. https://doi.org/10.1037/cou0000093

18. Kotliar DM. Depression Narratives in Blogs: A Collaborative Quest for Coherence. Qualitative Health Research. 2016;26(9):1203-1215. doi:10.1177/1049732315612715

19. Wang Y, Willis E. Examining theory-based behavior-change constructs, social interaction, and sociability features of the weight watchers' online community. *Health Educ Behav*. 2016;43(6):656-664. doi:10.1177/1090198116629415

20. Teoh D. The power of social media for hpv vaccination-not fake news!. *Am Soc Clin Oncol Educ Book*. 2019;39:75-78. doi:10.1200/EDBK_239363

21. Russell Neuman W, Guggenheim L, Mo Jang S, Bae SY. The dynamics of public attention: Agenda-setting theory meets big data. Journal of Communication. 2014 Apr 1;64(2):193-214. https://doi.org/10.1111/jcom.12088

22. National Geographic. 2021. Coronavirus in the U.S.: Where cases are growing and declining. https://www.nationalgeographic.com/science/graphics/graphic-tracking-coronavirus-infections-us.

23. Statista. 2021. Most popular social networks worldwide as of July 2021, ranked by number of active users. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

24. Omnicore. 2021. Twitter by the numbers: Stats, demographics & fund facts. https://www.omnicoreagency.com/twitter-statistics/.

25. Wojcik S, & Hughes A. 2019. Sizing up Twitter users.

https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research. 2003 Mar 1;3:993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB_iframe=true&width=370.8&height=658.8

27. Shao, S. 2020. Contextual topic identification for steam reviews. https://github.com/Stveshawn/contextual_topic_identification.

28. Hugging Face. 2021. The AI community building the future.
https://huggingface.co/.

29. Reimers N, Gurevych I. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084. 2019 Aug 27.

30. Berger AA. Media research techniques (2nd ed.). Sage; 1998 May 5.

31. McKee A. *Textual analysis: A beginner's guide*. Sage; 2003 Mar 4.

32. Miles MB, Huberman AM. Qualitative data analysis: An expanded sourcebook. Sage; 1994 Jan 12.

33. Hu X, Tang J, Gao H, Liu H. Unsupervised sentiment analysis with emotional signals. In Proceedings of the 22nd international conference on World Wide Web. 2013 May 13 (pp. 607-618). https://doi.org/10.1145/2488388.2488442

34. Liu B. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge university press; 2020 Sep 30.

35. AARP. 2021. Covid-19's fourth wave. https://www.aarp.org/health/conditions-treatments/info-2021/covid-4th-wave.html.

36. Centers for Disease Control and Prevention. 2021. Trends in number of Covid-19 cases in the US reported to the CDC. https://stacks.cdc.gov/view/cdc/101044.

37. Bertakis KD, Azari R. Patient-centered care is associated with decreased health care utilization. *J Am Board Fam Med*. 2011;24(3):229-239. doi:10.3122/jabfm.2011.03.100170

38. Epstein RM. The science of patient-centered care. *J Fam Pract*. 2000;49(9):805-807.

39. Benetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. *Patient Educ Couns*. 2018;101(3):439-444. doi:10.1016/j.pec.2017.08.015

40. Chen YY, Li CM, Liang JC, Tsai CC. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. *J Med Internet Res*. 2018;20(2):e47. Published 2018 Feb 12. doi:10.2196/jmir.9370

41. Henry Akintobi T, Jacobs T, Sabbs D, et al. Community engagement of African Americans in the

era of covid-19: considerations, challenges, implications, and recommendations for public health. *Prev Chronic Dis*. 2020;17:E83. Published 2020 Aug 13. doi:10.5888/pcd17.200255

42. Vraga EK, Radzikowski JR, Stefanidis A, Croitoru A, Crooks AT, Delamater PL, Pfoser D, Jacobsen KH. Social media engagement with cancer awareness campaigns declined during the 2016 US presidential election. World Medical & Health Policy. 2017 Dec;9(4):456-65. https://doi.org/10.1002/wmh3.247

43. Gollust SE, Fowler EF, Niederdeppe J. Television news coverage of public health issues and implications for public health policy and practice. *Annu Rev Public Health*. 2019;40:167-185. doi:10.1146/annurev-publhealth-040218-044017

44. Brabers AE, Rademakers JJ, Groenewegen PP, van Dijk L, de Jong JD. What role does health literacy play in patients' involvement in medical decision-making?. *PLoS One*. 2017;12(3):e0173316. Published 2017 Mar 3. doi:10.1371/journal.pone.0173316

45. Hickey KT, Masterson Creber RM, Reading M, et al. Low health literacy: Implications for managing cardiac patients in practice. *Nurse Pract*. 2018;43(8):49-55. doi:10.1097/01.NPR.0000541468.54290.49

46. Fortini-Campbell L. The consumer insight workbook: How consumer insights can inspire better marketing and advertising. Journal of Consumer Marketing. 1992 Apr 1. https://doi.org/10.1108/eb008240

47. Southgate D. The emergence of Generation Z and its impact in advertising: Long-term implications for media planning and creative development. Journal of Advertising Research. 2017 Jun 1;57(2):227-35.

48. Grybauskas A, Pilinkienė V, Stundžienė A. Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *J Big Data*. 2021;8(1):105. doi:10.1186/s40537-021-00476-0

