Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: a spatiotemporal perspective

Tao Hu, Siqin Wang, Wei Luo, Mengxi Zhang, Xiao Huang, Yingwei Yan, Regina Liu, Kelly Ly, Viraj Kacker, Bing She, Zhenlong Li

Submitted to: Journal of Medical Internet Research
on: June 01, 2021

Table of Contents

Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: a spatiotemporal perspective

Tao Hu[1]; Siqin Wang[2]; Wei Luo[3]; Mengxi Zhang[4]; Xiao Huang[5]; Yingwei Yan[3]; Regina Liu[6]; Kelly Ly[7]; Viraj Kacker[8]; Bing She[9]; Zhenlong Li[10]

[1]Harvard University Cambridge US
[2]University of Queensland Brisbane AU
[3]National University of Singapore Singapore SG
[4]Ball State University Muncie US
[5]University of Arkansas Fayetteville US
[6]Mercer University Macon US
[7]University of Massachusetts Lowell Lowell US
[8]Georgia Institute of Technology GA US
[9]University of Michigan Ann Arbor US
[10]University of South Carolina SC US

Corresponding Author:
Wei Luo
National University of Singapore
21 Lower Kent Ridge Rd
Singapore
SG

Abstract

Background: The COVID-19 pandemic has imposed a large, initially uncontrollable, public health crisis both in the US and across the world, with experts looking to vaccines as the ultimate mechanism of defense. The development and deployment of COVID-19 vaccines have been rapidly advancing via global efforts. Hence, it is crucial for governments, public health officials, and policy makers to understand public attitudes and opinions towards vaccines, such that effective interventions and educational campaigns can be designed to promote vaccine acceptance

Objective: The aim of this study is to investigate public opinion and perception on COVID-19 vaccines by investigating the spatiotemporal trends of their sentiment and emotion towards vaccines, as well as how such trends relate to popular topics on Twitter in the US

Methods: We collected over 300,000 geotagged tweets in the US from March 1, 2020 to February 28, 2021. We examined the spatiotemporal patterns of public sentiment and emotion over time at both national and state scales and identified three phases along the pandemic timeline with the significant changes of public sentiment and emotion, further linking to eleven key events and major topics as the potential drivers to induce such changes via cloud mapping of keywords and topic modelling

Results: An increasing trend of positive sentiment in parallel with the decrease of negative sentiment are generally observed in most states, reflecting the rising confidence and anticipation of the public towards vaccines. The overall tendency of the eight types of emotion implies the trustiness and anticipation of the public to vaccination, accompanied by the mixture of fear, sadness and anger. Critical social/international events and/or the announcements of political leaders and authorities may have potential impacts on the public opinion on vaccines. These factors, along with important topics and manual reading of popular posts on eleven key events, help identify underlying themes and validate insights from the analysis

Conclusions: The analyses of near real-time social media big data benefit public health authorities by enabling them to monitor public attitudes and opinions towards vaccine-related information in a geo-aware manner, address the concerns of vaccine skeptics and promote the confidence of individuals within a certain region or community, towards vaccines

(JMIR Preprints 01/06/2021:30854)
DOI: https://doi.org/10.2196/preprints.30854

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

Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: a spatiotemporal perspective

Tao Hu[1,2†], Siqin Wang[3†], Wei Luo[4†*], Mengxi Zhang[5], Xiao Huang[6], Yingwei Yan[4], Regina Liu[7], Kelly Ly[8], Viraj Kacker[9], Bing She[10], Zhenlong Li[11]

[1]Department of Geography, Oklahoma State University, Stillwater, OK, US. taohu.gis@gmail.com
[2]Center for Geographic Analysis, Harvard University, Cambridge, MA, US. taohu.gis@gmail.com
[3]School of Earth and Environmental Sciences, University of Queensland, Brisbane, Queensland, Australia. s.wang6@uq.edu.au
[4]Department of Geography, National University of Singapore, Singapore, Singapore. geowl@nus.edu.sg; yingwei.yan@nus.edu.sg
[5]Department of Nutrition and Health Science, Ball State University, Indiana, US, mzhang2@bsu.edu
[6]Department of Geosciences, University of Arkansas, Fayetteville, Arkansas, US, xh010@uark.edu.
[7]Department of Biology, Mercer University, Macon, GA, US. Regina.liu@live.mercer.edu
[8]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, US. kelly_Ly@student.uml.edu
[9]College of Computing, Georgia Institute of Technology, GA, US. vkacker@gatech.edu
[10]Institute for Social Research, University of Michigan, MI, US. coolnanjizhou@gmail.com
[11]Geoinformation and Big Data Research Laboratory, Department of Geography, University of South Carolina, SC, USA. zhenlong@sc.edu

[†] These authors contributed equally to this work
*Corresponding author

Abstract

Background: The COVID-19 pandemic has imposed a large, initially uncontrollable, public health crisis both in the United States and across the world, with experts looking to vaccines as the ultimate mechanism of defense. The development and deployment of COVID-19 vaccines have been rapidly advancing via global efforts. Hence, it is crucial for governments, public health officials, and policy makers to understand public attitudes and opinions towards vaccines, such that effective interventions and educational campaigns can be designed to promote vaccine acceptance.
Objective: The aim of this study is to investigate public opinion and perception on COVID-19 vaccines in the US. We investigated the spatiotemporal trends of public sentiment and emotion towards COVID-19 vaccines, and analyzed how such trends relate to popular topics found on Twitter.
Methods: We collected over 300,000 geotagged tweets in the US from March 1, 2020 to February 28, 2021. We examined the spatiotemporal patterns of public sentiment and emotion over time at both national and state scales and identified three phases along the pandemic timeline with the sharp changes in public sentiment and emotion. Using sentiment analysis, emotion analysis (with cloud mapping of keywords), and topic modelling, we further identified 11 key events and major topics as the potential drivers to such changes.
Results: An increasing trend of positive sentiment in conjunction with the decrease in negative sentiment are generally observed in most states, reflecting the rising confidence and anticipation of the public towards vaccines. The overall tendency of the eight types of emotion implies that the public trusts and anticipates the vaccine. This is accompanied by a mixture of fear, sadness and anger. Critical social/international events and/or the announcements from political leaders and authorities may have potential impacts on the public opinion towards vaccines. These factors help identify underlying themes and validate insights from the analysis.

Conclusions: The analyses of near real-time social media big data benefit public health authorities by enabling them to monitor public attitudes and opinions towards vaccine-related information in a geo-aware manner, address the concerns of vaccine skeptics, and promote the confidence that individuals within a certain region or community have towards vaccines.

Keywords: Twitter; public opinion; COVID-19 vaccines; sentiment analysis; emotion analysis; topic modelling

1. Introduction

As of May 21, 2021, the coronavirus disease 2019 (COVID-19) pandemic has led to more than 160 million confirmed cases and more than three million deaths worldwide [1]. COVID-19 has continued to spread worldwide due to its highly contagious nature, diverse variants, and the mass public's inconsistent adherence to effective public health measures, such as wearing masks and maintaining social distance [2]. Meanwhile, the emergence of asymptomatic cases (which are difficult to detect) has become more frequent, potentially leading to a substantial accumulation in the number of infections over time [3]. As such, it is important to keep COVID-19 vaccines widely available and accessible [4].

Since January 2020, scientists and medical experts around the world have been developing and testing COVID-19 vaccines; 16 vaccines have been approved for emergency use around the world so far, but the progress of vaccination has been subject to hesitancy, distrust, and debate. Vaccine hesitancy has been identified by the World Health Organization (WHO) as one of the top ten global health threats in 2019 [5]. In many countries, such hesitancy, along with vaccine misinformation, have presented substantial obstacles towards vaccinating a sufficient amount of the population in order to establish herd immunity [6,7].

Therefore, it is crucial for governments, public health officials, and policy makers to understand the potential drivers that affect public opinion towards COVID-19 vaccines [8]. A number of campaigns against the anti-vaccination activists have been made through multiple channels since January 2020. Notably, the accelerated pace of vaccine development has further heightened public anxieties and could compromise the public's acceptance of the vaccine [9]. However, this acceptance varies across geographic contexts and the pandemic timeline. As governments put more effort into developing strategies for promoting vaccine acceptance and uptake, the key questions regarding the willingness to be vaccinated persist — what are the public's opinions and perceptions towards COVID-19 vaccines and what are the potential drivers that affect such opinions?

The internet and social media have provided rich user-generated data sources, in the form of infodemiology studies [10], in real-time for performing public health surveillance. Social media, especially Twitter, have been considered as major channels for the distribution of health information and opinion exchange, helping people to make intelligent decisions [11,12]. The analysis of big data derived from Twitter has been an emerging trend in recent COVID-19 vaccine related-studies. Geotagged Tweets (hereinafter termed as geotweets) provide a rich volume of cost-effective content, including news, events, public comments, and the locational information of Twitter users. Through sentiment analysis and topic modeling methods that have been widely used in existing studies, qualitative tweet contents can be retrieved to reflect public opinions and attitudes towards COVID-19 vaccines. Additionally, the users' location information enables researchers to investigate the spatiotemporal patterns of the public's opinions and attitudes. In general, existing studies have investigated people's reactions towards COVID-19 vaccines, with a geographical emphasis on the US [13-18]. Some papers have also studied other countries in the world, including China [19], South Africa [20], Australia [21], United Kingdom [22-23], Canada [24], Africa [25] and to a global scale

[26]. However, the study period of these works is relatively limited to or predominantly focused on the early stage of the pandemic or up to the end of 2020. None of these studies cover early 2021, the period of implementing mass systemic vaccine distribution. Furthermore, although sentiment analysis and topic modeling have been broadly applied, what remains less explored are the potential drivers that induce the change in public sentiment and opinion on vaccines, such as important events and the announcement of political leaders (e.g., the propaganda of vaccine success or vaccine conspiracy theories). There is a pressing need to investigate public opinion towards COVID-19 vaccines across a longer timeline and to explore the potential drivers that influence the change in such opinion over time.

To address these knowledge gaps, this study aims to analyze the spatiotemporal patterns of the public sentiment and emotion and explore the keywords and major topics of the tweets regarding COVID-19 vaccines that were tweeted by Twitter users. Drawing on more than 300,000 geotweets from March 1, 2020 to February 28, 2021 in the US, we employed sentiment and emotion analysis at both the national and state level. We identified three phases along the pandemic timeline that display sharp changes in public sentiment and emotion. Using cloud mapping of keywords and topic modeling, we identified eleven key events and major topics as the potential drivers that induced such changes. Findings from this study can help governments, policymakers, and public health officials understand factors that motivate and cause hesitance in public towards vaccination. With this understanding, these entities can better design potential interventions during their vaccination campaigns.

2. Data and Methodology

2.1 Data

Using the Twitter streaming Application Programming Interface (API), Harvard CGA (Center for Geographic Analysis) collected geotweets from March 1, 2020, to February 28, 2021. Geotweets provide the location information of the user-defined places. If users activate the Global Positioning System (GPS) function in Twitter, their longitude and latitude are provided. We used the keyword "vaccin*" to query vaccine-related tweets, generating a total of 308,755 geotweets. In the results, 1.43% (44,118) of geotweets' geographic locations are at a state-level (i.e., MA, USA), and others are geocoded at a city-level (i.e., Cambridge, MA) or at a finer geographical level (i.e., Uptown Coffee, Oxford, MS). We then conducted a series of data preprocessing to the geotweets' contents. First, we generalized the variations of COVID-related terms to "COVID-19", including "corona", "covid", "covid19" and "coronavirus"; second, we removed the unrelated website links from the searching results, including the links starting with the fragment of "https"; third, we removed punctuations (e.g., a period, question mark, comma, colon, and ellipsis) and other key symbols (e.g., a bracket, single and double quote) and converted capital case letters into lower case letters; fourth, we removed the inflectional endings (e.g., "ly") and returned words into their root or their dictionary form (e.g., "peopl" from people, "dai" from daily, and "viru" from virus), by employing the function of *word lemmatization* provided in the Python package Natural Language Toolkit 3.6.2 [27].

2.2 Methodology

To explore the spatiotemporal patterns of the public sentiment and emotion towards COVID-19 vaccines, we conducted four sets of analyses, including sentiment analysis, emotion analysis, topic modeling, and word cloud mapping. For the sentiment analysis, we applied Valence Aware Dictionary for Sentiment Reasoning (VADER), a well-known rule-based model, to estimate sentiment compound scores [28]. The sentiment compound score is computed by summing the score of each word in the lexicon, adjusted according to the rules. The rules embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. Then the score is normalized to be between -1 (most extreme negative) and +1 (most extreme positive). To reclassify sentences as positive, neutral, or negative sentiment, threshold values are set as follows: a tweet with a compound score larger than 0.05 is classified as positive sentiment; a tweet with a compound score smaller than -0.05 is classified as negative sentiment; otherwise, it is classified as neutral sentiment [28]. We then cross-tabulated the three types of sentiment on a daily and weekly basis with the number of geotweets. We generated line graphs at the national level and in the top 10 states with the largest number of geotweets.

Different from sentiment analysis, which detects positive, neutral, or negative feelings from tweet contents, emotion analysis aims to recognize the types of feelings more specifically through the content expression, such as anger, fear, and happiness. The emotion analysis of this study was performed based on the National Research Council Canada Lexicon (NRCLex) [29]. NRCLex examines four pairs of primary bipolar emotions: joy (feeling happy) versus sadness (feeling sad); anger (feeling angry) versus fear (feeling of being afraid); trust (stronger admiration and weaker acceptance) versus disgust (feeling something is wrong or nasty); and surprise (being unprepared for something) versus anticipation (looking forward positively to something). We then examined the temporal patterns of these eight types of emotion at both national and state levels.

In order to investigate the potential drivers of such changes, we applied the Latent Dirichlet Allocation (LDA) model [30] to detect popular topics based on a certain number of key dates as the turning points of sentiment scores or with the sharp change of the number of geotweets. The LDA model generates automatic summaries of topics in terms of a discrete probability distribution over words for each topic, and further infers per-document discrete distributions over topics [31]. Each topic is treated as a cluster, and each document will be assigned to a cluster that represents its dominant topic. LDA is an unsupervised algorithm [32], meaning that, prior to running the model, users need to predefine the number of topics. To estimate the optimal number of topics, we used the Python package [33] and pyLDAvis [34] to compare the results with topic numbers from 3 to 10, and found that the smallest overlap among topics occurs when the topic number is 3. We further visualized the topic modeling results in bar graphs with the Y-axis, which indicates the top ten keywords associated with that topic, and the X-axis, which shows the weight of each keyword (to reveal the extent to which a certain keyword contributes to that topic). Based on the top ten most relevant keywords to each topic, we generalized and presented a name of each topic at the bottom of each graph.

We then categorized the study period into three phases based on two iconic events: the results of Phase 1 clinical trials of Moderna that were published in *The New England Journal of Medicine* on July 14, 2020 [35] and the first COVID-19 vaccine shots that were given in the US on December 14, 2020 [36]. Phase 1, dating from March 1, 2020 to July 13, 2020, is the stage in which the public was waiting for official announcements regarding the effectiveness of COVID-19 vaccines; Phase 2, ranging from July 14, 2020 to December 13, 2020, is when the positive news of COVID-19 vaccine development began to arrive; Phase 3 starts from December 14, 2020, when the first vaccine shots were given in the US. We then aggregated sentiment scores at the state level and analyzed the changes in sentiment over the three phases in the top 10 states. Finally, we produced word cloud maps over three pre-defined phases based on the frequency of keywords appearing in Tweets contents, with the size of a keyword reflecting its frequency and popularity.

