

Public sentiment towards COVID-19 vaccines on Twitter

Aminata Bangura, Ehren Dietrick, Huong Giang Ho

CIS428-001

12/2/2025

Fall Semester 2025

Introduction

Understanding how people react during major crises has become increasingly important, especially as social media now serves as a primary space for public conversation. Pandemics and other large-scale natural disasters don't just create physical and economic disruption but also shape emotions, opinions, and behaviors in ways that spread rapidly online. Because platforms like Twitter capture real-time reactions from millions of users, they offer a unique window into how people process uncertainty, evaluate risk, and respond to evolving information.

This project uses text analytics to explore how public sentiment changed throughout different stages of the COVID-19 pandemic. By focusing on global weekly death toll trends, we look at how the severity of the crisis influenced what people talked about and how they expressed it. Topic modeling helps us identify the major themes that dominated conversations in each period, while sentiment analysis allows us to track whether the overall tone became more positive, negative, or neutral as conditions changed.

Our broader goal is to understand not only what people said, but how conversations shifted the way they did. We believe pandemics can trigger fear, speculation, and political debate, and those reactions can spread as quickly as the disease itself. By examining patterns in language, we aim to discover how uncertainty, information availability, and perceived progress shape public response over time. These insights can help us understand reactions to future pandemics or natural disasters, where social media will continue to play a major role in shaping public perception and behavior.

Data Collection

The “Covid-19 Vaccine Tweets” dataset was collected by using Twitter’s API using the hashtag #covidvaccine. This dataset is a collection of 209,930 tweets from February 12, 2020, to October 22, 2020. In total, there are 209,929 observations (rows), with 36 variables (columns), including core tweet data: id, content, username, datetime, and engagement metrics (likes, retweets, replies), and tweet composition: additional hashtags, mentions, URLs, and media (photos/videos), etc (Yadav, 2020).

Divide the dataset based on the trend of global death tolls per week

Purpose: Break down the dataset into more manageable parts with specific context, providing a chance for deeper analysis. The reason is death toll is a concrete metric that drives fear and consequently discussion and action more than the infected toll. The English-speaking population with access to Twitter is concentrated in a few regions, but is still spread out across the globe, so looking at the global death tolls makes sense in this context.

The datasets are: Feb 12 - March 14 (fluctuated under 1000 weekly deaths), March 15 - April 25 (sharp increase in weekly death tolls 2.7k to 50k), April 26 - June 6 (gradual decline from 50k to 33k), June 7 - July 11 (fluctuated around 35k), July 12 - Sep 19 (fluctuated around 40k), Sep 20 - Oct 22 (a sharp decline) (World Health Organization, 2024; Yadav, 2020).

Data transformation

Converted the tweet column into a text corpus using the tm package's Corpus(VectorSource()) function, forming the foundation for all text mining operations

Cleaning and normalizing the text by applying a systematic cleaning sequence to the corpus, converted all text to lowercase for consistency, removed punctuation, numbers, and extraneous whitespace, eliminated common stopwords, converted the text into a Term Document Matrix (TDM), and turned the cleaned corpus into a Term-Document Matrix (TDM)

Results

LDA (Blei et al., 2003) (topics are defined to provide context for sentiment analysis)

Feb 22 - March 14: fluctuated under 1000 weekly deaths (Covid 6)

Topic 1: Personal Initial Acknowledgement of COVID (flu, get, like, just, dont, think, shot, year, don't, people, hope, antivaxxers, going). This topic reflects personal emotion and debate about vaccination. The language is informal (just, like, don't), emotional (hope), and related to getting a shot. The presence of antivaxxers signals a polarized debate.

Topic 2: Vaccine Development & Accessibility (get, flu, one, people, research, amp, first, available, developed, everyone, free, months, year). This topic is focused on the logistics and science of the vaccine itself. Keywords like research, developed, first, available, months, and free likely point to discussions about the speed of research development.

Topic 3: Political and Social Commentary (people, get, now, make, trump, years, see, free, much, know, can, scientists, big, flu, year, virus, like, new, every, even). This topic centers on broader societal and political responses with terms like make, big, scientists, and free suggesting discussions about government mandates, rules, regulations, and economic impacts ("make" money), and "big pharma."