49. Bowen J, Whalen E. Trends that are changing travel and tourism. Worldwide Hospitality and Tourism Themes. 2017 Dec 4. https://doi.org/10.1108/WHATT-09-2017-0045
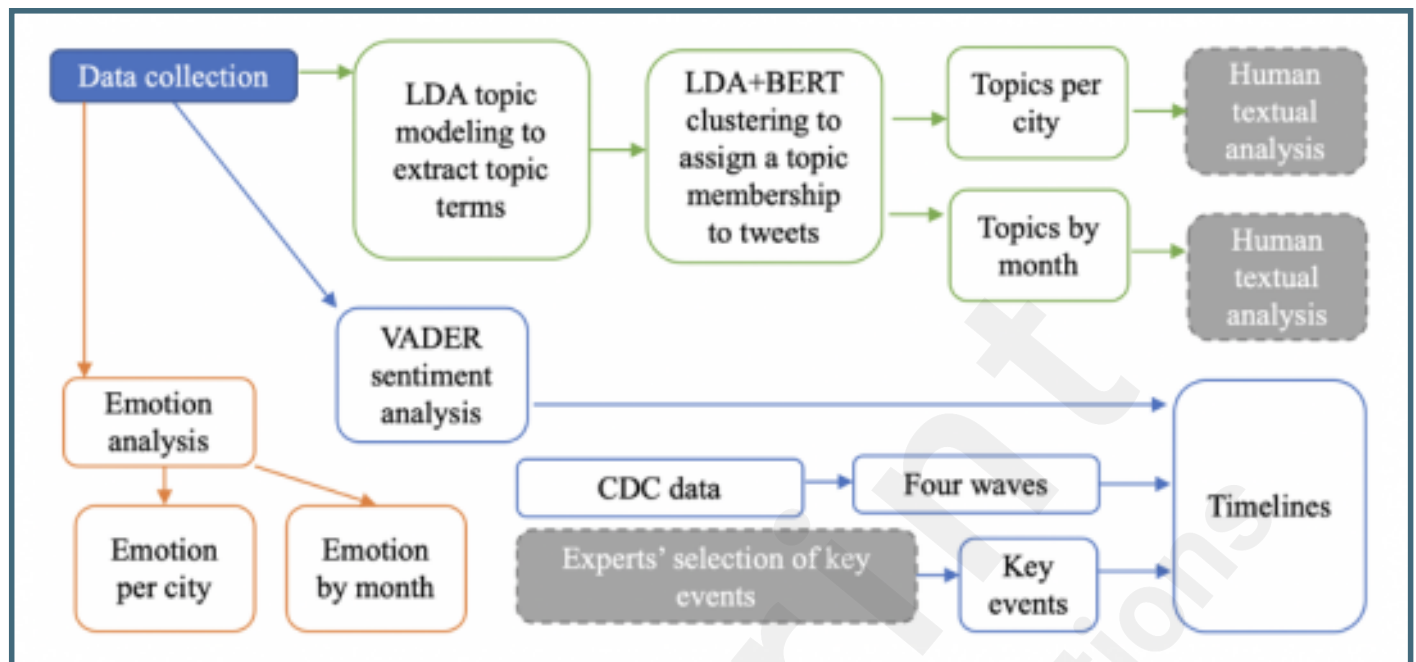
50. Colladon AF, Guardabascio B, Innarella R. Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. Decision Support Systems. 2019 Aug 1;123:113075. https://doi.org/10.1016/j.dss.2019.113075
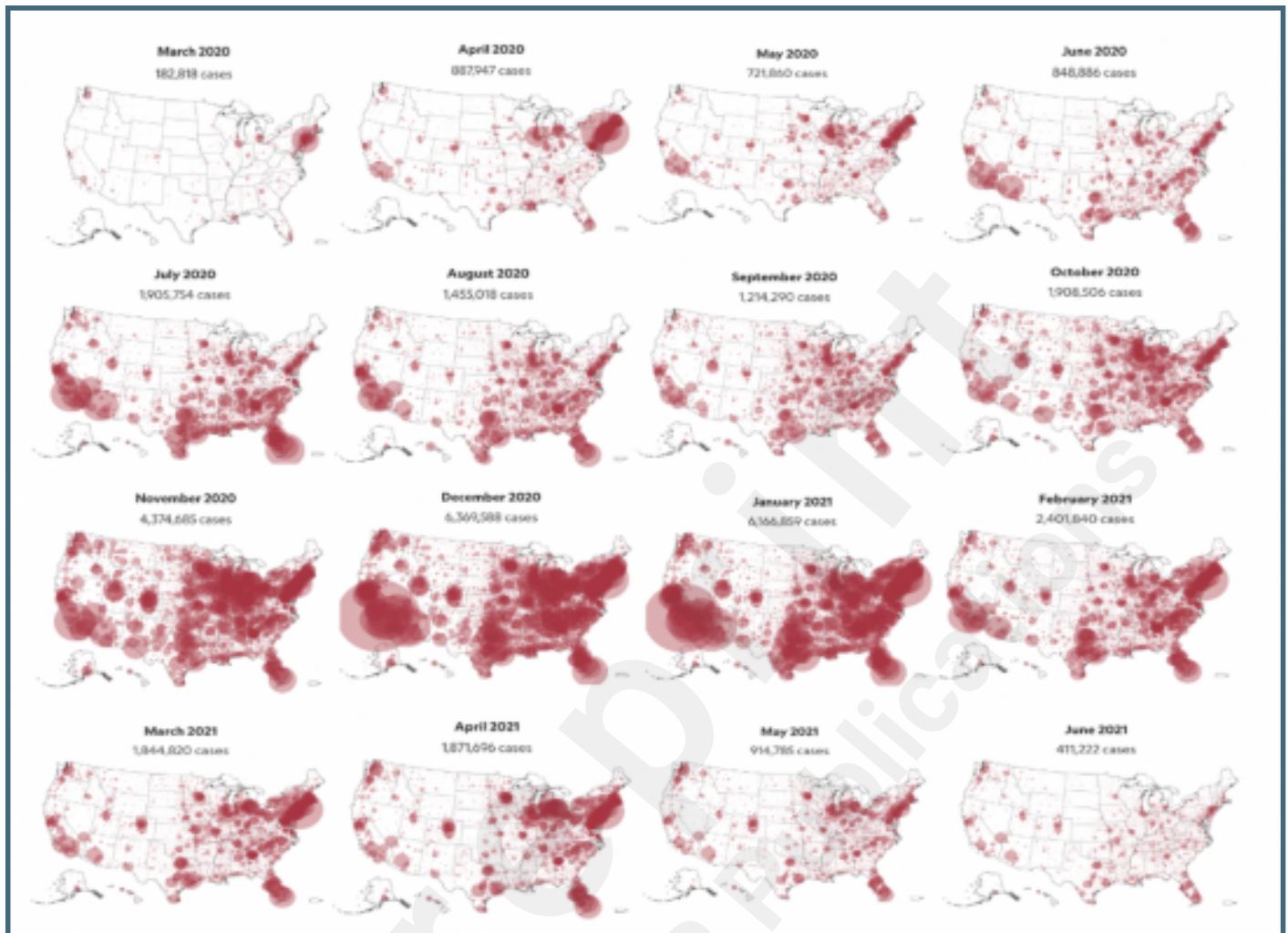
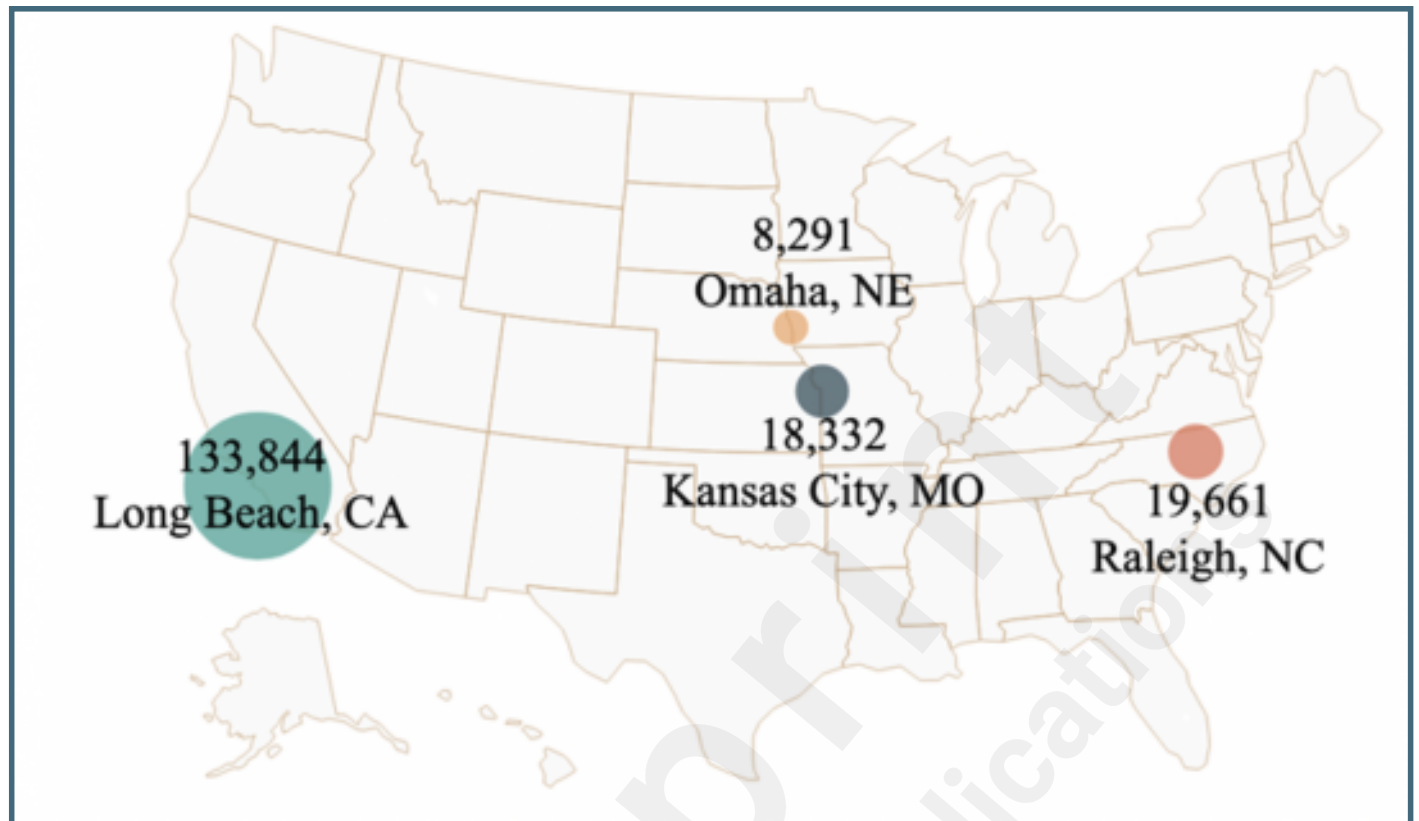# Supplementary Files