4.Results

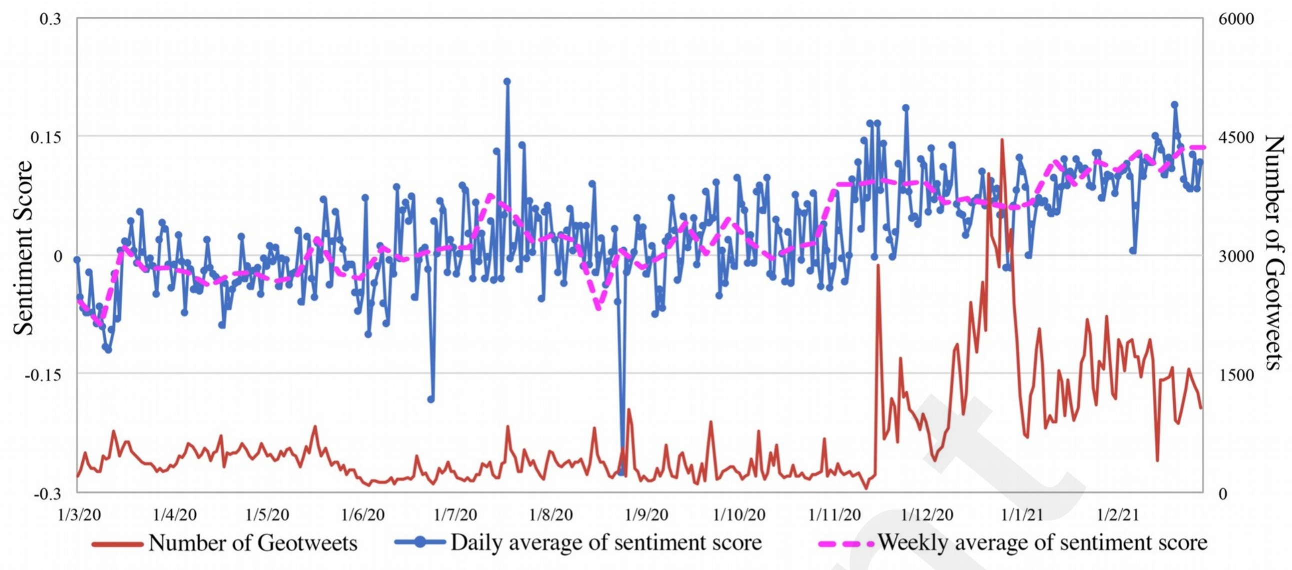
4.1 Sentiment analysis and topic modeling

Figure 1 shows the overall trends of the weekly sentiment scores, unveiling the increased positive attitude towards COVID-19 vaccines within the study period. Eleven key dates are identified as the turning point in sentiment scores or in the number of geotweets. Correspondingly, a total of 33 topics on these 11 key dates are summarized and presented in Figure 2-4. In Phase 1, changes in the sentiment score have been relatively stable, except for a sharp drop on June 21, 2020. This drop could have resulted from the misinformation and conspiracy theories related to Bill Gates. Vaccine-adverse conspiracy related to Gates claimed that the pandemic is a cover for his plan to implant trackable microchips made by Microsoft [37]. Topic modeling suggests that Gates was referred to as “satan”, “terrorist”, and “evil” on that day (Figure 2).

In Phase 2, the first stimuli was observed on July 14, 2020, when the results of Phase 1 clinical trials of Moderna were published [35]. However, we did not observe a dramatic change in sentiment score until July 15, 2020, when Donald Trump tweeted “Great News on Vaccines!” [35]. Topic modeling suggests that keywords related to “good”, “trial”, “promis”, and “test” were widely discussed on July 15, 2020 (Figure 2). Speculation suggests that compared to key events of the development of COVID-19 vaccines, comments from public figures on vaccination could trigger bigger changes in public sentiment.

Another sharp increase in sentiment score was observed on July 22, 2020, when the partnership between Pfizer and the US government accelerated the production and delivery of 100 million doses of COVID-19 vaccines [38]. The keywords “pfizer”, “govern”, and “million” were widely discussed and identified through topic modeling (Figure 3). On August 20, 2020, the sentiment score dropped dramatically after Kamala Harris formally accepted the Democrats’ vice-presidential nomination at the 2020 Democratic National Convention. Harris advocated, “There is no vaccine for racism”, mentioning the context of the racism protest for George Floyd and Breonna Taylor [39]. Of the keywords, “racism” and “kamala” were observed through topic modeling. Another increase in sentiment score appeared on November 9, 2020, when Pfizer announced that its vaccine is 90% effective (Figure 3) [40]. On the same day, Trump tweeted “STOCK MARKET UP BIG, VACCINE COMING SOON. REPORT 90% EFFECTIVE. SUCH GREAT NEWS!”. Amid positive news from Pfizer, people questioned whether Pfizer purposefully released study results after Election Day, though Pfizer’s CEO claimed that the releasing time has nothing to do with politics [41]. On that day, widely discussed keywords included “trump”, “pfizer”, and “elect” (Figure 3).

In Phase 3 on December 14, 2020, an increased sentiment score was observed when an Intensive Care Unit (ICU) nurse received the first COVID-19 vaccine in New York. On the same day, the Electoral College voted to cement Biden’s victory over Trump. Discussion regarding COVID-19 vaccines (“pfizer”, “nurs”, “receive”) quickly increased on Twitter, while other related discussions regarding mask wearing (“wear” and “mask”) and the presidential election (“house”, “trump”, “biden”) remained popular (Figure 4). By December 18, 2020, the sentiment score remained high as both Pfizer and Moderna were authorized for emergency use by The US Food and Drug Administration [42]. Trump tweeted “Moderna vaccine overwhelmingly approved. Distribution to start immediately,” Additionally, the fact that the former Vice President Pence and second lady Karen Pence received COVID-19 vaccine [43] was widely discussed (“penc” and “receiv”). Expectations for the COVID-19 vaccines were also discussed (“need” and “want”) (Figure 4). On January 30, 2021, The Department of Defense paused a plan to give COVID-19 vaccines to detainees in the Guantanamo Bay prison camp [44], which raised queries of COVID-19 vaccine delivery, leading to a moderate decrease in the sentiment score. Keywords were observed, including “terrorist” and “distribut” through topic modeling (Figure 4). On February 12, 2021, an increased sentiment score was observed after the Biden administration announced the purchase of 200 million COVID-19 vaccine doses from Pfizer and Moderna [45]. Discussion surrounding the administration of COVID-19 vaccines was extensive (“wait”, “get”, “need”) (Figure 4). Topic modeling also suggests that complaints were pervasive (“teacher”, “school” and “get”) (Figure 4), because teachers were not prioritized for vaccination in states despite CDC’s recommendation.



1. June 21, 2020: COVID-19 vaccine conspiracy of Bill Gates which calls Gates as an international terrorist with fake vaccine. (a sharp decline of sentiment score)
2. July 14, 2020: Publish initial Phase I/II clinical trial data on July 14th for Moderna.
3. July 15, 2020: Donald Trump tweets: Great news on vaccines! (a sharp increase of sentiment score)
4. July 22, 2020: The US Government engages Pfizer to produce millions of doses of COVID-19 vaccine. (a sharp increase of sentiment score)
5. August 20, 2020: Kamala Harris has formally accepted the Democrats' vice-presidential nomination at the 2020 Democratic National Convention; Mrs. Harris said voters must elect a president who will brin something different and that there was "no vaccine for racism". (a sharp decline of sentiment score)
6. November 9, 2020: Pfizer shows more than 90% effectivity rate in preventing COVID19 in patients without evidence of prior SARS-CoV-2 infection. (a sharp increase of geotweets number)
7. November 18, 2020: Primary efficacy analysis demonstrates BNT162b2 to be 95% effective against COVID-19. (a moderate increase of sentiment score)
8. December 14, 2020: ICU nurse in New York among the first people in the US to get authorized coronavirus vaccine. (a sharp increase of geotweets number)
9. December 18, 2020: The US Food and Drug Administration issued an emergency use authorization (EUA) of Moderna. (a sharp increase of geotweets number)
10. January 30, 2021: A day after The Department of Defense plan to give the COVID-19 vaccine to detainees in the Guantanamo Bay prison camp (a moderate decline of sentiment score)
11. February 12, 2021: The Biden administration announces the purchase of 200 more million doses of Moderna and Pfizer vaccines, bringing the country's total purchase at this point to 600 million, or enough t vaccinate 300 million people. (a moderate increase of sentiment score)

Figure 1. Sentiment scores and the number of geotweets over the entire study timeline at the national level

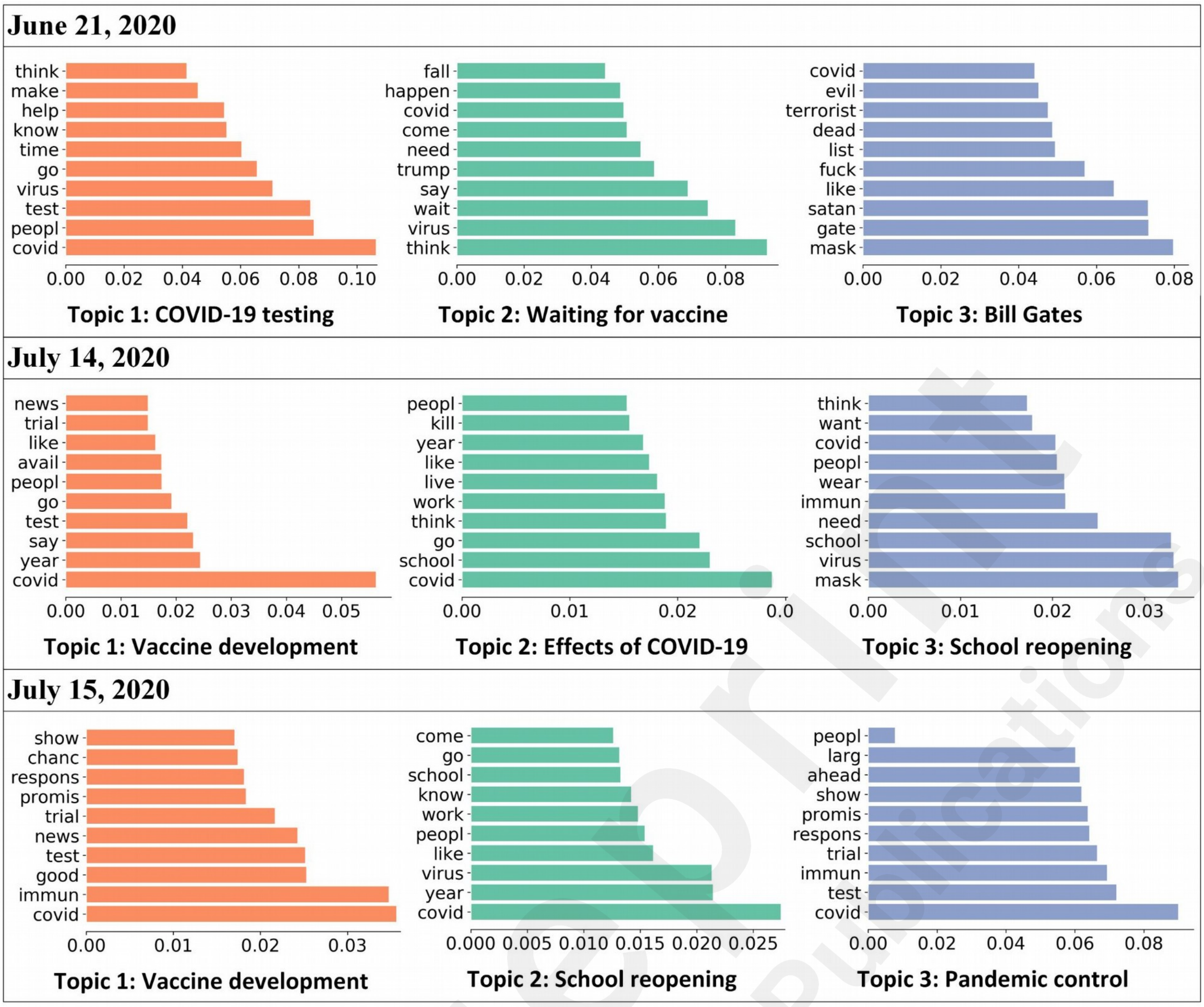


Figure 2. Three topics discussed on key dates: June 21, July 14, and July 15, 2020

Note: Within each graph, X axis indicates the weight of each word; Y axis indicates the top ten keywords most relevant to that topic.

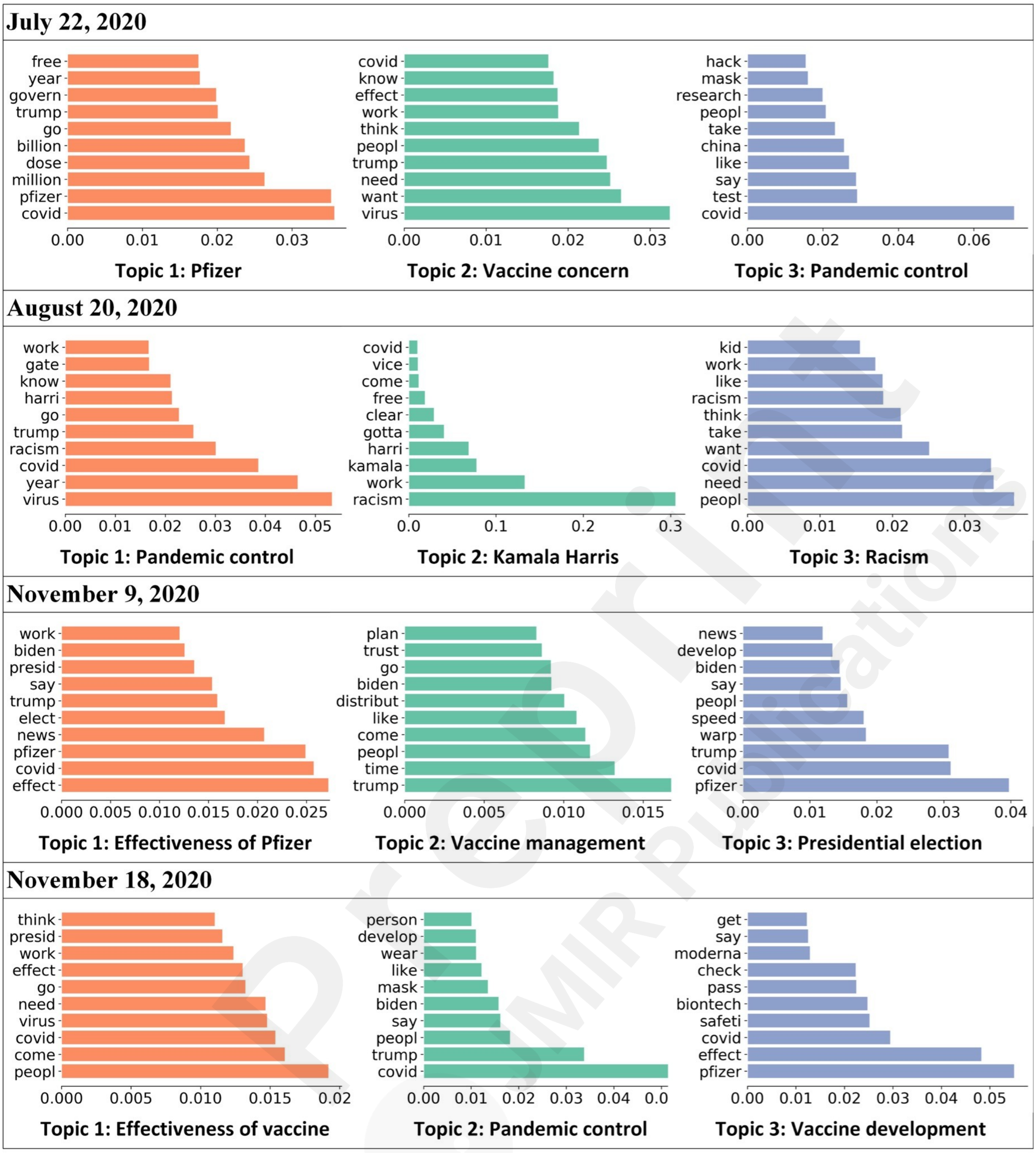


Figure 3. Three topics discussed on key dates: Jul. 22, Aug. 20, Nov. 9, and Nov. 18, 2020
Note: Within each graph, X axis indicates the weight of each word; Y axis indicates the top ten keywords most relevant to that topic.