March 15 - April 25: sharp increase in weekly death tolls 2.7k to 50k (Covid 5)

Topic 1: Globalized discussion (amp, flu, can, get, just, people, know, 's, need, believed, virus, make, time, gates, world, india, ingenuity, skin, bill, says). This topic is dominated by the Bill Gates narrative as the cluster of gates, bill, world, india, ingenuity (likely from "Ingenuity Labs" or similar), and skin (possibly "microchip" or "5G" conspiracies) points to viral theories blaming global elites for creating or exploiting the pandemic. Flu persists as a misleading comparison.

Topic 2: Skepticism around COVID (take, gates, one, can, now, test, amp, even, see, year, months, bill, think, time, available, get, working, say, refuse, virus). This topic centers on immediate personal actions (take, test, refuse) and skepticism (think, even). The presence of gates/bill again links refusal to conspiratorial thinking. The test is a focal point, indicating debates over testing availability, accuracy, and compliance.

Topic 3: Vaccine Development (get, people, first, make, 's, want, good, going, 'm, test, dont, trials, say, need, now, new, gates, well, hope, africa). This topic is forward-looking and pragmatic, focused on the pathway to a solution. Key terms trials, test, new, and hope point to vaccine development. However, the high weight of people, first, and africa introduces a strong theme of equity and "vaccine nationalism". There's tension between wanting a good solution and fearing inequitable distribution.

Topic 4: Political Response (get, world, just, trump, going, new, trials, make, testing, 'm, one, bill, developed, wait, research, think, skin, let, american, virus). Terms like trump, world, american, testing, research, and developed conversation around the U.S. national response in a global context. Words like wait, just, and going convey frustration and impatience with

leadership. It's a discourse of comparing national responses and blaming domestic leadership for perceived failures.

Topic 5: Media Critique (like, people, 's, just, flu, one, take, never, can, year, via, think, news, right, working, virus, first, shot, africa, bill). This topic frames the pandemic through media and historical lenses. Like, just, and flu are used to minimize the severity. News and via suggest a focus on media coverage and information sharing.

April 26 - June 6: gradual decline from 50k to 33k (Covid 4)

Topic 1: Personal Choice & Skeptical Hesitancy (get, just, flu, take, can, one, dont, like, well, everyone, amp, 'm, use, know, want, 's, good, bill, gates, may). This topic centers on the first-person perspective of acceptance. The high frequency of get, take, can, don't, and mention intimate personal choices that are COVID-related and possibly vaccine-related. The continued use of flu as a comparator (like) shows minimization persists even as deaths decline. Terms everyone and may indicate social pressure and uncertainty.

Topic 2: Global Discussion (people, can, 's, work, oxford, say, take, china, world, one, flu, get, think, trials, now, trump, moderna, made, bill, new). The focus here is on the actors and geopolitics of vaccine development. Oxford, Moderna, and China name specific players in a global race. World, say, trials, made, and new points to international announcements, trial results, and manufacturing claims. Trump reappears, linking the effort to U.S. politics.

Topic 3: Progresses (bill, now, going, amp, still, trial, via, like, get, research, first, one, shot, need, moderna, test, see, really, gates, trump). This topic is characterized by a sense of process and anticipation. Key temporal words: now, still, first, going. It revolves around tracking the stages (trial, research, test) of frontrunner vaccines (moderna). Words like see, need, and really convey public impatience and scrutiny.

Topic 4: Evaluating Vaccine (trial, first, take, just, flu, like, even, people, available, virus, work, come, dont, know, via, realdonaldtrump, sure, news, effective, 's). This is the discourse of assessment and debate. Central terms trial, first, effective, available, and virus indicate discussions are moving from if a vaccine will come to how good and safe it will be. Realdonaldtrump shows political figures are directly shaping this evaluation. Sure, dont, know, and news reflect public skepticism and the role of media in reporting scientific results.

Topic 5: Scientific Explanation and Theories (gates, get, people, 's, amp, via, think, now, one, technology, experimental, use, new, dna, better, explains, alter, good, first, global). Terms like technology, experimental, dna, alter, and global point to narratives about vaccines or COVID capability for genetic modification (alter dna), transhumanism, or global surveillance/control.

June 7 - July 11: fluctuated around 35k (Covid 3)

Topic 1: Anticipation around Covid (get, can, first, make, flu, world, like, 's, just, human, think, hope, trial, going, icmr, new, now, august, amp, good). This topic expresses a future-

oriented, cautiously optimistic outlook. The key is the focus on human trial progress and a specific timeline (August). Terms like world, first, new, and hope convey a global race for a solution.