Figures
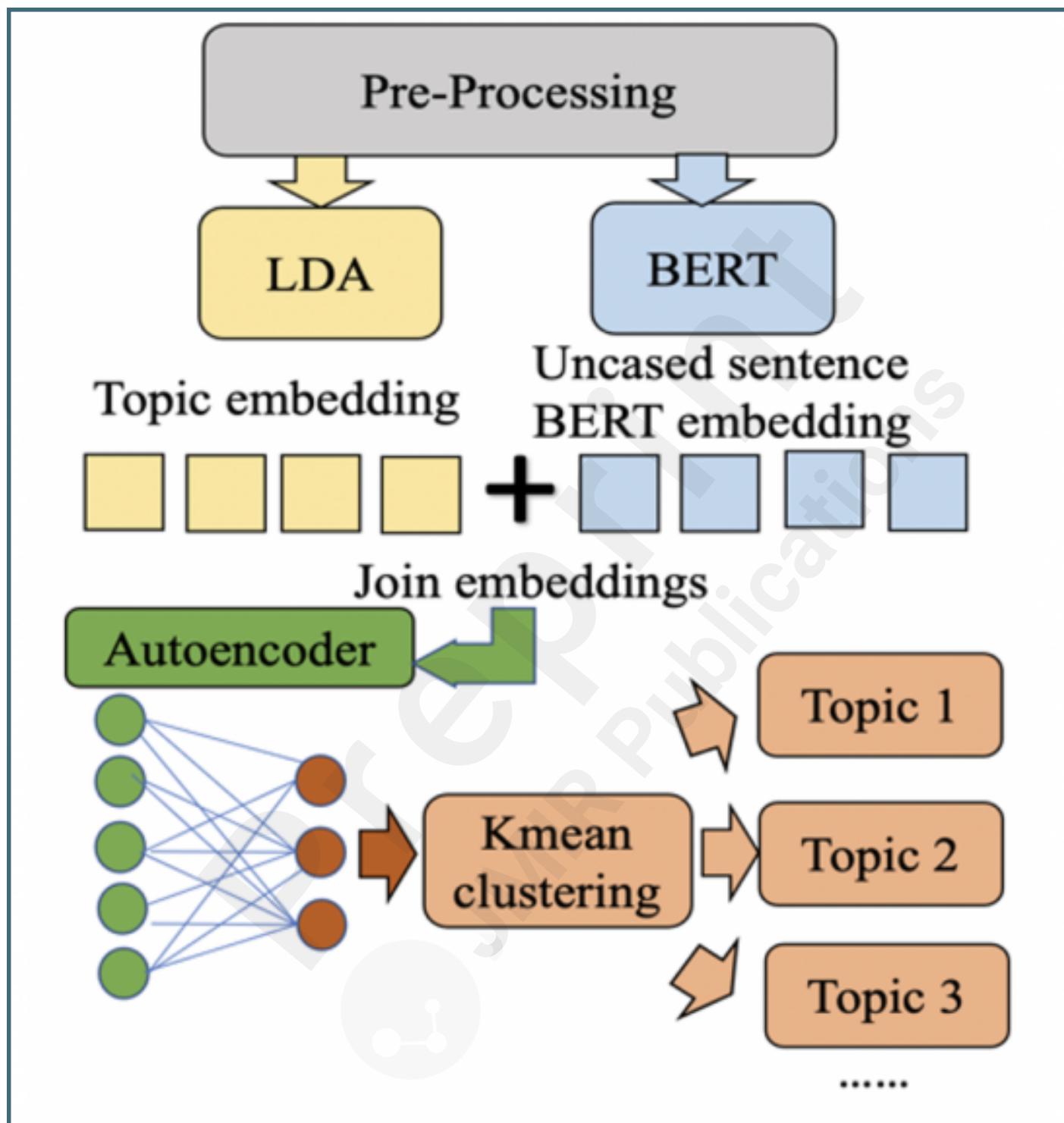
The workflow of the data analysis.