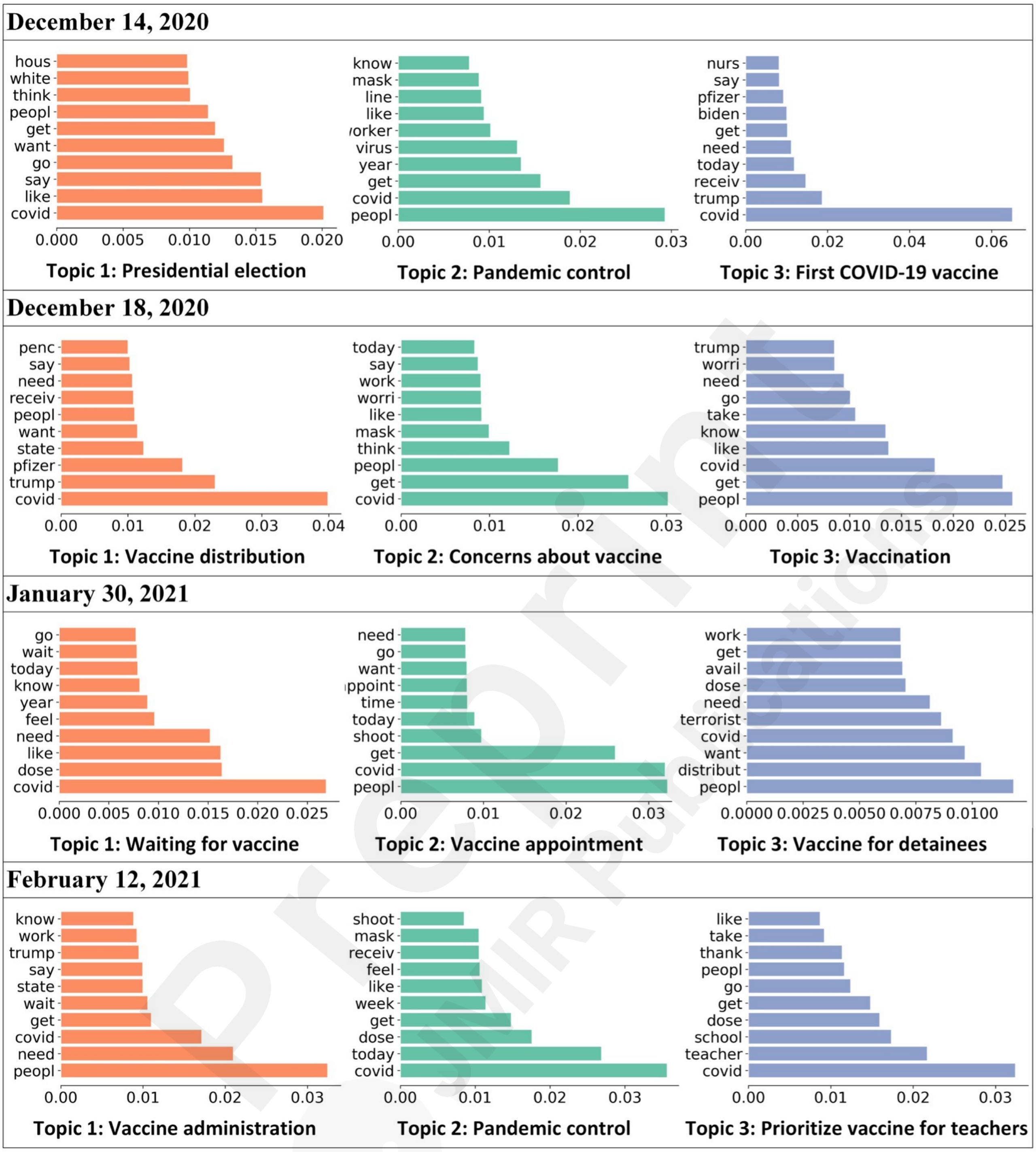


Figure 4. Three topics discussed on key dates at phase 3

Note: Within each graph, X axis indicates the weight of each word; Y axis indicates the top ten keywords most relevant to that topic.

We then broke down the sentiment scores by the state in tandem, along with the pandemic timeline. We presented the results in the top ten states with the largest number of geotweets (Figure 5), including California, New York, Texas, Florida, Illinois, Ohio, North Carolina, Pennsylvania, Georgia, and Virginia. The temporal patterns of sentiment scores vary across states, with more obvious fluctuations before November 2020 in Illinois, Ohio, North Carolina, Georgia, Pennsylvania, and Virginia. A number of sharp decreases in sentiment scores were observed in June 2020 in Illinois, North Carolina, Ohio, Pennsylvania, Georgia, and Virginia, in line with the tendency of sentiment drops at the national level. The states with relatively larger numbers of geotweets (i.e., California, New York, Texas, and Florida) are more stable in the temporal trends and in sentiment scores, compared to the states with relatively smaller numbers of geotweets (e.g., Ohio, North Carolina, Pennsylvania, Georgia, and Virginia).

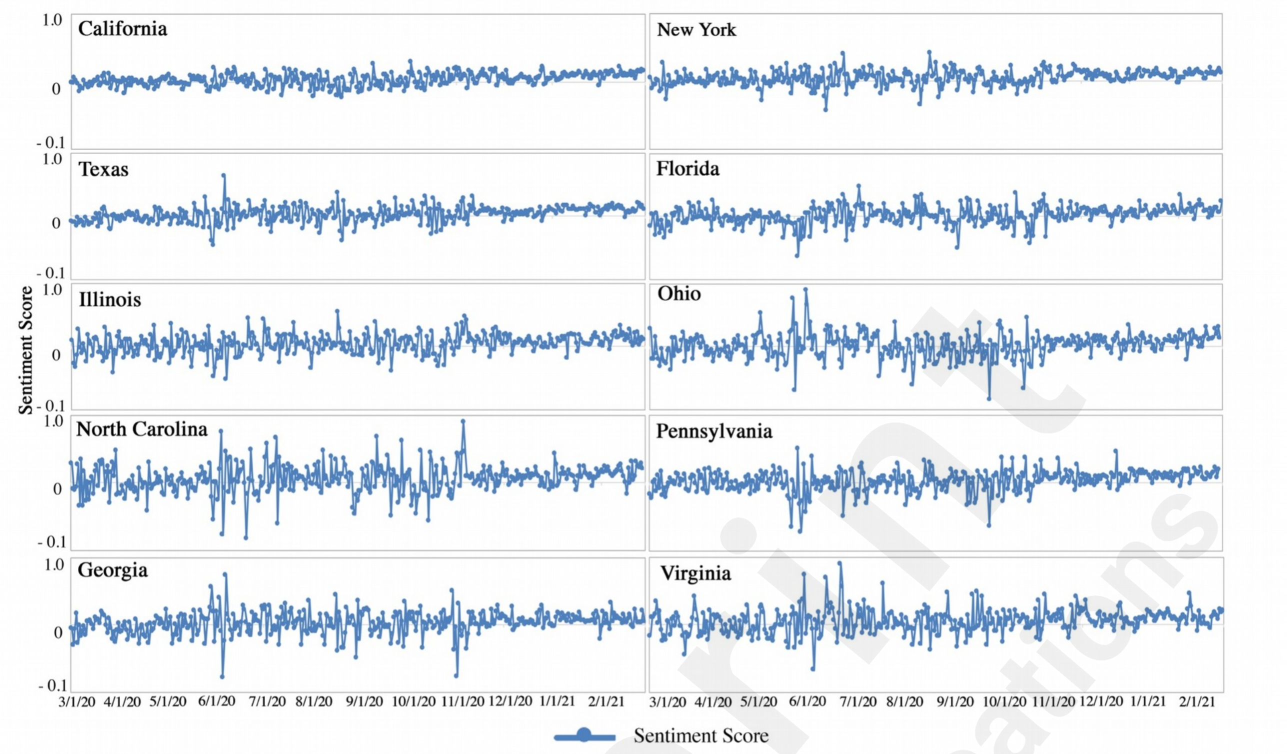


Figure 5. Sentiment scores in ten selected states

We further examined the absolute value of the average positive and negative sentiment scores by states in Figure 6. In the majority of the states, the absolute value of the positive sentiment score is larger than that of the negative sentiment score. The difference between the positive and negative sentiment scores is relatively more obvious in the mainland states of Alabama, Utah, Nebraska, Minnesota, West Virginia (highlighted in dark grey in Figure 6), as well as in Hawaii and Alaska; the potential drivers triggering such differences across states may either relate to information or news spreading locally or be subject to the variations caused by the different sampling size in each state.

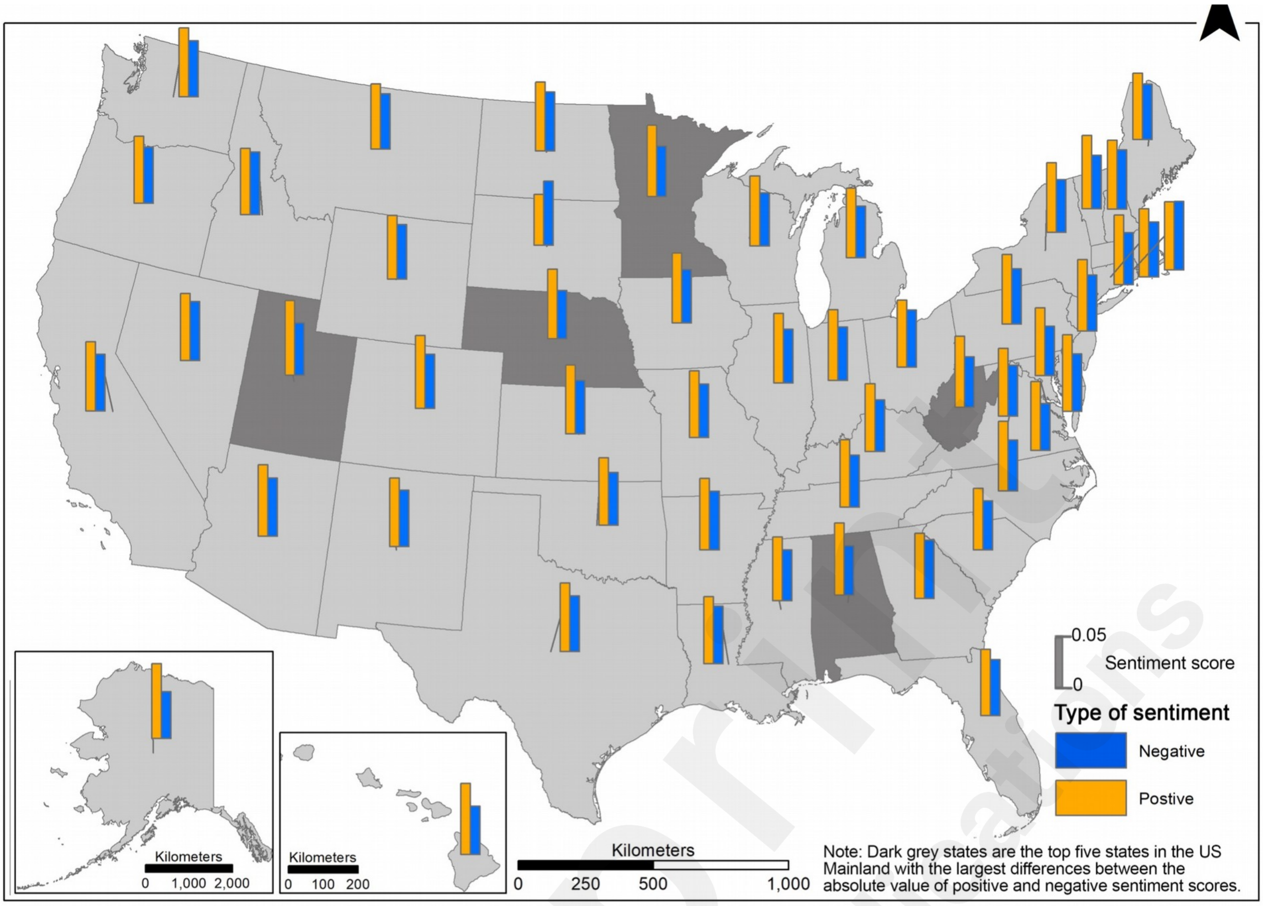


Figure 6. Absolute value of negative and positive sentiment scores at the state level

The change in positive and negative sentiment scores over two periods of time (Phase 1 to Phase 2; Phase 2 to Phase 3) is compared and presented in Figure 7. From Phase 1 to Phase 2, the increase in positive sentiment scores (orange bars) appears in most states, most obviously in South Dakota, followed by North Dakota and Arkansas; meanwhile, the decrease in negative sentiment scores (dark blue bars) is also observed in the majority of states, most obviously in South Dakota and Rhode Island, followed by Montana, North Dakota and Arkansas. From Phase 2 to Phase 3, the decrease in negative sentiment scores (light blue bars) appears in most states, most obviously in Idaho and Rhode Island, followed by North Dakota, Vermont, and New Hampshire. However, the change in positive sentiment scores (red bars) from Phase 2 to Phase 3 varies across states, with a slight increase that is more obviously observed in Idaho, North Dakota, and New Mexico, while a slight decrease is more obviously observed in South Dakota, Rhode Island, and Connecticut. In addition, the magnitude of both positive and negative sentiment scores from Phase 1 to Phase 2 (the height of dark blue and orange bars) is more obvious in most states than that of Phase 2 to Phase 3 (the height of light blue and red bars). It indicates that the fluctuation in people’s opinion towards vaccines becomes less obvious with the gradual development of vaccines and more encouraging news.