Topic 2: Personal Participation (get, amp, going, 's, people, trials, time, take, just, 'm, know, make, company, back, trial, getting, gonna, human, also, read). The high weight of get, going, take, 'm, getting, and gonna suggests discussion about signing up for or participating in clinical trials. Company and human trials are central. It captures the experience and nervous uncertainty of volunteers (back, know, read).

Topic 3: Skeptical of Safety & Process (one, now, people, gates, 's, trials, via, amp, news, right, can, clinical, even, don't, like, first, think, year, safe, testing). The cluster clinical, trials, testing, safe, and year centers on concerns about rushed safety protocols ("should take years"). Gates links this skepticism to conspiracy. News and via highlight the role of media in amplifying these concerns.

Topic 4: Geopolitical Demand (people, need, take, just, india, want, trials, news, get, one, like, gates, come, flu, bill, many, china, new, can, amp). This topic has a strong geopolitical and nationalistic focus, centered on India. India is the anchor, combined with need, want, people, and take. It reflects a domestic demand for vaccines for India's population, likely in response to fears of Western hoarding. Mentions of gates, bill, and china show this demand is framed within a context of distrust towards Western elites and geopolitical rivals.

July 12 - Sep 19: fluctuated around 40k (Covid 2)

Topic 1: Global Vaccine Politics (take, people, can, via, 's, just, amp, one, flu, health, trump, time, big, pharma, first, india, world, news, make, oxford). This topic frames vaccines within a skeptical, global political context. Big pharma is a key target of distrust. India and Oxford (AstraZeneca) indicate specific players, while world and trump place it in a geopolitical arena. The core sentiment is caution (take, can, just) about vaccines driven by corporate (pharma) and political (trump, world) motives, rather than pure health.

Topic 2: Personal Hesitancy & Media-Driven Fear (get, people, don't, 'm, news, dont, just, need, 's, like, amp, going, may, one, russian, think, safe, also, trial, now). This is the voice of personal vaccine refusal, grounded in media skepticism. The high weight of first-person negation (don't, 'm, dont) is striking. News is a direct source of distrust. The new term russian enters here, linked to safety concerns.

Topic 3: The Russian Vaccine & Crisis of Trust (get, trial, russia, one, flu, now, need, know, new, dont, trust, first, trump, china, 'm, trials, shot, just, india). This topic is dominated by Russia's vaccine announcement. Russia and trial are central, sparking debate about speed (first, now) and credibility (trust, dont, know). Trump and China are mentioned, showing it's immediately compared to the U.S.-China rivalry.

Topic 4: Russian-related discussion (russia, amp, can, trials, trump, want, first, just, take, like, human, make, get, says, putin, year, people, said, think, even). This topic focuses on Russia's aggressive claim and the West's response. Russia, Putin, says, said, first paints a picture

of a bold national announcement. Human trials can reflect Western skepticism about the validity of their process. Trump is positioned as a reacting counterpart. It captures the geopolitical drama and one-upmanship of the moment.

Topic 5: U.S. Political Pressure & Regulatory Scrutiny (trump, says, take, first, via, trial, get, news, like, people, cdc, now, see, russia, amp, new, gates, phase, trials, got). This topic shows U.S. political leaders (Trump) and agencies (CDC) responding to the Russian news. Trial, phase, and cdc indicate the scrutiny of the American regulatory process being contrasted with Russia's approach.

Sep 20 - Oct 22: a sharp decline (Covid 1)

Topic 1: Corporate Vaccine Rollout & Public Messaging (amp, people, trial, get, johnson, 's, health, via, first, just, know, trials, like, says, wont, said, can, safe, public, many). This topic is dominated by corporate and official communications. Johnson (Johnson & Johnson) joins the list of vaccine developers. Health, public, says, said indicate statements from companies and health agencies (FDA, CDC). The focus is on reporting trial results, safety (safe) announcements, and public guidance about who will get the vaccine first.

Topic 2: Political and Legal Pressure around Regulatory (trump, people, says, just, get, first, take, flu, fda, can, via, news, safe, life, work, ready, know, won't, health, now). This topic captures the intense political pressure on the vaccine approval process. Trump is directly linked to fda, safe, ready, and first. Terms like says, news, and just frame it as a public, media-driven narrative where the President is urging, promising, or demanding expedited authorization before the election, raising safety concerns among critics. Flu can be seen as downplaying risks in this push.

Topic 3: Election-Timed Skepticism (get, one, flu, make, year, take, going, just, now, see, research, first, like, next, can, trials, back, election, study, everyone). This topic expresses deep personal caution explicitly tied to the political calendar. Election is the critical new term, which can be linked to rushed, safe, and trust concerns. Phrases like make, take, see, next, year, and research suggest a future outlook, indicating politically-motivated discussion.