The confirmed cases of coronavirus in the U.S. by month, March 2020 to June 2021.
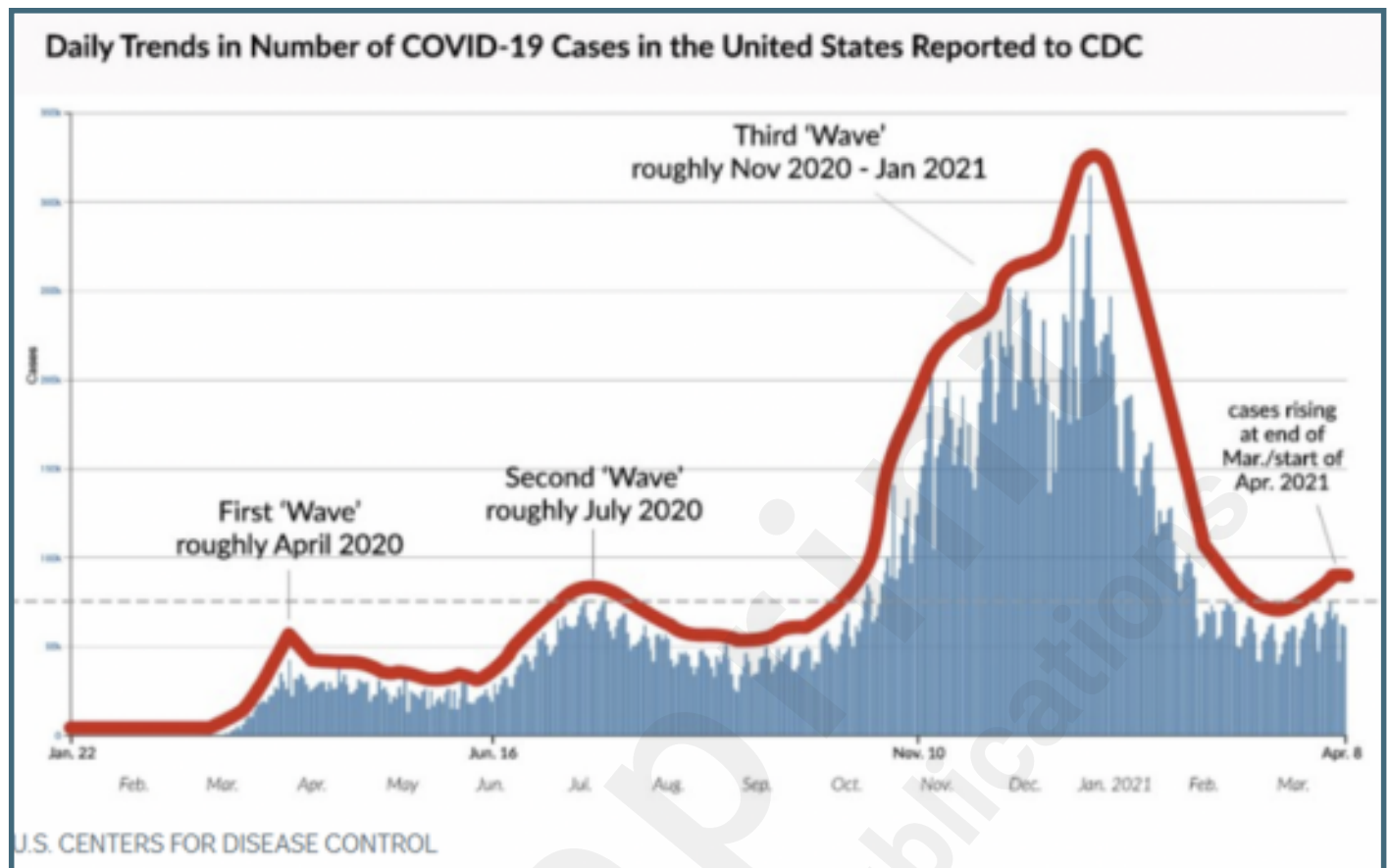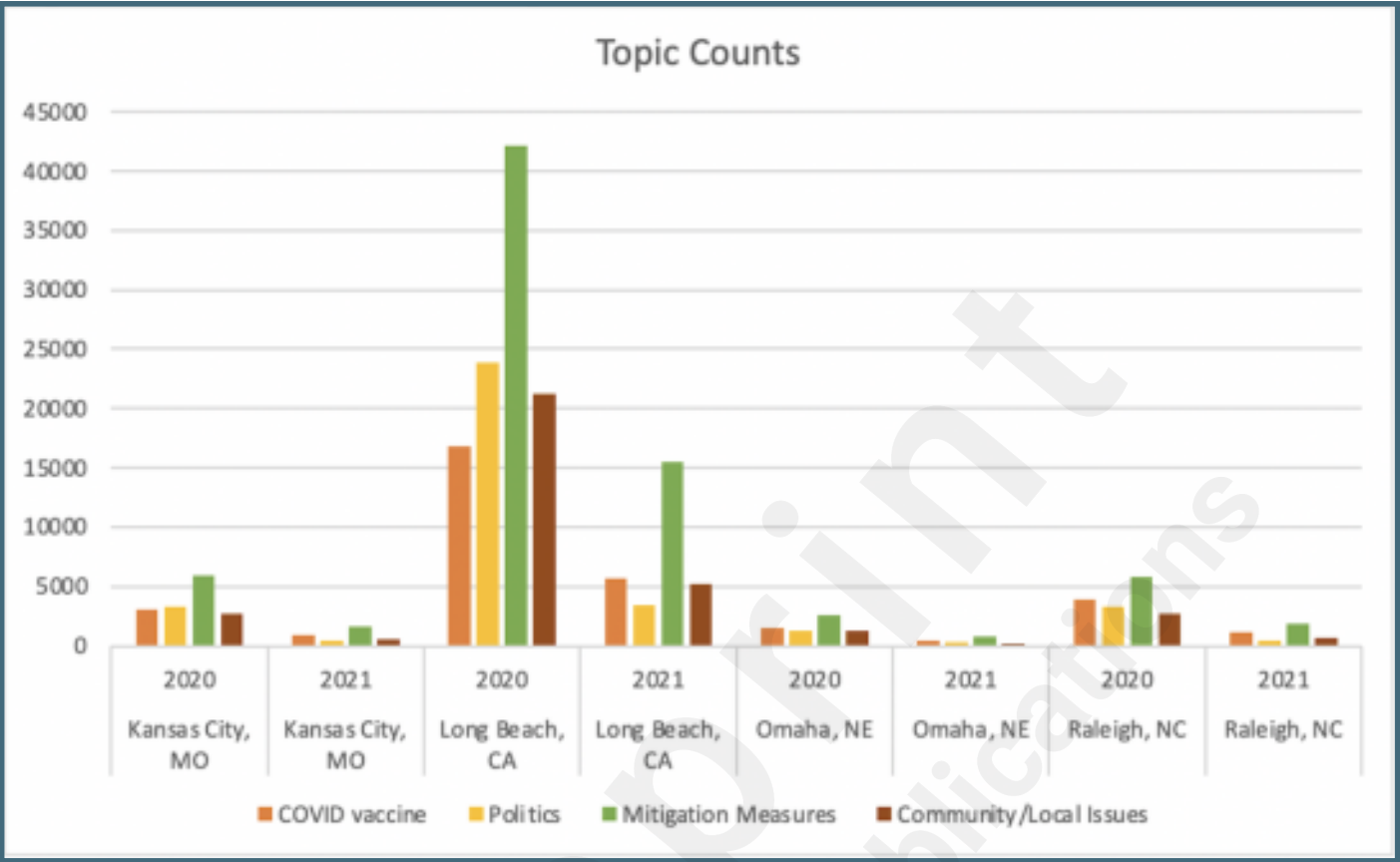
Geographic distribution of the sampled tweets.
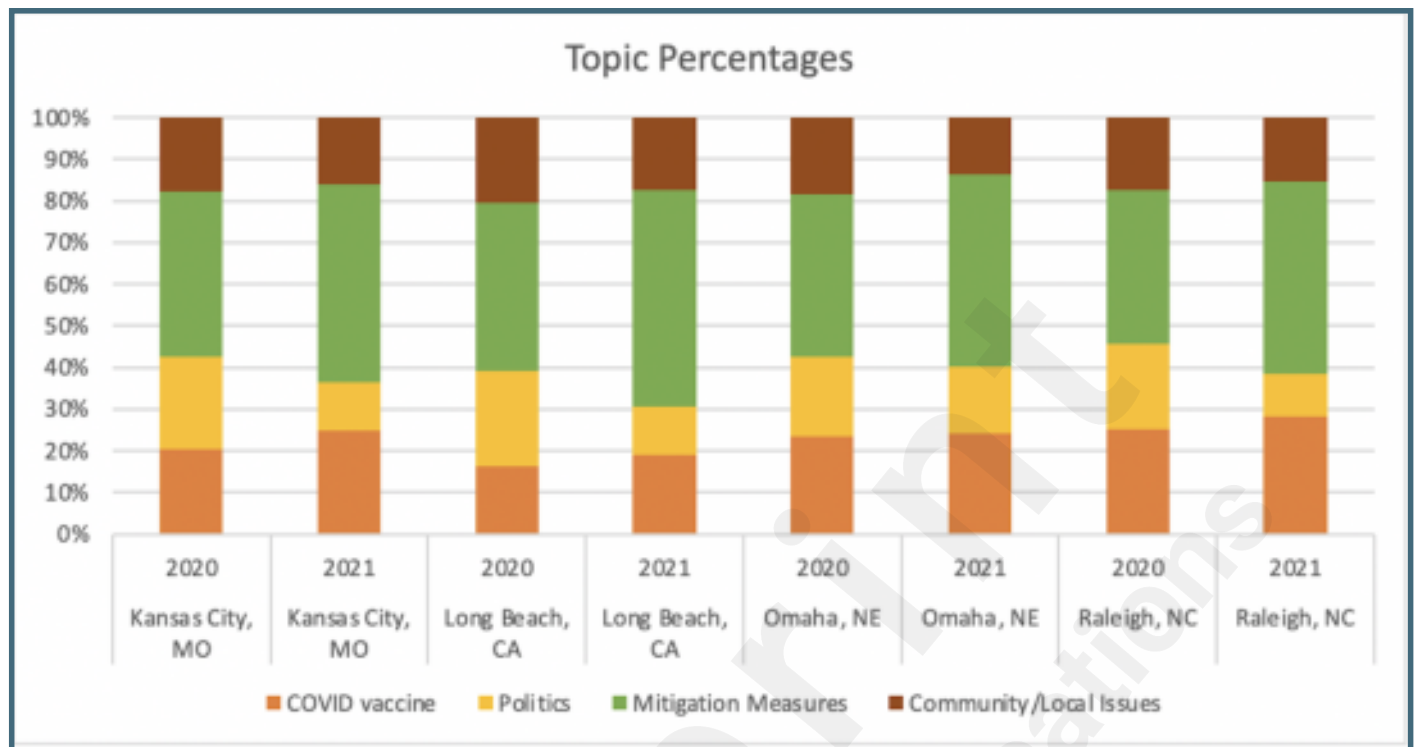
The LDA+BERT clustering model.