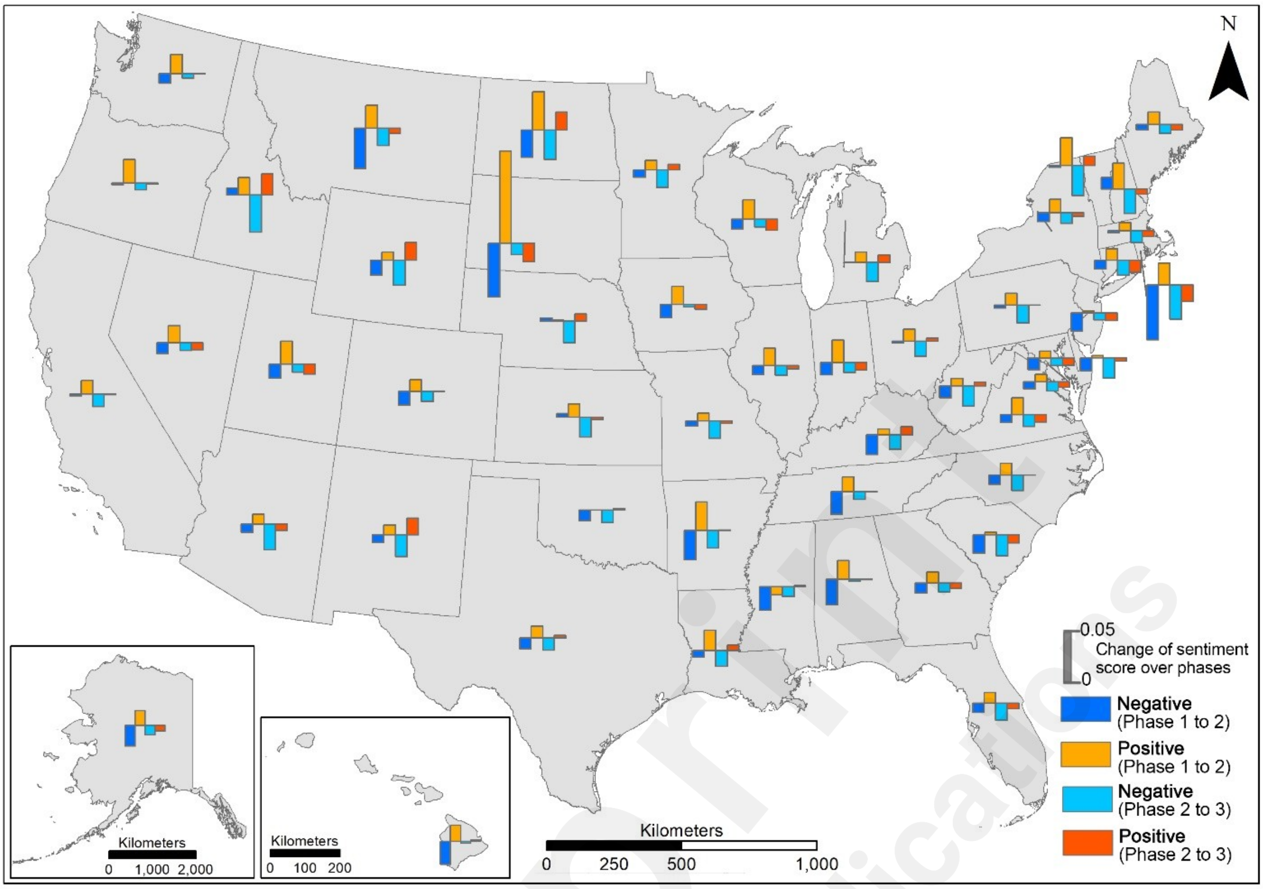


Figure 7. Change of sentiment scores over three phases at the state level

4.2 Emotion analysis

Figure 8 shows the temporal patterns of the eight types of emotion, including joy, trust, anticipation, trust, surprise, disgust, sadness, and fear. Through the vertical comparison of the weekly average trend lines (dash lines), we find that the emotion with the highest weekly average scores along the majority of the timeline is trust (blue dash line), followed by fear, anticipation, sadness, anger, joy, disgust, and surprise. It is worth noting that the weekly average emotion score of fear is higher than that of trust before mid April 2020, possibly due to rapid COVID-19 infection and ineffective control of viral spread at the early stage of the pandemic. These events may have caused fear, uncertainty, or even feelings of panic [46]. Although fluctuations in emotion scores (e.g., local peaks and valleys) can be found within each type of the eight emotions, the general trend implies that the public's trustiness and anticipation towards vaccination is accompanied by a mixture of fear, sadness, and anger.

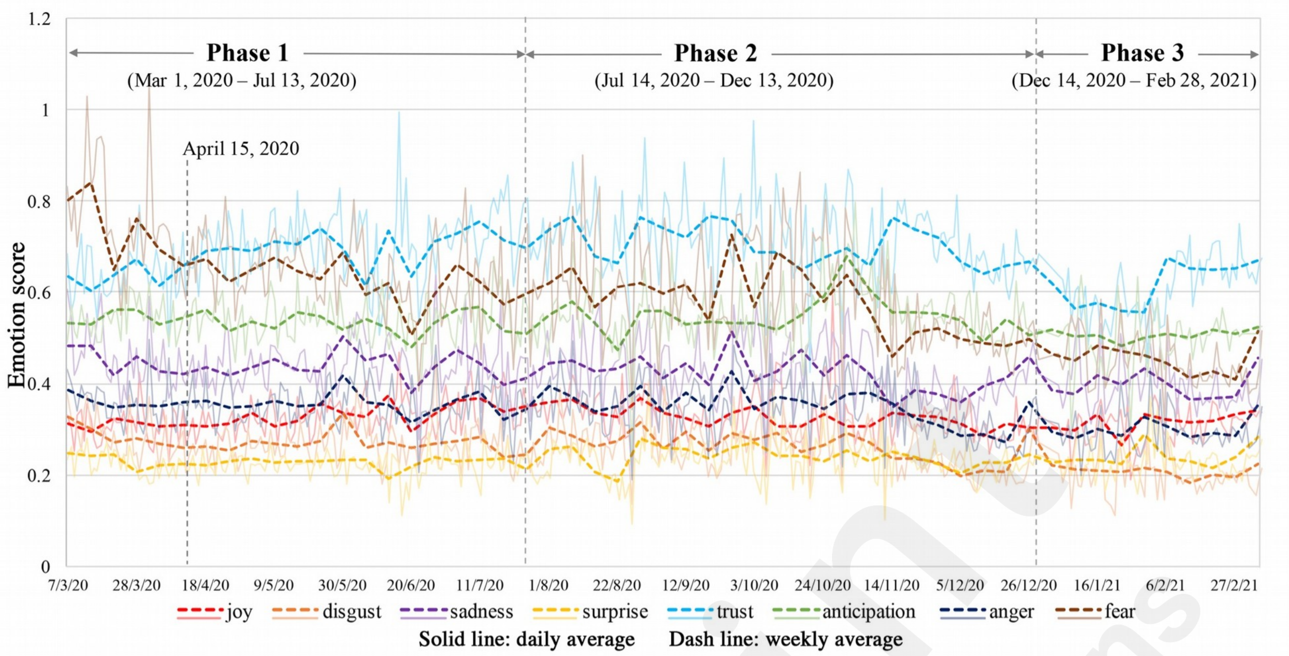


Figure 8. Daily and weekly average of emotion scores over the entire study timeline at the national level

We further investigate the relative distributions of eight emotions in each state, as indicated by the percentage of emotion scores for each type with different colors (Figure 9). The overall patterns of the eight emotions are consistent across most states. Throughout the entire timeline and in each of the three phases of the pandemic, trust is the dominant emotion towards vaccination over the full timeline of the pandemic. It is followed by anticipation, fear, sadness, anger, joy, disgust, and surprise. The state-level patterns largely align with the national pattern as depicted in Figure 9, although there are some exceptions, such as fear overweighting anticipation, joy, and trust (e.g., Washington) and with fear, anger and sadness overweighting other emotions (e.g., Maine). As shown in Figures 10 and 11, the emotion of trust stays consistent over time, while the change in trends for other types of emotion is distinct across phases and by states.

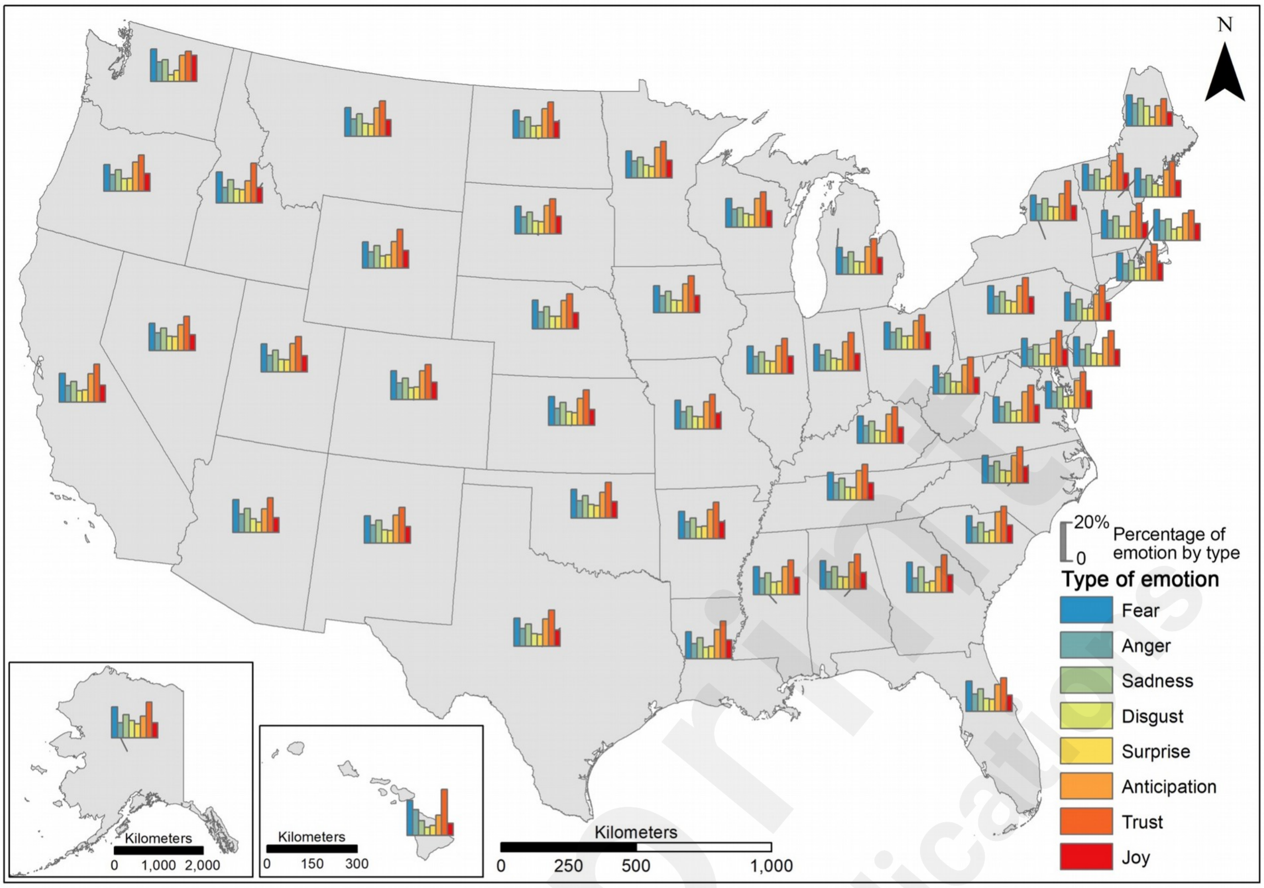


Figure 9. Percentage of eight emotions at the state level

We further compared the change in the percentage of emotion over two periods of time (Phase 1 to Phase 2; Phase 2 to Phase 3). From Phase 1 to Phase 2 (Figure 10), the decrease in fear (dark blue bars) is observed in most states, though its magnitude varies across states. This decrease is most obvious in South Dakota, followed by North Dakota, Arkansas, Mississippi, North Carolina, and South Carolina. The change in anger, sadness, and disgust varies across states, with a general decrease in most states but sporadic increases in some others (e.g., Idaho, New Mexico, and New Hampshire). Furthermore, the combination of a decrease in fear and an increase in joy, trust, and anticipation is observed in most states except South Dakota. Throughout the period from Phase 2 to Phase 3 (Figure 11), it is difficult to generalize the pattern of emotion change across states in terms of its type and magnitude. The increase in joy, trust, anticipation, and surprise along with the decrease in fear, anger, sadness, and disgust are the most notable (high bars) in Idaho and Rhode Island, followed by Missouri, Vermont, and New Hampshire. On the contrary, some states encounter a decrease in trust and anticipation in tandem with an increase in anger and sadness, including South Dakota, North Dakota, Montana, Kansas, Indiana, Maine, and Delaware. The complexity of emotion change from Phase 2 to Phase 3 varies across states, reflecting the diversity in people's opinion and psychological reaction to vaccination, which should be subject to an in-depth investigation of causality.

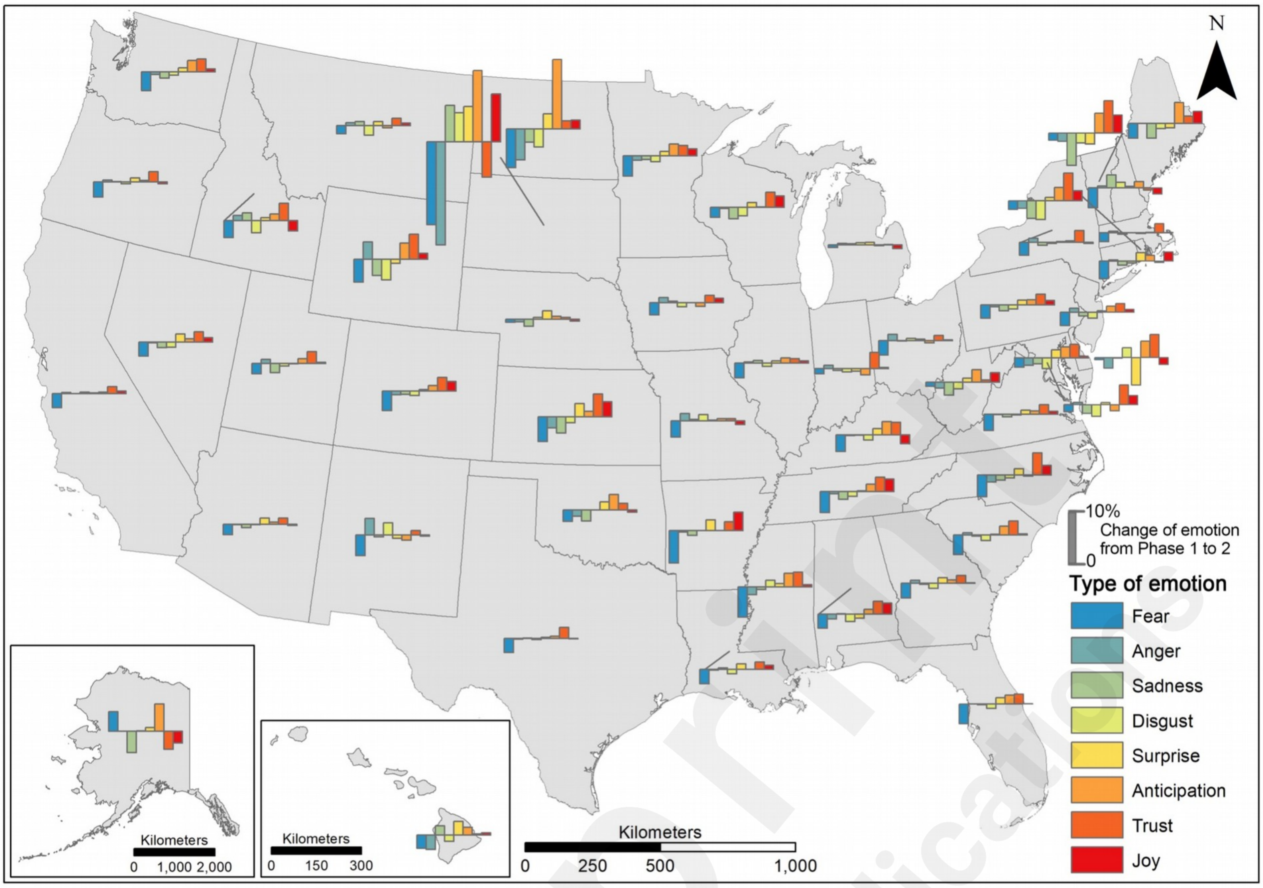


Figure 10: Change of emotion from Phase 1 to Phase 2 at the state level

Figure 11: Change of emotion from Phase 2 to Phase 3 at the state level

4.3 Word cloud visualization

We produced word cloud mappings of 50 popular words associated with positive and negative sentiment over the three phases (Figure 12). The size of a word represents its popularity and the frequency with which it appears in tweets. Among the words associated with the positive sentiment, the popular ones are “hope”, “help”, “thank”, “love”, “safe”, “cure”, and “free”, although the word “peopl” with a more neutral nature, appears to be the largest one. Throughout the three phases, “hope”, “safe”, and “thank” grow larger from Phase 1 to Phase 3; in particular, “thank” becomes the most popular word in Phase 3. On the contrary, “flu”, “death”, “trump”, “fuck”, “lie”, “die”, “kill”, “shit”, and “stupid” are popular words associated with negative sentiment. Over the three phases, “flu” becomes smaller from Phase 1 to Phase 3 whereas “die”, “fuck”, “shit”, and “trump” evolve to be larger from Phase 1 to Phase 3; in particular, “trump” becomes predominant in Phase 2 possibly due to Trump’s increasing popularity, caused by the 2020 Presidential Election. More specifically, while people were waiting for the news of COVID-19 vaccine development during Phase 1, their uncertainties on potential vaccines were reflected in the included keywords, which are related to the coronavirus and public’s frustration of the pandemic (e.g., “viru”, “death”, “cure” and “test”). Some keywords related to the COVID-19 vaccine were also observed, including “hope” and “develop”. Positive news about the development of COVID-19 vaccines appeared in Phase 2, which brought hope as well as misinformation regarding the vaccines to the public. At this stage, more specific information of COVID-19 vaccines was discussed (e.g., “Pfizer”, “effect”, “risk”, “develop”, and “approve”), as compared to Phase 1. With Pfizer and Moderna approved during Phase 3, the public’s attention moved from vaccine development towards vaccine distribution (“distribution”, “wait” and “free”), effectiveness (“safe” and “risk”), and priority (“teacher”). In all three phases, public figures (e.g., “Trump”, “Biden”, and “Bill Gates”) have contributed to hot topics with impacts on both positive and negative sentiment.