Topic 4: Logistical Realities & Side Effect Concerns (get, trials, trial, like, via, take, johnson, amp, available, india, may, says, need, news, weeks, effects, can, mandatory, dont, new). This topic focuses on distribution (available, weeks, india), real-world safety (effects), and policy (mandatory). Johnson again appears alongside trials, indicating discussions of specific vaccine properties. May, need, dont reflect public uncertainty about access and requirements.

Topic 5: Crisis of Trust in Institutions (trump, 's, take, amp, people, can, realdonaldtrump, flu, get, don't, trust, news, world, like, getting, available, new, says, make, want). This topic displays the theme of trust in institutions. Trust is the core word, positioned among trump/realdonaldtrump, don't, news, and world.

“Bing” sentiment analysis (Hu & Liu, 2004)

To measure the emotional tone across the six COVID-19 datasets, the Bing sentiment lexicon was used. This lexicon categorizes English words into a binary classification of positive and negative sentiment, allowing for a quantitative analysis of the overall affective direction within the public discourse during each distinct phase of the pandemic.

“Like”, “trump” are stop words being removed due to likely misunderstanding of context. “Like” in these discussions is often used as a word for comparison, not a positive feeling. “Trump” is mostly being mentioned as a person’s last name, and also not a positive feeling. “Virus” is also being removed due to its being mentioned a lot, yet it provides neutral context rather than portraying a negative feeling.

Sentiment analysis per topics

Dataset	Period	Negative Words	Positive Words	Net Sentiment	Total Words	Pos/Neg Ratio	Sentiment Character
Covid 6	Feb 22 - Mar 14 <1k)	16	71	.55	87	.067	lightly positive
Covid 5	Mar 15 - Apr 25 2.7k → 50k)	880	372	.492	2252	.0837	lightly positive
Covid 4	Apr 26 - Jun 6 50k → 33k)	2536	2934	.398	5470	.0317	lightly positive
Covid 3	Jun 7 - Jul 11 ~35k)	2223	2199	.24	4422	.998	Most Neutral
Covid 2	Jul 12 - Sep 19 ~40k)	1107	4008	.2901	45115	.0408	lightly positive
Covid 1	Sep 20 - Oct 22 Sharp Decline)	1610	9973	.1637	1583	.9482	lightly negative

The sentiment results show that although the overall tone of COVID-19 discussions on Twitter remained relatively neutral with only slightly positive or negative tone for the entirety of each period throughout the pandemic, the intensity and meaning of that positivity changed as conditions shifted.

In the early months, the small number of positive and negative words reflected a period of limited information, where people were mostly sharing personal reactions and comparisons to seasonal flu.

As deaths surged sharply from mid-March to late April, both negative and positive word counts increased dramatically, driven by heated debates around conspiracies, political blame, and testing controversies. Even though the net sentiment remained technically positive, this positivity was largely tied to hopeful or speculative language (“want,” “hope,” “need”) occurring alongside

a spike in criticism and distrust. The slight positive skew in this period should therefore be understood not as optimism, but as intense emotional engagement.

The Apr 26 - Jun 6 period presents a critical divergence: sentiment remained slightly positive (1.032) even as deaths fell sharply from their peak. This period's positivity seems to be fueled by intense, forward-looking debate. The discourse pivoted to vaccine development timelines, geopolitical rivalry between developers, and the emergence of sophisticated conspiracies. The positive language here was charged with anticipation, competition, and speculative hope for a scientific solution, demonstrating that sentiment was driven more by the narrative of a future exit strategy than by the current reality.

The analysis for Jun 7 - Jul 11 shows sentiment reached its most neutral point (0.998) of the entire dataset. This dip in positive language reflects a shift in discourse from the initial surge's emotional urgency to a more pragmatic, weary phase. Conversations centered on the logistics of human trials, vaccine nationalism, and volunteer participation, topics that generate less polarizing, advocacy-driven language and more procedural or uncertain discussion. This neutral sentiment captures a public catching its breath, focusing on complex realities rather than clear-cut hope or blame.

As the pandemic moved into a plateau and a gradual decline phase, sentiment trends aligned with major topic shifts observed in the LDA modeling. Conversations became more technical and progress-oriented, which kept only sentiment slightly positive, though increasingly neutralized by persistent safety concerns. The period from July to September stands out clearly: despite high death rates, this window produced a relatively positive sentiment ratio. This aligns with a major shift toward solution-focused discussions, with global vaccine announcements creating genuine optimism that overshadowed ongoing skepticism.