Four waves identified from the CDC's daily trends' data.

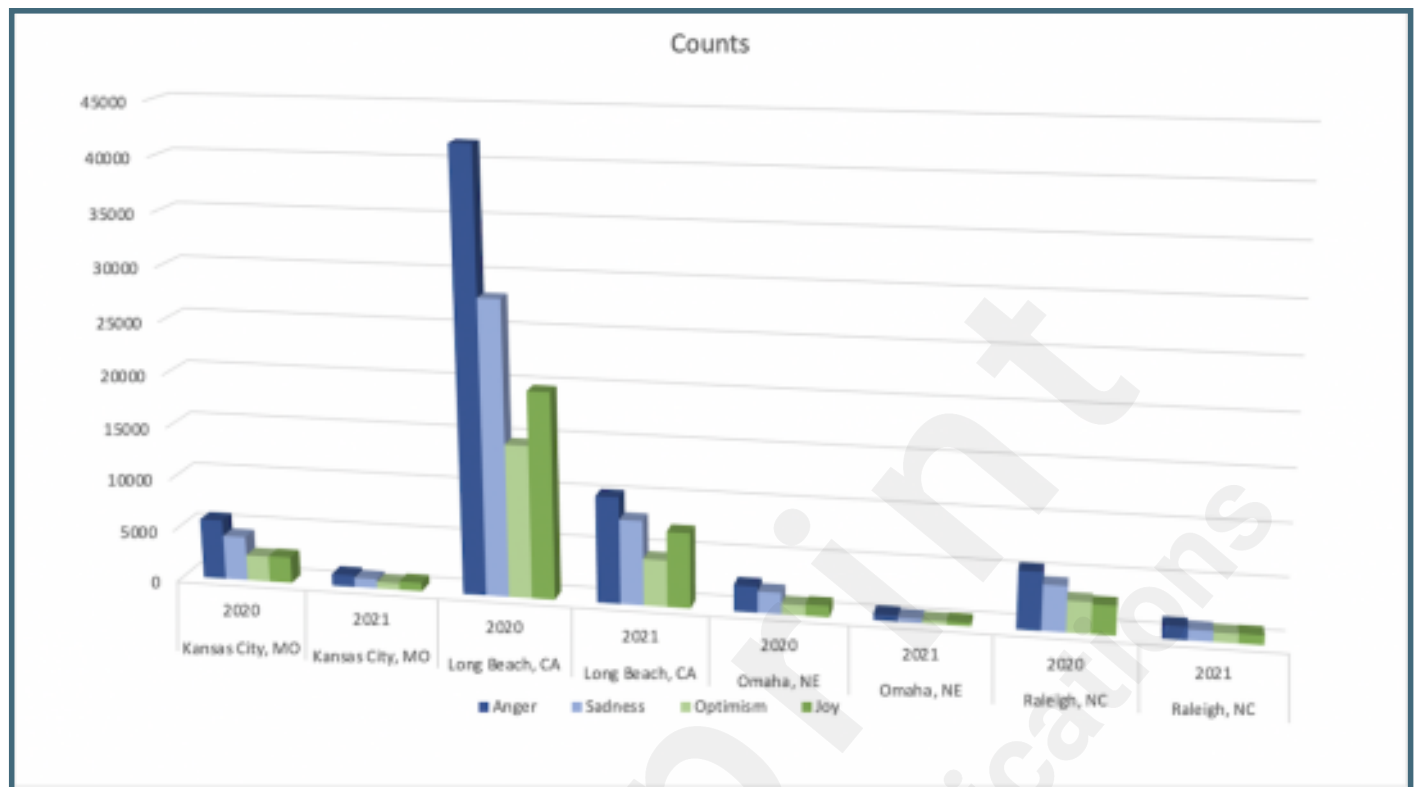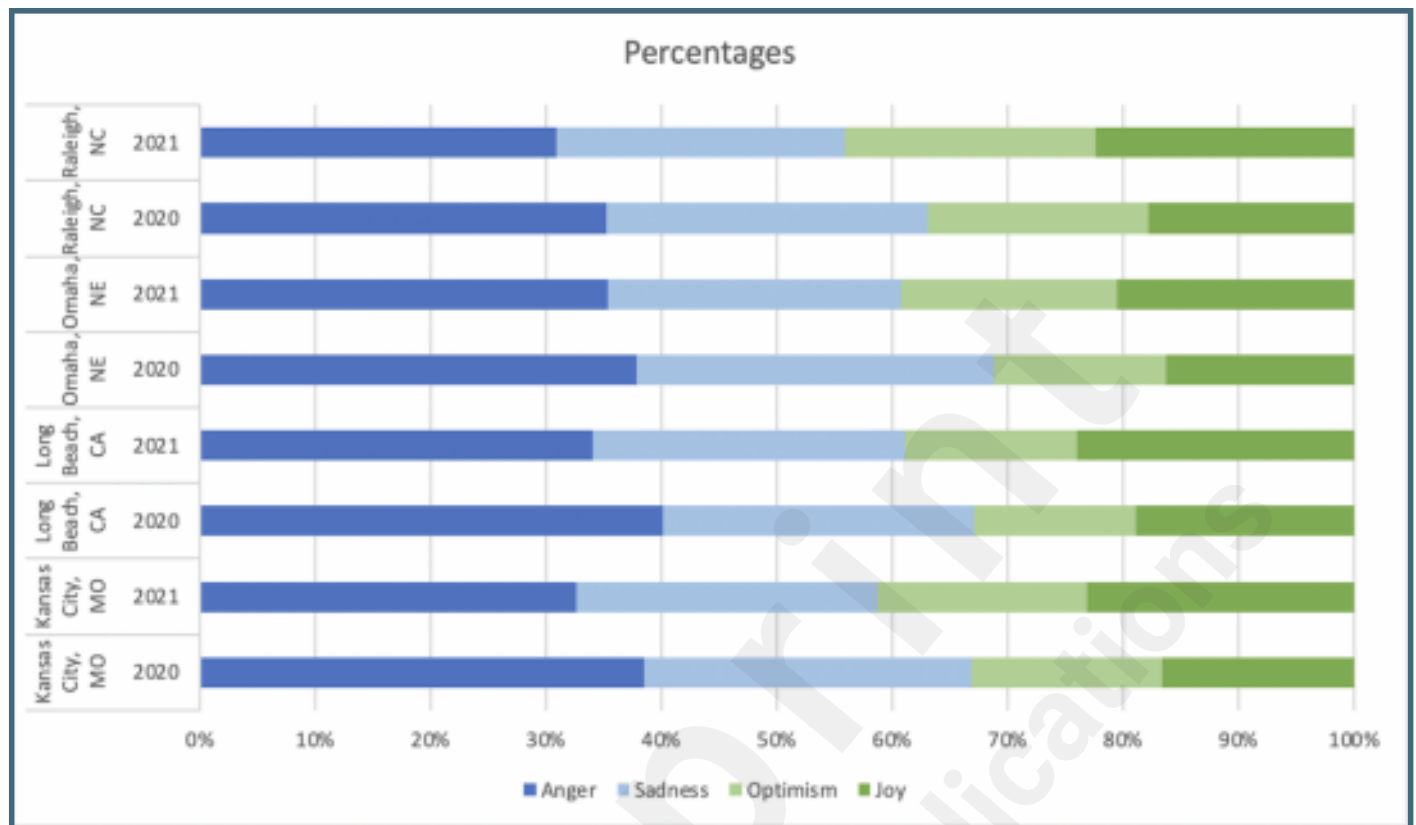Tweet counts of topics.