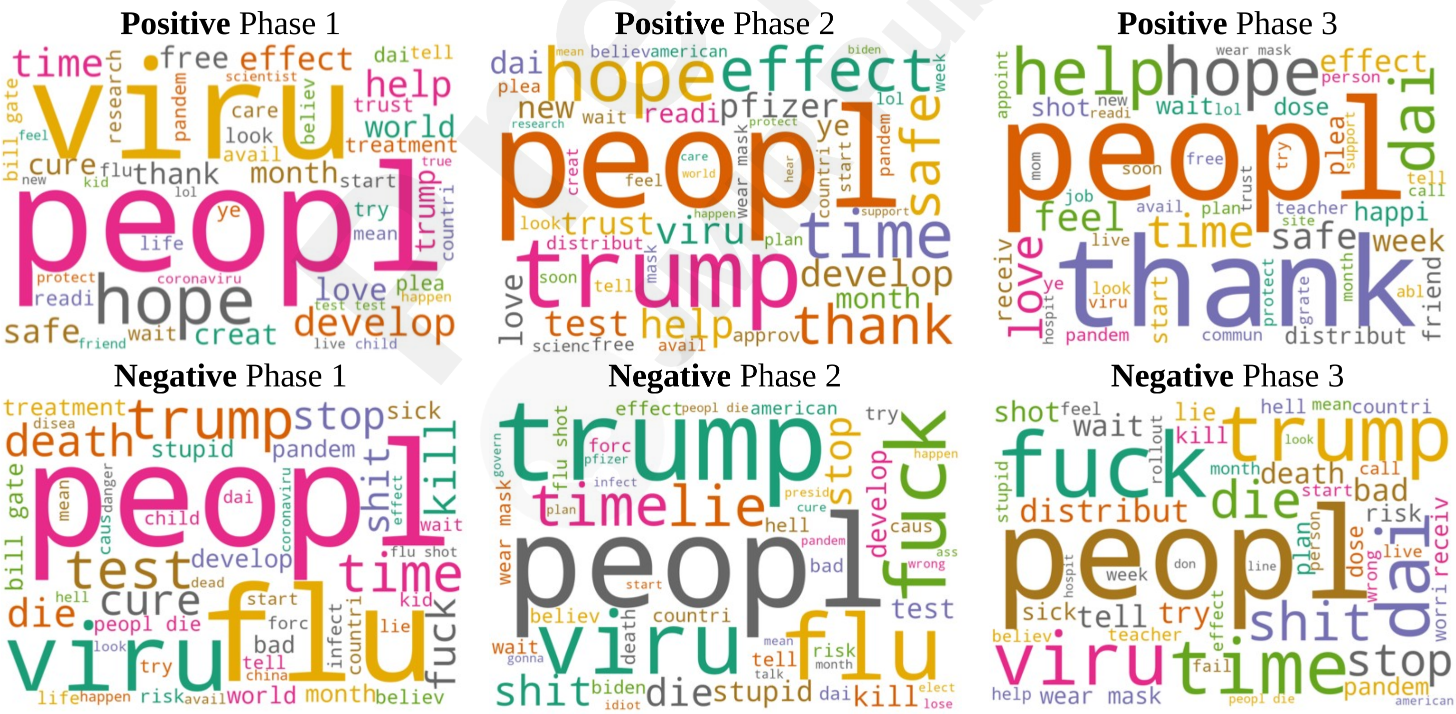


Figure 12. Popular keywords associated with positive and negative sentiment over three phases
(Note: some keywords are extracted and presented as the root of a word, e.g., “peopl” from people, “dai” from daily, and “viru” from virus)

5.Discussion and conclusion

5.1. Key findings

Drawing on geotweets from March 1, 2020 to February 28, 2021, this study examines public opinion on COVID-19 vaccines in the US, through unveiling the spatiotemporal patterns of public sentiment and emotion over time, modeling the popular keywords and topics of Twitter contents, and analyzing the potential drivers of public opinion on vaccines. Our findings indicate that critical social/international events and/or the announcements of political leaders and authorities may have potential impacts on public opinion towards COVID-19 vaccines. Such examples include the vaccine-adverse conspiracy related to Bill Gates on June 21, 2020, the tweet by Donald Trump of “Great News on Vaccines!” on July 14, 2020, Kamala Harris’s advocacy of “There is no vaccine for racism” on August 20, 2020, Biden’s victory of the presidential election over Trump on December 14, 2020, and the authorized emergent usage of Pfizer and Moderna on December 18, 2020. In the proposed three phases over the study time, changes in public opinions on vaccines vary across space and time. More specifically, the fluctuation in people’s sentimental response to the vaccine during the earlier stage of the pandemic is more obvious compared to that of the later stage in the pandemic. However, the increase in positive sentiment in parallel with the decrease in negative sentiment are generally observed in most states, reflecting the rising confidence and anticipation of the public towards COVID-19 vaccines. Furthermore, the public’s eight types of emotion towards the COVID-19 vaccine display a general trend that the combination of trustiness and anticipation with a mixture of fear, sadness, and anger. Moreover, the word cloud mapping shows that positive keywords including “hope”, “safe”, and “thank” grew larger from Phase 1 to Phase 3; in particular, “thank” became the most popular word in Phase 3, indicating the public’s increasingly positive response towards vaccination. In all three phases, public figures (e.g., “Trump”, “Biden”, and “Bill Gates”) have contributed to the most popular topics, impacting both positive and negative sentiment. The aforementioned findings reveal the diversity and complexity of people’s perception on and their psychological reaction towards COVID-19 vaccines, which indicates a further need to be cautious in the interpretation of analytical outcomes and to initiate an additional in-depth investigation of the causality.

Our findings have been partially supported by the current literature. Hussain et al. observed a marked increase in the positive sentiment trend toward COVID-19 vaccines in the US from March 1 to November 22, 2020 [12, 46]. Guntuku et al. and Roy et al. found that republican legislators became more engaged in public discussion on vaccine progress, which may have implications for COVID-19 vaccine uptake among their followers [48, 49]. Germani et al. revealed that anti-vaccination supporters have been heavily engaged in discussions and dissemination of misinformation and conspiracy theories [16]. Considering the limitations (i.e., random sample) inherent in Twitter data, it is important to propose alternative data that provide a complementary understanding of public opinions towards the COVID-19 vaccine to promote vaccination in the US.

5.2. Implication and recommendation

The emergence of the Internet and social media has provided new platforms for persuasion and the rapid spread of (mis)information, which leads to new opportunities and challenges to the communication of vaccine information [50]. There are over 4.3 billion people using the internet nowadays, with 3.8 billion of these individuals as social media users [51]. The popularity of social media platforms coupled with the advent of digital detection strategies benefit public health authorities by enabling the monitoring of public sentiment towards vaccine-relevant information in a geo-aware, (near) real time manner. This can inform more effective policy-making and promote participatory dialogue to establish confidence towards the vaccine, in order to maximize vaccine uptake. Some of our findings add new value to the current scholarship and also provide new insights and suggestions for policy implications with regard to safeguarding societal and economic health.

First, our findings indicate that public figures, especially politicians, play a crucial role in impacting the public's opinions on vaccination. Negative opinions expressed by public figures about a vaccine could impact a large population of people, especially those who do not hold an unswayable opinion [50]. People tend to believe public figures' opinions, as they are elected officials who can influence healthcare systems and are perceived to have more information about a vaccine [52, 53]. Thus, public figures have a responsibility to disseminate accurate health information and should be cautious in expressing their opinions in public. This also highlights the necessity of considering the impact that public figures within vaccine campaigns have on upholding the public's confidence towards the concept of vaccination.

Second, our study reveals that vaccine-adverse conspiracy theories led to a sharp decline in sentiment scores. We need to be aware of the fact that social media platforms with a massive number of users, to some degree, 'disrupted' traditional vaccine information communication [54], allowing anti-vaccination advocates to disseminate misleading messages to a certain audience, whose views on vaccination could be susceptible to change. However, it also means that governmental officials should consider using these platforms to communicate with individuals directly about the vaccination via geo-tailored messages to address the concerns specific to a certain region.

Third, different states demonstrate various trends in sentimental and emotional scores. Our geospatial analysis and map visualization [55] better portray more aspects of users' attitudes towards COVID-19 vaccine. This helps identify the areas with high negative sentimental and emotional scores that require further research to understand the public's underlying fears and concerns of COVID-19 vaccines. We also recommend government and public health agencies to conduct COVID-19 vaccines campaigns in these areas to address people's fears and concerns on COVID-19 vaccines and provide guidance to access the available vaccines.

5.3. Limitation and future work

Our study has several limitations that can be improved in future studies. First, the demographics of Twitter users is typically characterized by younger users who are avid users of mobile phone apps and the Internet, and such users may not be able to reflect the opinion and perception of the general public with varying demographics and socioeconomic status [56, 57]. In addition, the representativeness of Twitter users is not stationary but geographically varying [58, 59]. Like other studies that rely on digital devices, the "Digital Divide" [60] issue needs to be acknowledged. This study only accounts for the reactions from Twitter users to vaccines, which, to some degree, neglect the underprivileged members of society (especially the poor and elderly), inhabitants of rural areas (who do not have access to digital devices), and those who are not willing to share their thoughts on social media platforms. Additionally, the Twitter API that we used allows access to approximately only 1% of the total records [61]. As Padilla et al. demonstrated, tweet sentiment can be impacted based on attraction visits throughout the course of a day [62]. Hence, future works need to increase the sample size to reduce the uncertainties and fluctuations of sentiment scores and emotions. Efforts are also needed to distinguish between local residents and visitors and also conduct investigations under finer temporal scales. In early 2021, Twitter released a new Twitter API (academic research product track) that grants free access to full-archive search with enhanced features and functionality for researchers to obtain more precise, complete, and unbiased data for analyzing the public conversation [63]. Further efforts can be made to explore the potential of this new API in mining public opinions towards COVID-19 vaccines at finer-level scales. Since emotion is a complex and integrated product of human feelings [64], future research efforts can be put into exploring more diverse dimensions of emotion, on top of the eight primary types of emotion. Moreover, disaster and crisis management include four phases, namely prevention (capacity building), preparation (early warning), response (search, rescue, and emergency relief), and recovery (rehabilitation) [65].

Management of the COVID-19 pandemic is still in the response phase. For policy and decision-making endeavors that are pertinent to COVID-19 crisis management, it will be highly beneficial if researchers and practitioners continuously monitor emotional and perspective variations throughout the response and also extend the study timeline to the recovery phase or massive vaccination phase in the post-pandemic years. More importantly, to understand the impact of vaccination on countries, the workflow, and methodology used in this study can be applied in multiple languages to global-scale geotweets.

Conflict of Interests

There is no conflict of interest.

Funding Sources

This research was partially funded by the National University of Singapore Start-up Grant under WBS R-109-000-270-133 awarded to WL, and NSF under Grant 1841403, 2027540, and 2028791. This research has also been partially supported by Ball State University Digital Fellowship Funding, and the Faculty of Arts & Social Sciences Staff Research Support Scheme FY2021 of National University of Singapore (WBS: C-109-000-222-091).

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020 May;20(5):533-4. PMID: 32087114. doi: 10.1016/S1473-3099(20)30120-1.
2. Brown TS, Walensky RP. Serosurveillance and the COVID-19 Epidemic in the US: Undetected, Uncertain, and Out of Control. JAMA. 2020 Aug 25;324(8):749-51. PMID: 32692350. doi: 10.1001/jama.2020.14017.
3. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung SM, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis. 2020 May;94:154-5. PMID: 32179137. doi: 10.1016/j.ijid.2020.03.020.
4. Haynes BF. A New Vaccine to Battle Covid-19. N Engl J Med. 2021 Feb 4;384(5):470-1. PMID: 33378607. doi: 10.1056/NEJMe2035557.
5. World Health Organization. World Health Report 2019: Ten threats to global health in 2019. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019. [accessed May 30, 2021].
6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med. 2021 Feb;27(2):225-8. PMID: 33082575. doi: 10.1038/s41591-020-1124-9.
7. Lane S, MacDonald NE, Marti M, Dumolard L. Vaccine hesitancy around the globe: Analysis of three years of WHO/UNICEF Joint Reporting Form data-2015-2017. Vaccine. 2018 Jun 18;36(26):3861-7. PMID: 29605516. doi: 10.1016/j.vaccine.2018.03.063.

8. Cornwall W. Officials gird for a war on vaccine misinformation. Science. 2020 Jul 3;369(6499):14-5. PMID: 32631873. doi: 10.1126/science.369.6499.14.
9. Fadda M, Albanese E, Suggs LS. When a COVID-19 vaccine is ready, will we all be ready for it?: Springer; 2020. doi: 10.1007/s00038-020-01404-4
10. Zeraatkar, K., & Ahmadi, M. (2018). Trends of infodemiology studies: a scoping review. Health Information & Libraries Journal, 35(2), 91-120.
11. Kwak H, Lee C, Park H, Moon S, editors. What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web; 2010.
12. Paul MJ, Dredze M, editors. You are what you tweet: Analyzing twitter for public health. Fifth international AAAI conference on weblogs and social media; 2011.
13. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the US: Observational Study. J Med Internet Res. 2021 Apr 5;23(4):e26627. PMID: 33724919. doi: 10.2196/26627.
14. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. J Med Internet Res. 2021 Feb 10;23(2):e25431. PMID: 33497352. doi: 10.2196/25431.
15. Engel-Rebitzer E, Stokes DC, Buttenheim A, Purtle J, Meisel ZF. Changes in legislator vaccine-engagement on Twitter before and after the arrival of the COVID-19 pandemic. Hum Vaccin Immunother. 2021 May 10:1-5. PMID: 33970786. doi: 10.1080/21645515.2021.1911216.
16. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. PLoS One. 2021;16(3):e0247642. PMID: 33657152. doi: 10.1371/journal.pone.0247642.
17. Guntuku SC, Purtle J, Meisel ZF, Merchant RM, Agarwal A. Partisan Differences in Twitter Language among US Legislators during COVID-19: A Cross-Sectional Study. J Med Internet Res. 2021 Apr 16. PMID: 33939620. doi: 10.2196/27300.
18. Roy KC, Hasan S. Modeling the dynamics of hurricane evacuation decisions from twitter data: an input output hidden markov modeling approach. Transportation research part C: emerging technologies. 2021;123:102976. doi:10.1016/j.trc.2021.102976
19. Wang J, Zhou Y, Zhang W, Evans R, Zhu C. Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging

Data. *J Med Internet Res*. 2020;22(11):e22152. Published 2020 Nov 26. doi:10.2196/22152

20. Mutanga MB, Abayomi A. Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. African Journal of Science, Technology, Innovation and Development. 2020:1-10. doi: 10.1080/20421338.2020.1817262
21. 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 May 19;23(5):e26953. PMID: 33886492. doi: 10.2196/26953.
22. 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;9:33203-23. doi:10.1109/ACCESS.2021.3059821
23. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the US: Observational Study. J Med Internet Res. 2021 Apr 5;23(4):e26627. PMID: 33724919. doi: 10.2196/26627.
24. 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 Apr 13;23(4):e26874. PMID: 33769946. doi: 10.2196/26874.
25. Gbashi S, Adebo OA, Doorsamy W, Njobeh PB. Systematic Delineation of Media Polarity on COVID-19 Vaccines in Africa: Computational Linguistic Modeling Study. JMIR Med Inform. 2021 Mar 16;9(3):e22916. PMID: 33667172. doi: 10.2196/22916.
26. Chang CH, Monselise M, Yang CC. What Are People Concerned About During the Pandemic? Detecting Evolving Topics about COVID-19 from Twitter [published online ahead of print, 2021 Jan 17]. *J Healthc Inform Res*. 2021;1-28. doi:10.1007/s41666-020-00083-3
27. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit: " O'Reilly Media, Inc."; 2009. ISBN: 0596555717.
28. Hutto C, Gilbert E, editors. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media; 2014.
29. Mohammad SM, Turney PD. Nrc emotion lexicon. National Research Council, Canada. 2013;2. http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf.
30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning

research. 2003;3:993-1022. doi: *10.5555/944919.944937*

31. Hu T, She B, Duan L, Yue H, Clunis J. A systematic spatial and temporal sentiment analysis on Geo-Tweets. IEEE Access. 2019;8:8658-67. doi: 10.1109/ACCESS.2019.2961100.
32. Yan Y, Feng C-C, Huang W, Fan H, Wang Y-C, Zipf A. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science. 2020;34(9):1765-91. doi: 10.1080/13658816.2020.1730848
33. Topic modelling for humans, Gensim,2021. https://radimrehurek.com/gensim/. [accessed May 30, 2021]
34. bmabey. pyLDAvis. https://github.com/bmabey/pyLDAvis. 2014 [accessed May 30, 2021]
35. Jackson LA, Anderson EJ, Rouphael NG, et al. An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl J Med*. 2020;383(20):1920-1931. doi:10.1056/NEJMoa2022483
36. Ben G, Ariana EC, Josh W, and Griff W. The weapon that will end the war': First coronavirus vaccine shots given outside trials in U.S. 2020. https://www.washingtonpost.com/nation/2020/12/14/first-covid-vaccines-new-york/. [accessed May 30, 2021].
37. Goodman J, Carmichael F. Coronavirus: Bill Gates 'microchip'conspiracy theory and other vaccine claims fact-checked. BBC Reality Check May. 2020.
38. **HHS Press Office. U.S. Government Engages Pfizer to Produce Millions of Doses of COVID-19 Vaccine. https://www.hhs.gov/about/news/2020/07/22/us-government-engages-pfizer-produce-millions-doses-covid-19-vaccine.html. [accessed May 30, 2021].**
39. Nagourney A, Glueck K. Kamala Harris Takes the Spotlight, a Moment for Her and History. New York Times. Aug 19, 2020. https://www.nytimes.com/2020/08/19/us/politics/kamala-harris-dnc.html [accessed May 30, 2021].
40. Callaway, E. What Pfizer's landmark COVID vaccine results mean for the pandemic. Nature: Nov 09, 2020. https://www.nature.com/articles/d41586-020-03166-8 [accessed May 30, 2021].
41. Benveniste, A. Pfizer CEO: Our vaccine timing had nothing to do with politics. CNN2020. https://www.cnn.com/2020/11/09/business/pfizer-covid-vaccine/index.html. [accessed May 30, 2021].
42. U.S. Food and Drug Administration. Pfizer-BioNTech COVID-19 Vaccine. Dec 11, 2020. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine [accessed May 30, 2021].

43. Shabad, R. Pence receives Covid vaccine in televised appearance, hails ‘medical miracle’. NBC News. Dec 18, 2020. https://www.nbcnews.com/politics/white-house/pence-set-receive-covid-vaccine-televi3ed-appearance-n1251655. [accessed May 30, 2021].
44. Herridge C, McDonald C. Department of Defense pauses plan to give COVID-19 vaccine to Guantanamo detainees. CBS News. Jan 30, 202. https://www.cbsnews.com/news/dod-pauses-plan-to-give-covid-vaccine-to-guantanamo-detainees/. [accessed May 30, 2021].
45. U.S. Department of Health and Human Services. Biden Administration purchases additional doses of COVID-19 vaccines from Pfizer and Moderna. Feb 11, 2021. https://www.hhs.gov/about/news/2021/02/11/biden-administration-purchases-additional-doses-covid-19-vaccines-from-pfizer-and-moderna.html. [accessed May 30, 2021].
46. Wang S, Liu Y, Hu T. Examining the Change of Human Mobility Adherent to Social Restriction Policies and Its Effect on COVID-19 Cases in Australia. Int J Environ Res Public Health. 2020 Oct 29;17(21). PMID: 33137958. doi: 10.3390/ijerph17217930.
47. Cotfas L-A, 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;9:33203-23.
48. Guntuku S, Purtle J, Meisel Z, Merchant R, Agarwal A. Partisan Differences in Twitter Language among United States Legislators during COVID-19: A Cross-Sectional Study. Journal of Medical Internet Research. 2021.
49. Roy S, Ghosh P, editors. A Comparative Study on Distancing, Mask and Vaccine Adoption Rates from Global Twitter Trends. Healthcare; 2021: Multidisciplinary Digital Publishing Institute.
50. Kang GJ, Ewing-Nelson SR, Mackey L, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine*. 2017;35(29):3621-3638. doi:10.1016/j.vaccine.2017.05.052
51. We Are Social Ltd. Digital in 2020. 2020. https://wearesocial.com/digital-2020. [accessed May 30, 2021].
52. Zhang EJ, Chughtai AA, Heywood A, MacIntyre CR. Influence of political and medical leaders on parental perception of vaccination: a cross-sectional survey in Australia. *BMJ Open*. 2019;9(3):e025866. Published 2019 Mar 26. doi:10.1136/bmjopen-2018-025866
53. Abu-Akel A, Spitz A, West R. The effect of spokesperson attribution on public health message sharing during the COVID-19 pandemic. PloS one. 2021;16(2):e0245100.
54. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new

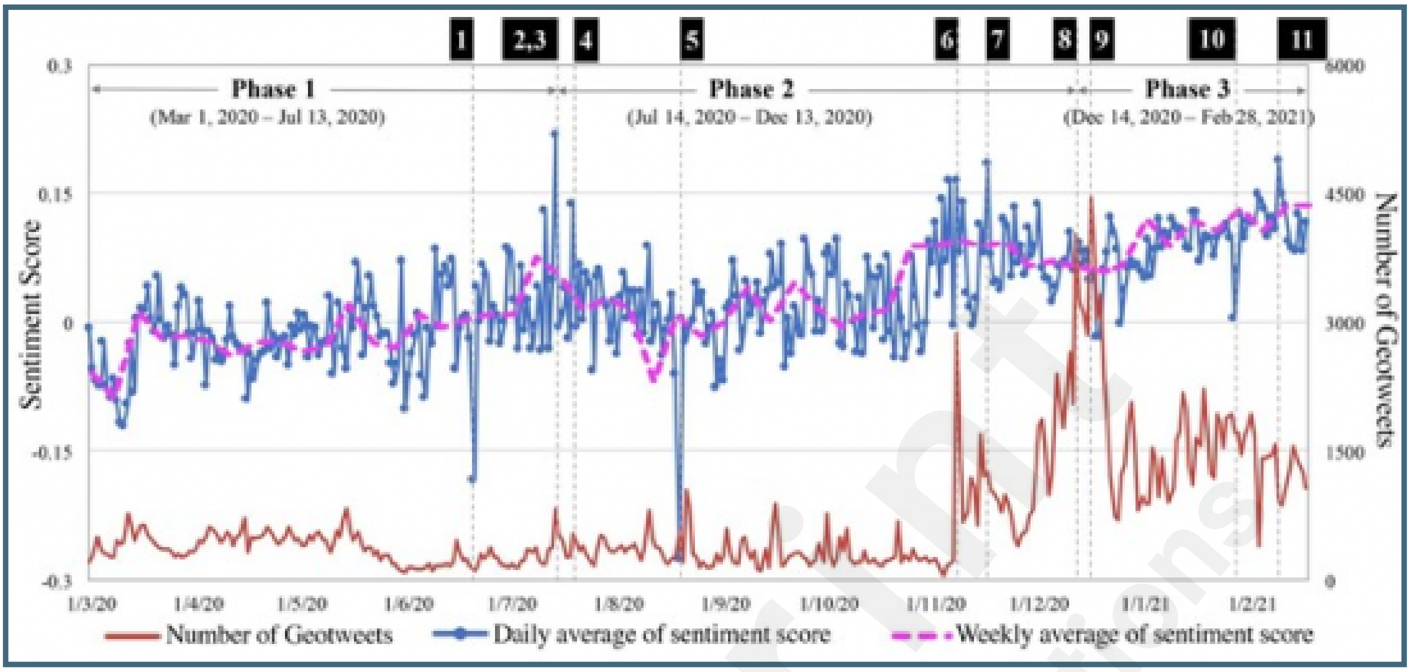
updates for the era of COVID-19 and globalized infectious diseases. *Hum Vaccin Immunother*. 2020;16(11):2586-2593. doi:10.1080/21645515.2020.1780846

55. Mocnik F-B, Raposo P, Feringa W, Kraak M-J, Köbben B. Epidemics and pandemics in maps–the case of COVID-19. Journal of Maps. 2020;16(1):144-52.
56. Wilson K, Atkinson K, Deeks S. Opportunities for utilizing new technologies to increase vaccine confidence. Expert Rev Vaccines. 2014 Aug;13(8):969-77. PMID: 24931799. doi: 10.1586/14760584.2014.928208.
57. Karami A, Kadari RR, Panati L, Nooli SP, Bheemreddy H, Bozorgi P. Analysis of Geotagging Behavior: Do Geotagged Users Represent the Twitter Population? ISPRS International Journal of Geo-Information. 2021;10(6):373.
58. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PloS one. 2015;10(9):e0133505.
59. Jiang Y, Li Z, Ye X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. Cartography and geographic information science. 2019;46(3):228-42.
60. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS One*. 2020;15(11):e0241957. Published 2020 Nov 10. doi:10.1371/journal.pone.0241957
61. Yan Y, Chen J, Wang Z. Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. Applied Geography. 2020;123:102306.
62. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PloS one. 2018;13(6):e0198857.
63. Yan Y, Kuo C, Feng C, Huang W, Fan H, and Zipf, A. Coupling maximum entropy modeling with geotagged social media data to determine the geographic distribution of tourists. *International Journal of Geographical Information Science*, 32(9), pp.1699-1736. doi: 10.1080/13658816.2018.1458989
64. Twitter products for Academic Researchers [database on the Internet]. Twitter. 2021. https://developer.twitter.com/en/products/twitter-api/academic-research. [accessed May 30, 2021].
65. Plutchik R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American scientist. 2001;89(4):344-50.