Tweet percentages of topics.

Topic Percentages

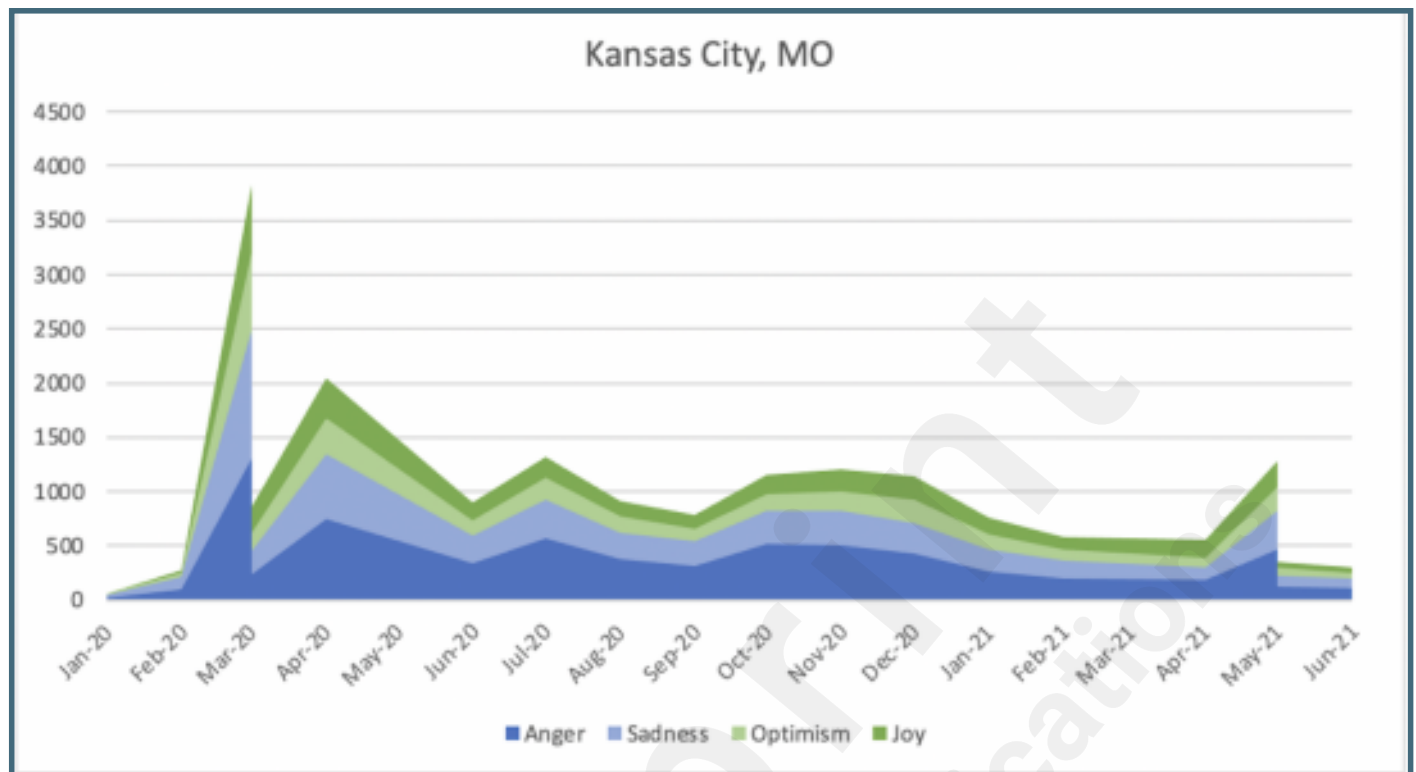Counts of tweets of different emotions.

Percentages of tweets of different emotions.

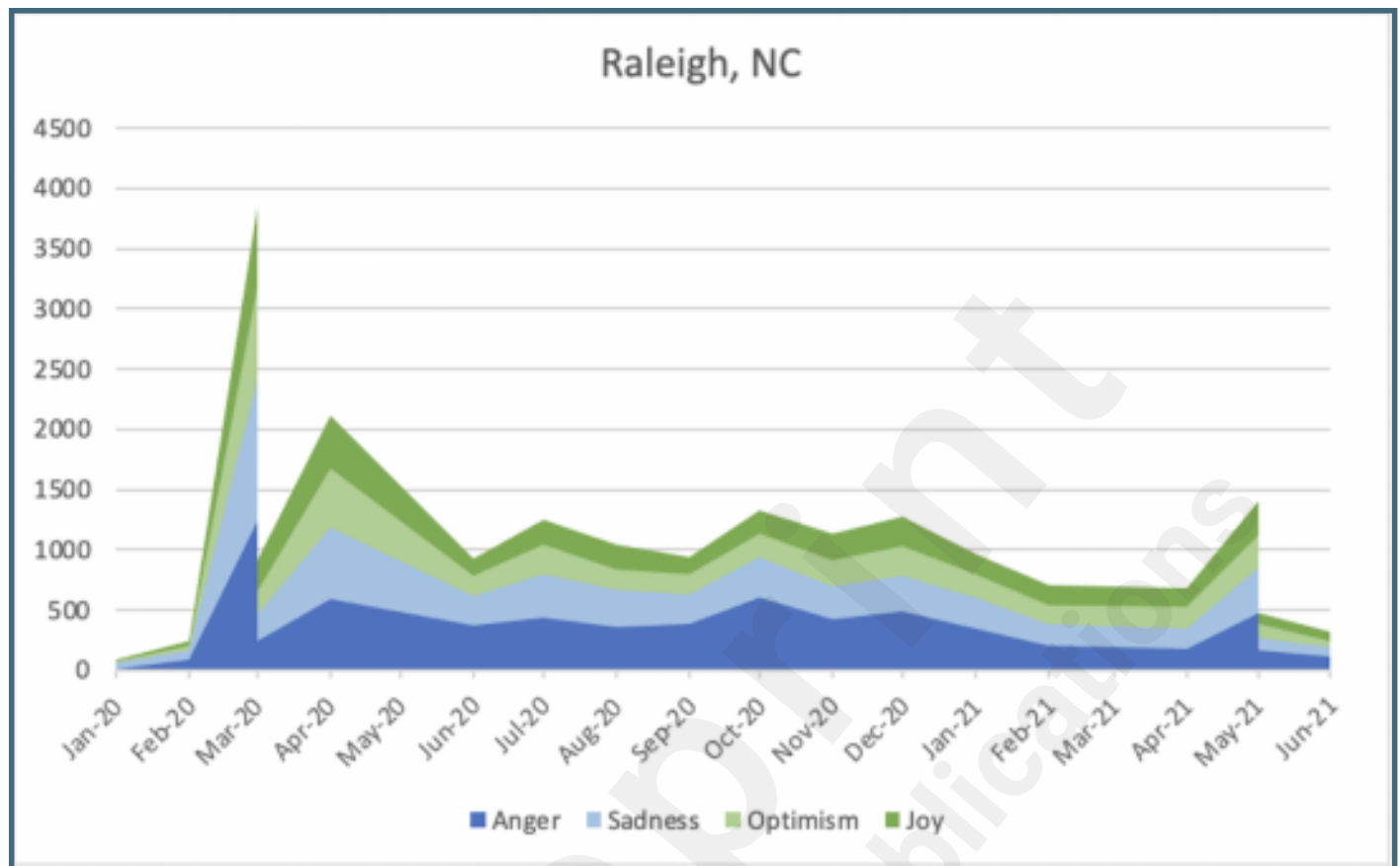Month by month emotions counts, Kansas City, MO, 2020-2021.