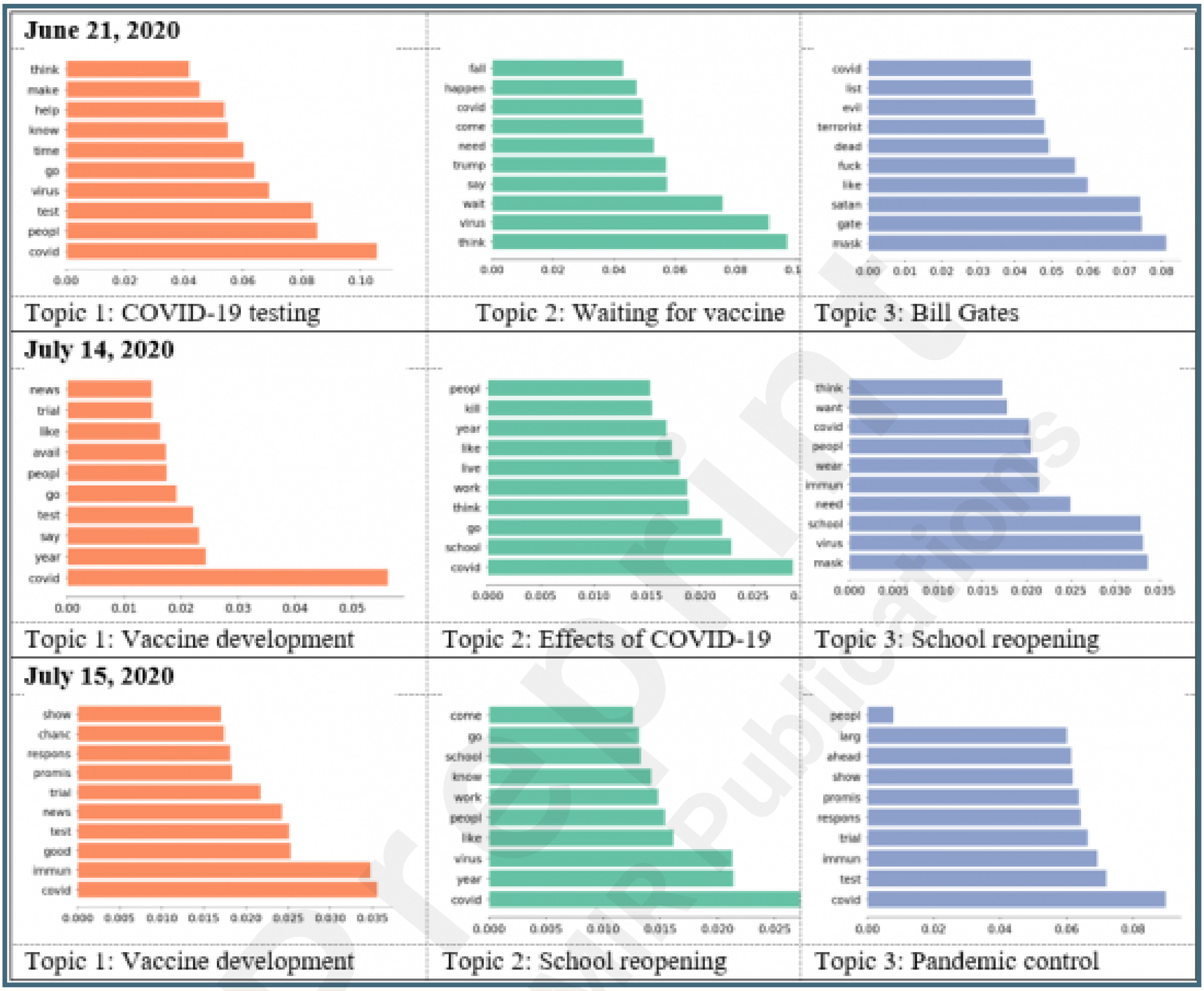
Supplementary Files

Figures

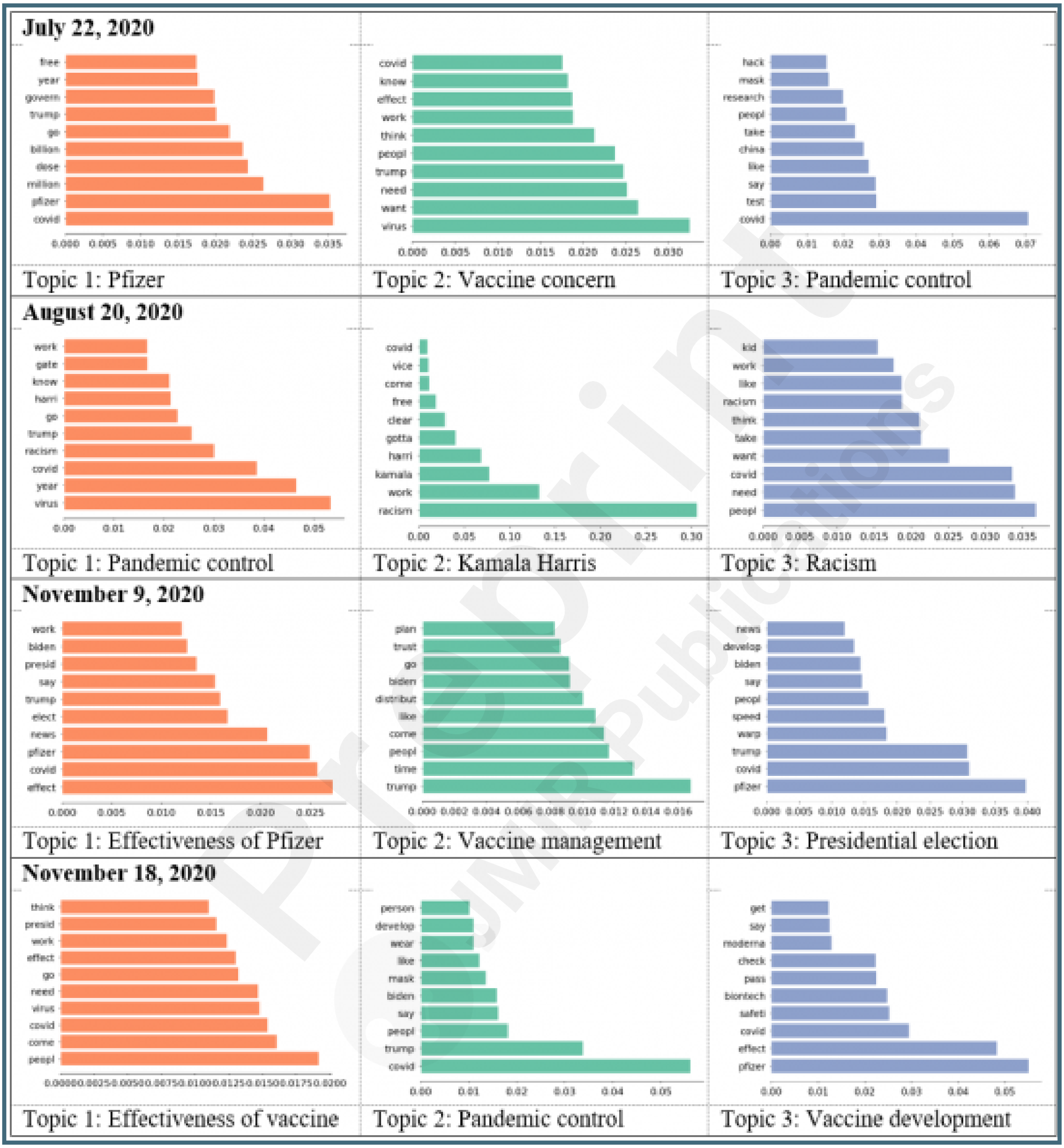
Sentiment scores and the number of geotweet over the entire study timeline with key dates.



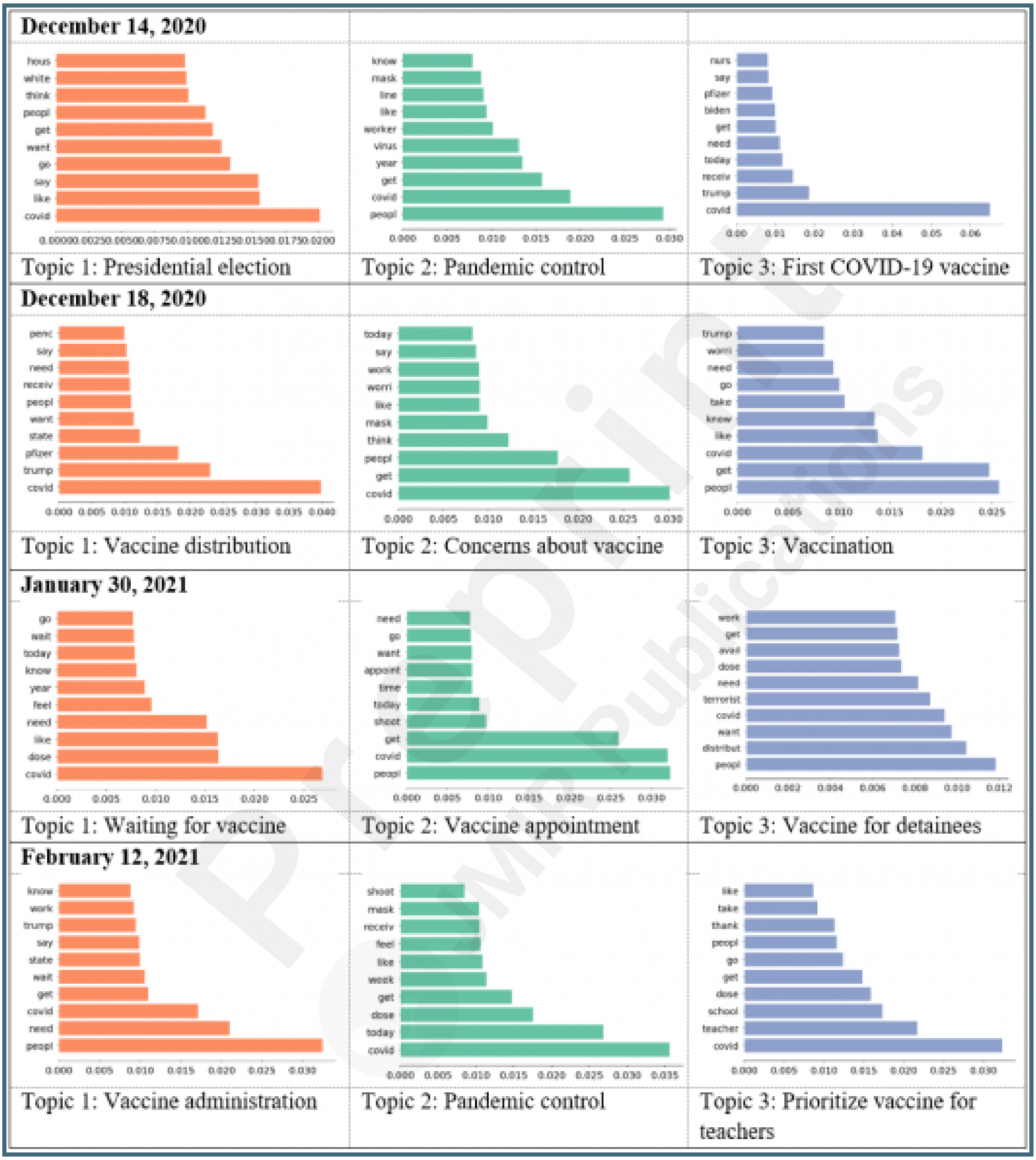
Three topics discussed on key dates: June 21, July 14, and July 15, 2020.



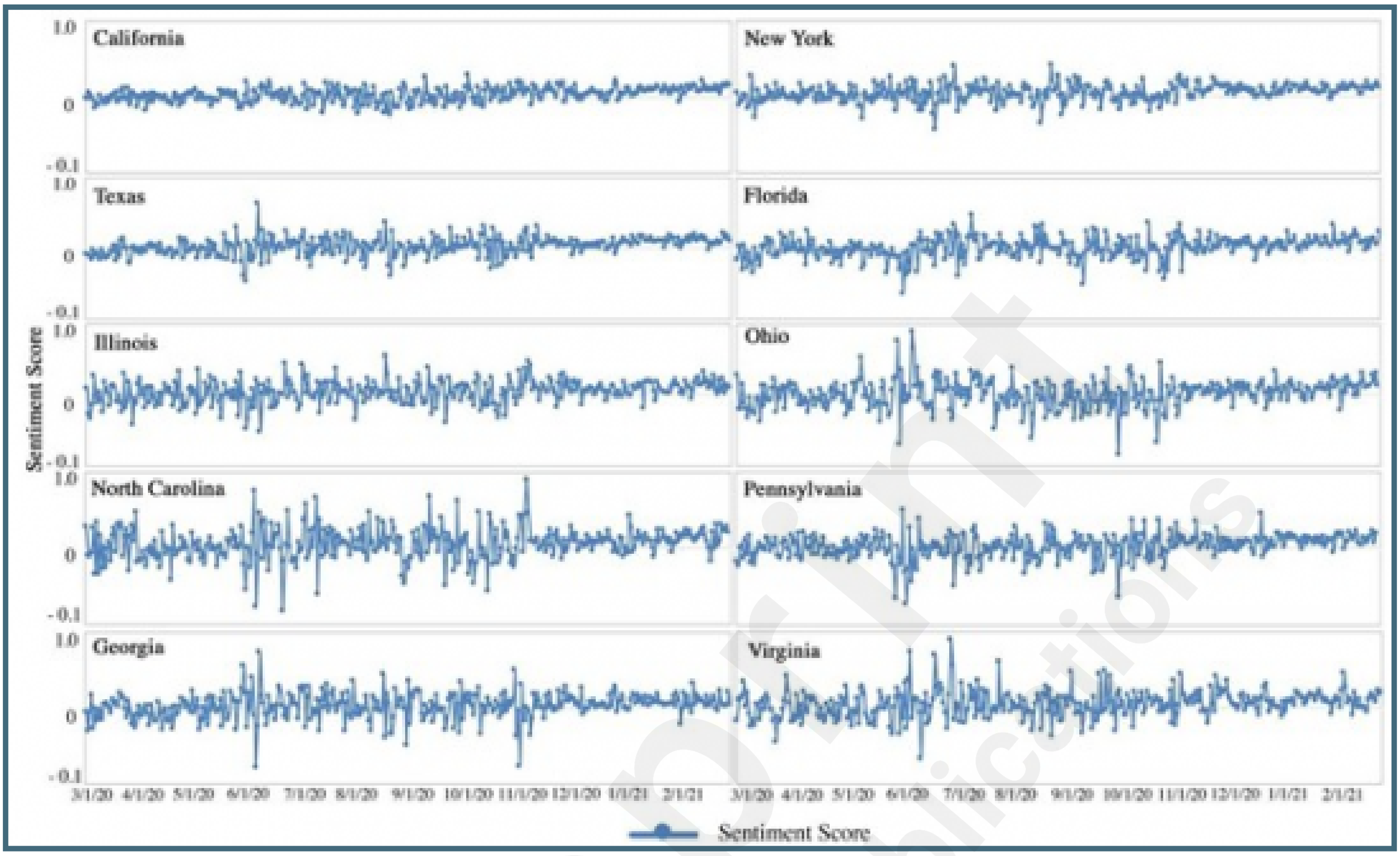
Three topics discussed on key dates: Jul. 22, Aug. 20, Nov. 9, and Nov. 18, 2020.



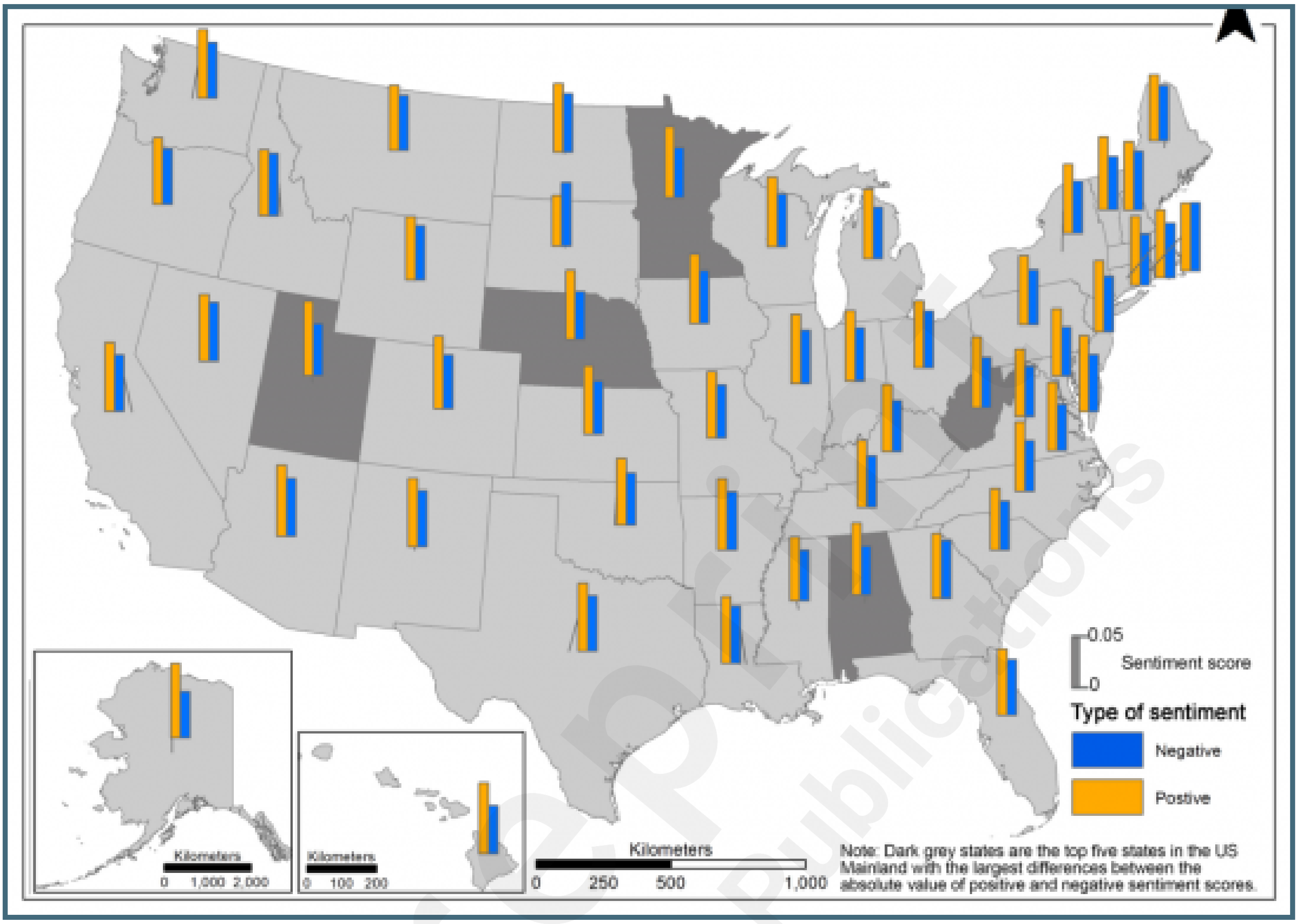
Three topics discussed on key dates at phase 3.