Month by month emotions counts, Long Beach, CA, 2020-2021.
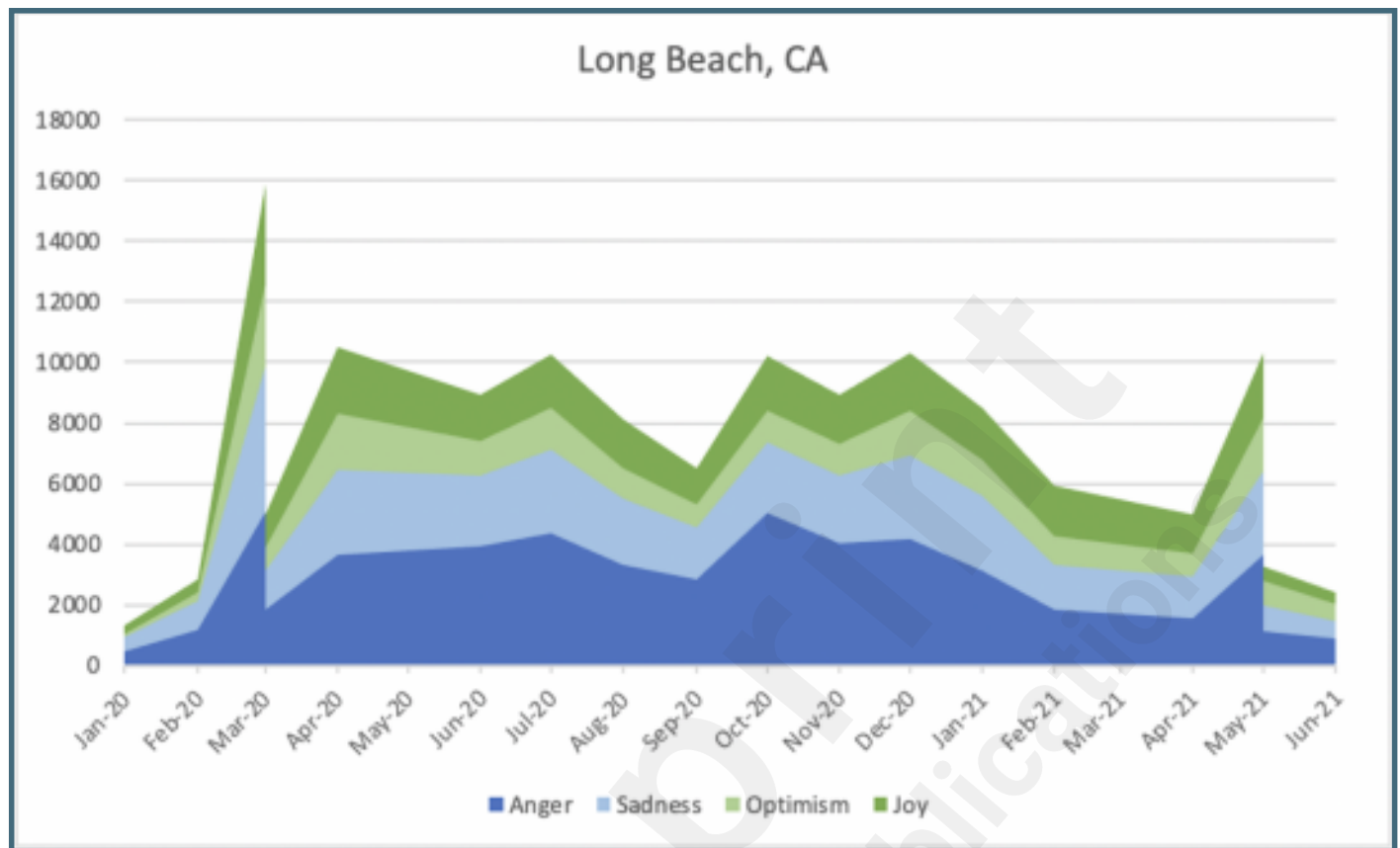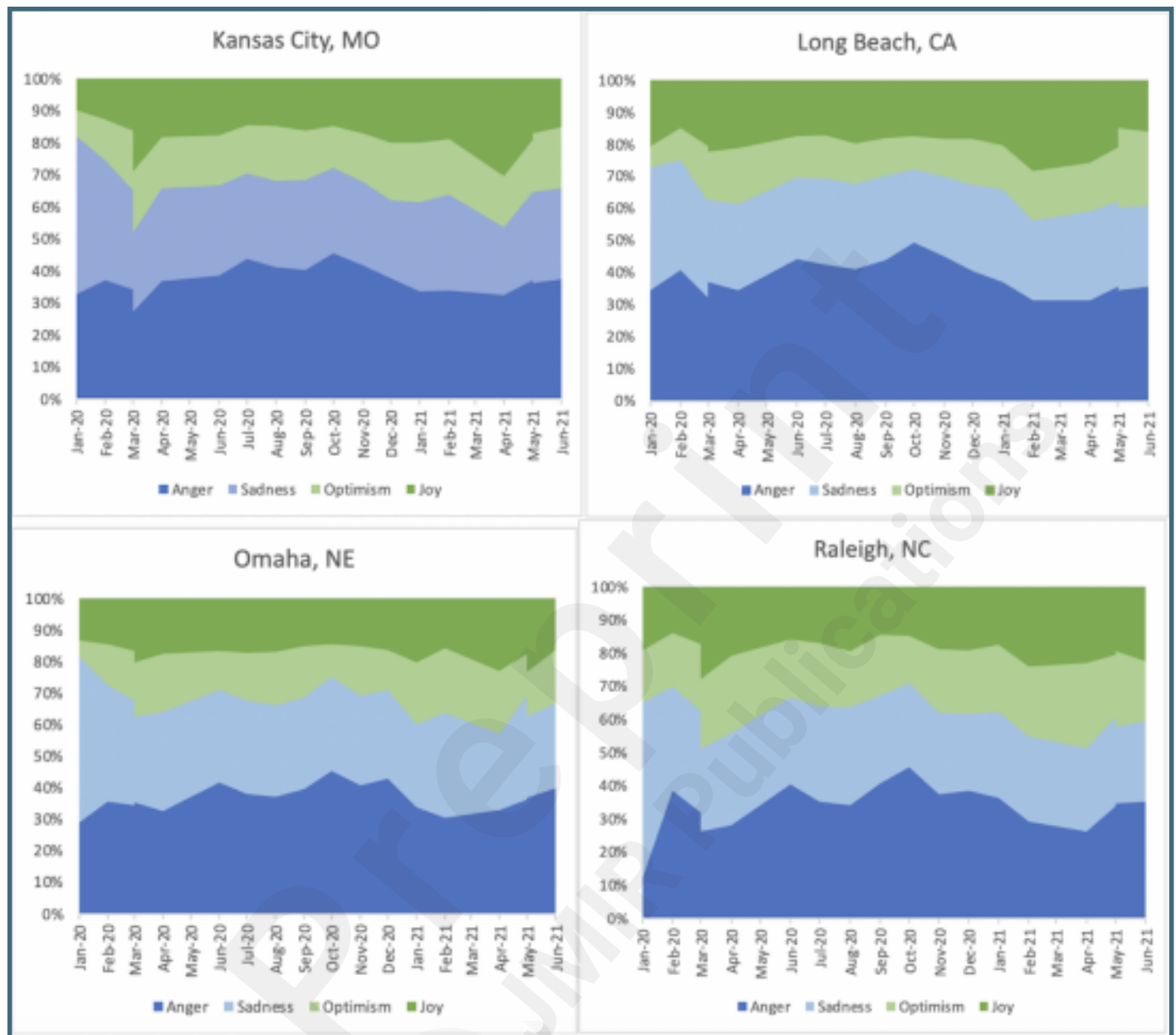
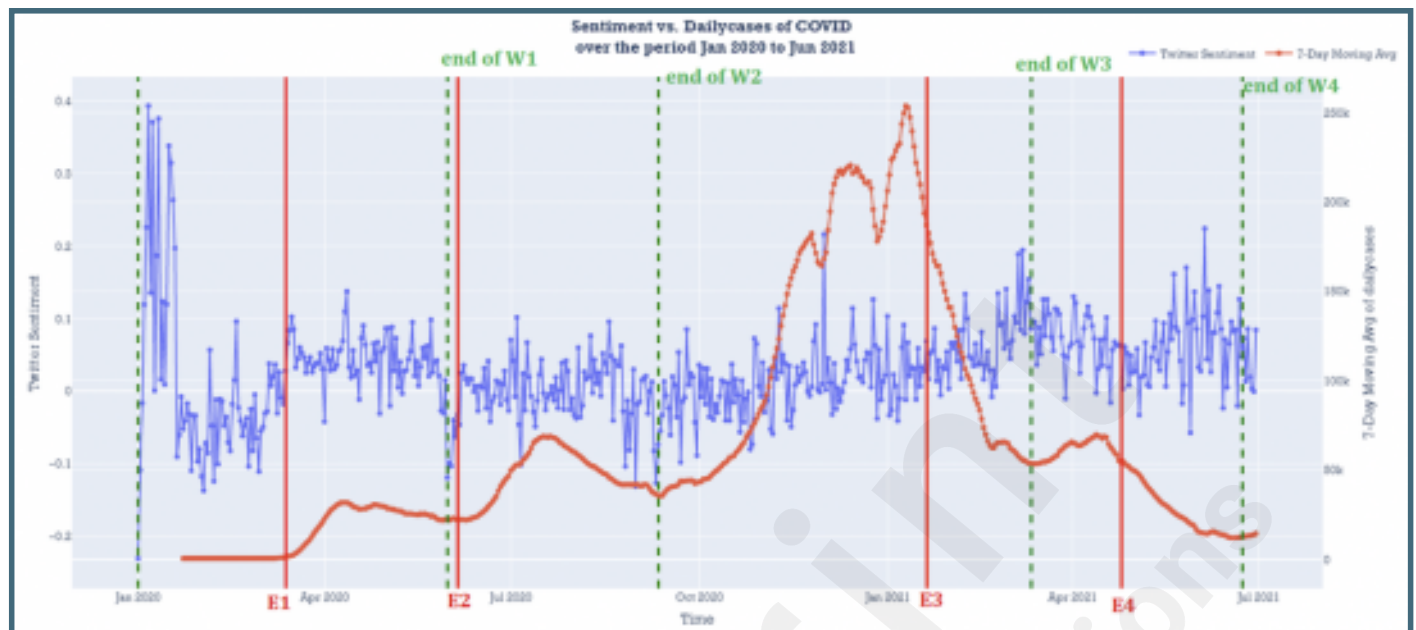Month by month emotions counts, Omaha, NE, 2020-2021.
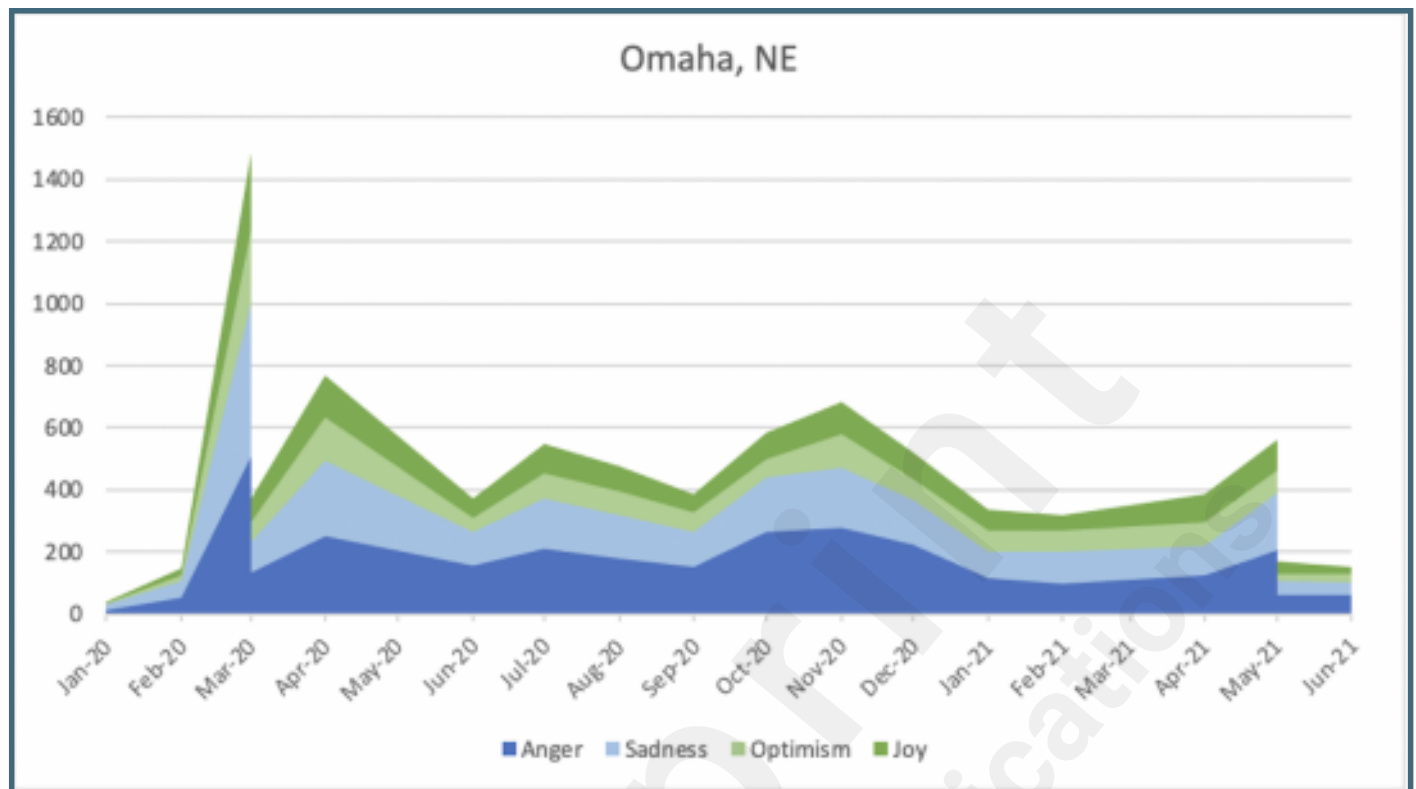
Month by month emotions counts, Raleigh, NC, 2020-2021.

Month by month emotions percentages, 2020-2021.

Sentiment and daily COVID case numbers displayed over the pandemic.

Google trends data.