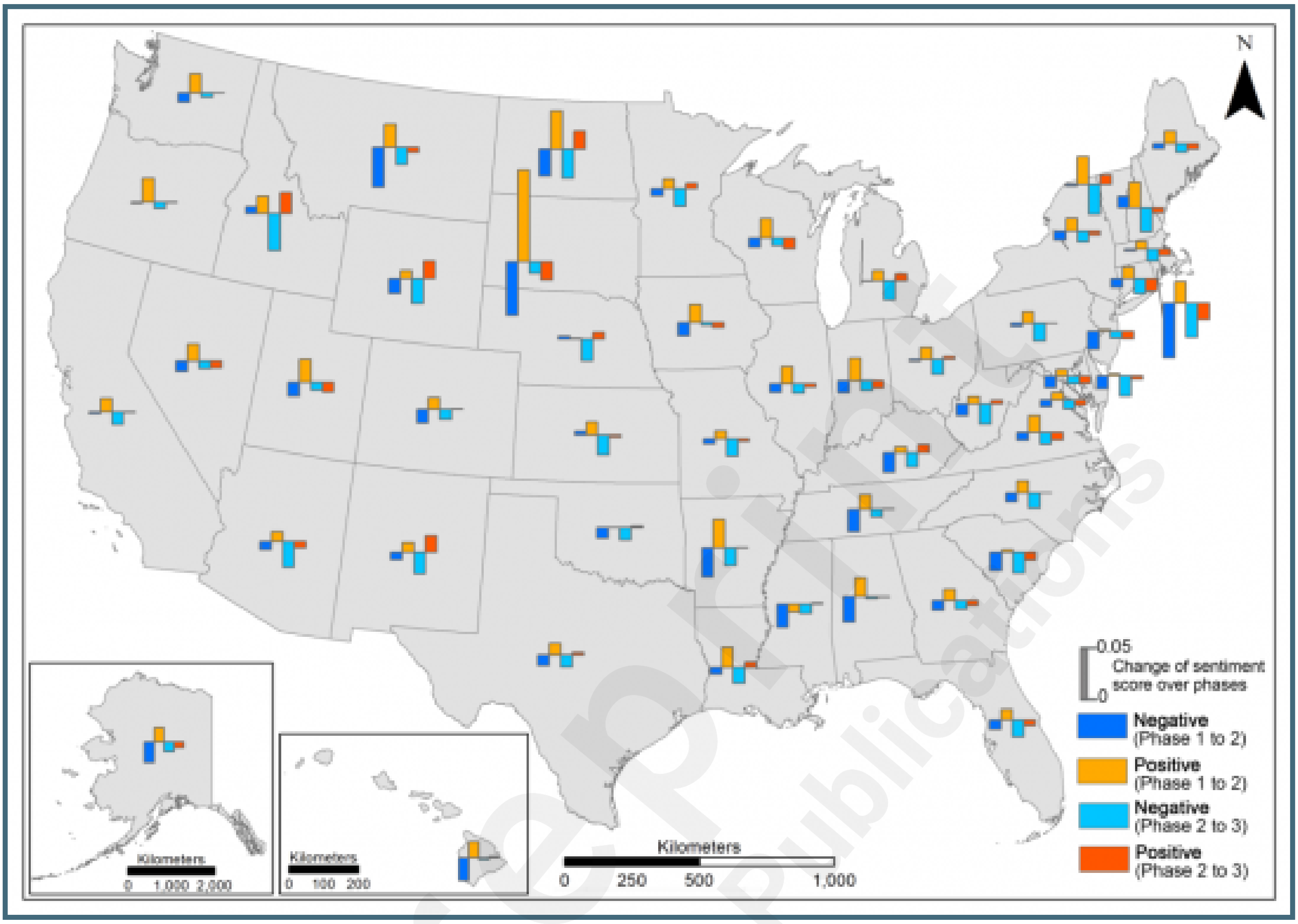
Sentiment scores and the number of geotweet in the selected ten states.



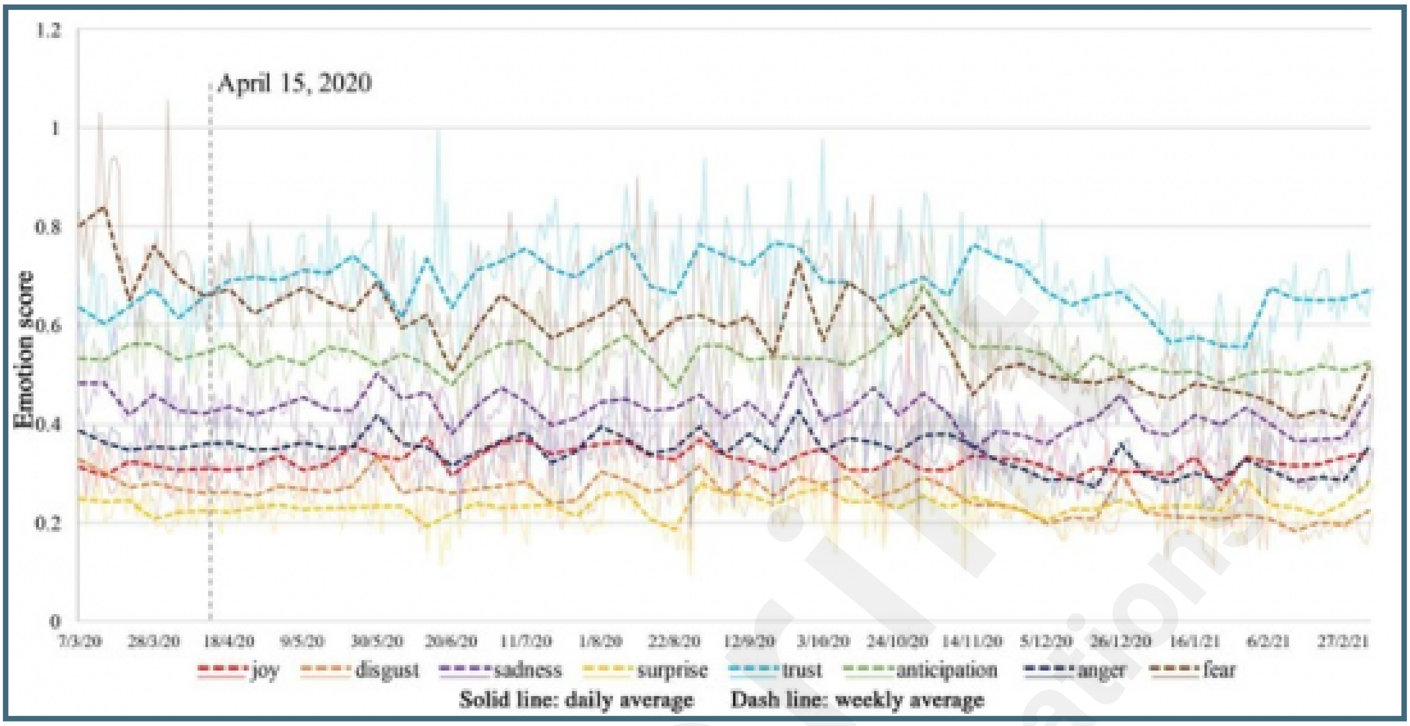
The absolute value of negative and positive sentiment scores at the state level.



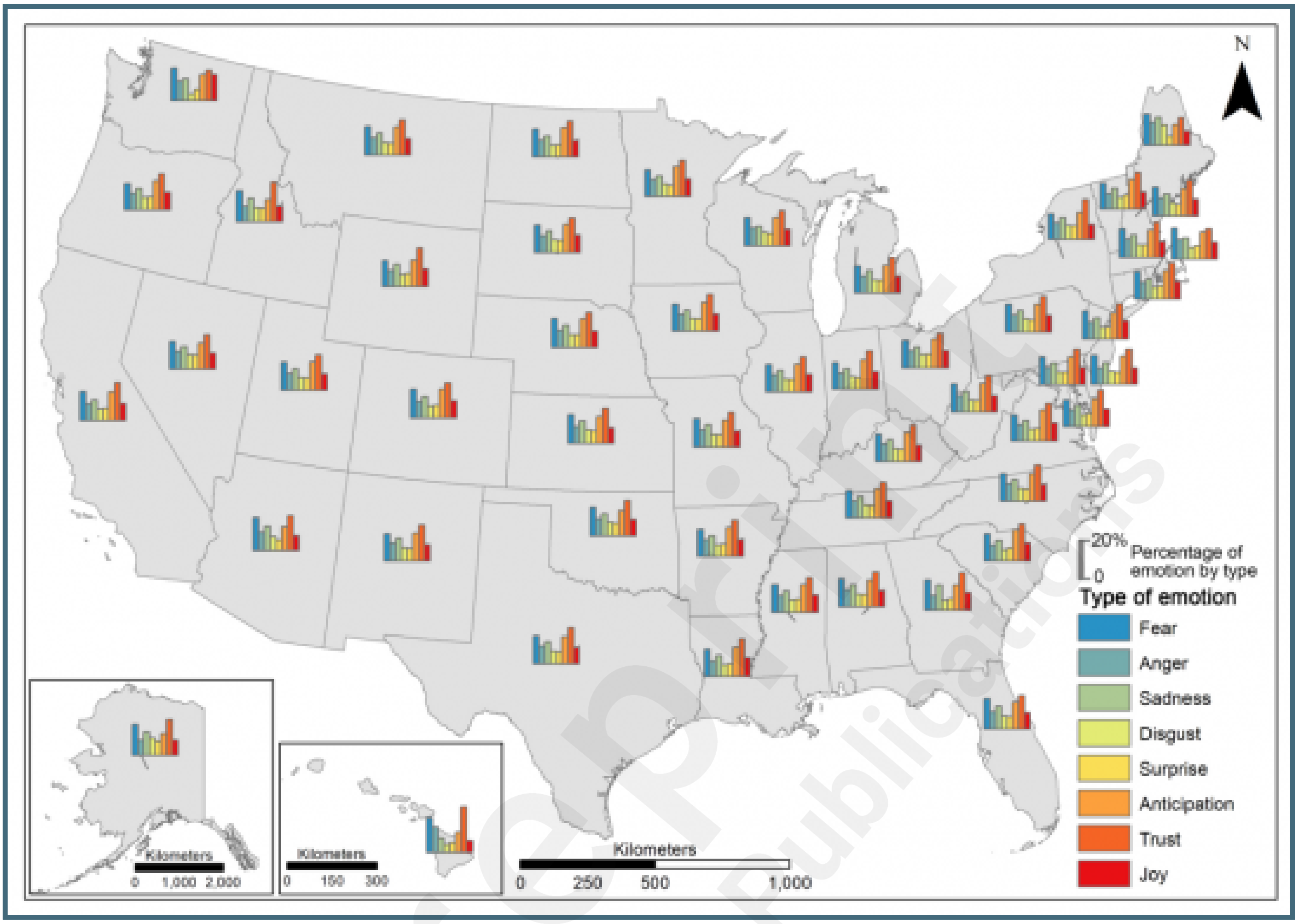
Change of sentiment scores over three phases at the state level.



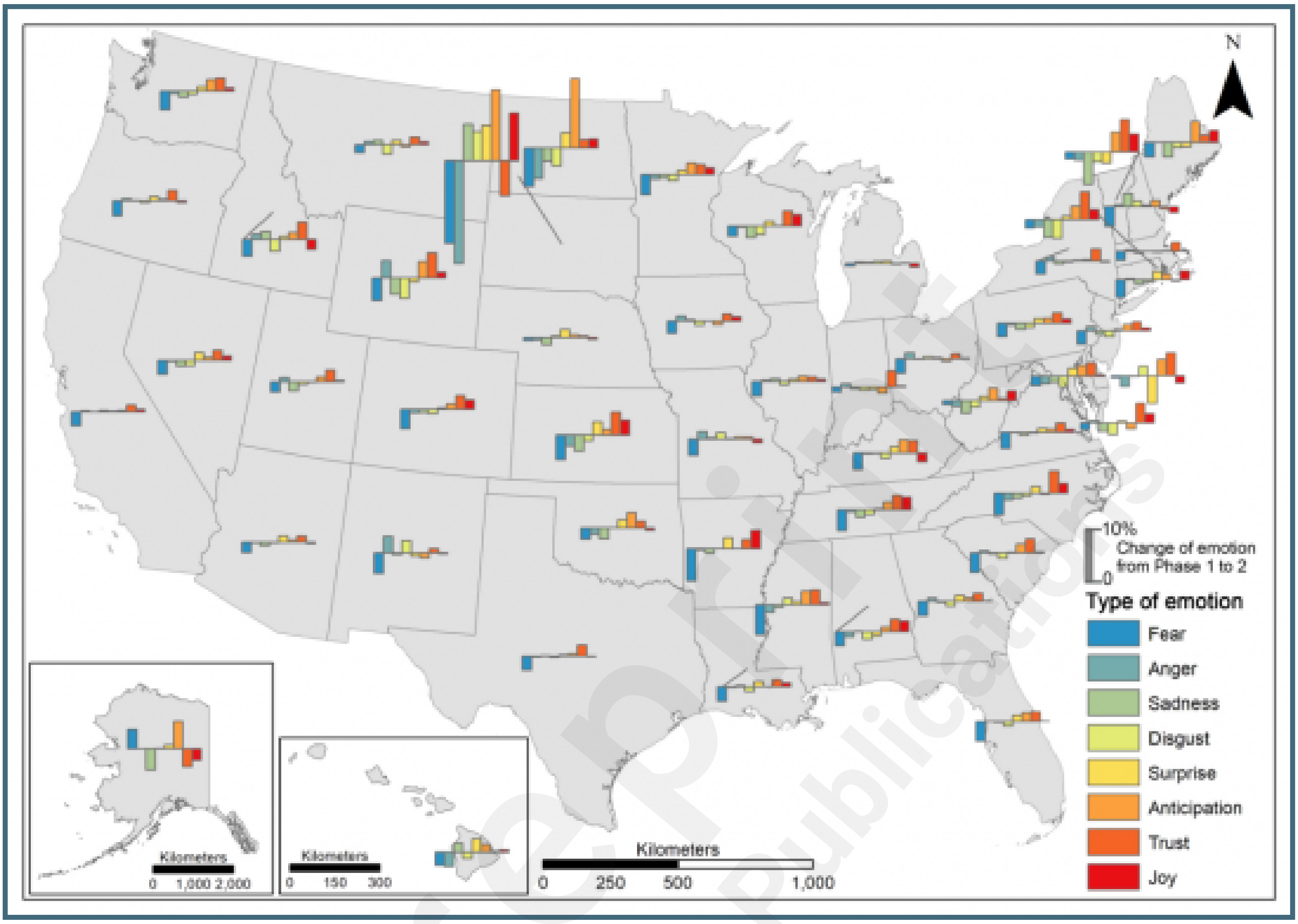
Daily and weekly average of emotion scores over the entire study timeline.



Percentage of eight emotions at the state level.



Change of emotion from Phase 1 to Phase 2 at the state level.



Change of emotion from Phase 2 to Phase 3 at the state level.

Popular keywords associated with positive and negative sentiment over three phases.