However, once deaths dropped sharply in late September, sentiment fell again. This drop corresponds to the re-emergence of divisive topics such as antivaxxer frustration, political interference, and distribution inequity. In other words, public discourse became less about the virus itself and more about social, political, and logistical conflicts, which reintroduced negativity despite improving real-world conditions.

Sentiment term analysis per topics

The top sentiment words reveal a clear evolution in the public's focus and framing of the pandemic crisis. In the initial phase (Feb-Mar), the negative lexicon was observational and concerned, featuring words like "strain," "outbreak," and "worry," while the positive words were logistical, dominated by "free," "available," and "ready." This reflects a society assessing a new threat and hoping for swift solutions.

As the crisis escalated (Mar-Apr), the negative tone became dramatically more confrontational and dire, with "refuse," "shit," "die," and "conspiracy" surging to the top. This shift captures the eruption of public backlash, fatalism, and misinformation. Concurrently, positive words like "work," "good," and the notably specific "ingenuity" illustrate a narrative of effort and hope, with scientific promise.

A pivotal shift occurs in the middle phases (Apr-Sep), where the negative sentiment contains sustained refusal as "refuse" remains the top negative word for three consecutive periods, and perceived danger ("risk," "untested"). This indicates that the core public debate had solidified into a conflict over compliance and safety. The positive narrative, however, evolves in meaning. The consistent dominance of "available" and "good" transitions from hope to expectation, while the introduction and rise of "safe" and "trust" in the later periods (Jul-Oct) become the central, charged battleground. These are no longer simple positives; they represent the key demands and points of contention in the vaccine rollout era, explicitly highlighting the public's paramount concerns about the solution itself.

By the final phase (Sep-Oct), negative words become more clinical and severe ("illness," "dies," "emergency," "unexplained"), moving from emotions to realistic outcomes and uncertainty. The positive frame is almost exclusively dominated by the rollout's prerequisites: "available," "safe," "trust," "ready," and "work." In the end, the word choices indicate that science succeeded in creating an answer, but society still faces the challenge of grappling with the reality of lives lost.

“Nrc” sentiment analysis (Mohammad, 2022)

To analyze emotional trends across the six Covid19 datasets, the NRC emotion lexicon was used. “The NRC Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive)” (Mohammad, 2022). This lexicon grouped words into ten emotion categories – fear, trust, anticipation, sadness, anger, surprise, joy, and disgust– along with positive and negative sentiment. Based on linking the words in each dataset to these categories, the emotional tone present during each stage of the pandemic was measured.

Covid 6: February 22-March 14

In this early dataset, emotions such as fear and anticipation were the most prominent. This aligns with the beginning of the global outbreak, when limited information and uncertainty influenced public reactions. Key findings: Since it was early 2020, when the virus broke out, people were worried, confused, and scared. Fear and trust are the top emotions hence the uncertainty of the virus and waiting on more information about it

Covid 5: March 15- April 25

As the timeline progresses to March and April, when lockdowns were happening worldwide, there is a huge rise in negative sentiment, including fear, sadness, and anger. This reflects widespread distress, frustration, and anxiety as daily life changed and case numbers increased.

Key findings: global lockdowns were happening and cases were rising, emotions were increasing, negative words increased from 425 to 7,703, fear increased from 306 to 4,956, and sadness and fear increased.

Covid 4: April 26 - June 6, Covid 3 June 7 - July 11, Covid 2 July 12 - September 19

From late April through the months of Summer, the emotional patterns start shifting. Although negative emotions continue to appear, there is a clear rise in positive emotions such as trust, anticipation, and joy. This aligns with the start of the vaccine trial announcements and the reopening of certain regions, which influenced feelings of hope. By mid-summer, emotional expression is at its highest level, indicating heightened public attention and extensive conversation about safety measures, increasing case numbers, and the politically driven discussions surrounding the pandemic. Key findings from April 26-June 6 include lockdown continuing and deaths increasing, people getting angry with the government, having fear of the virus spreading, and having anxiety about their safety, and the sentiments becoming higher. Key findings from June 7-July 11 include emotions beginning to soften after hearing news about vaccine trials and reopenings, plus the summer weather, joy rises from 3,142 (March 15) to 6,230 (June 7), and positive emotions increased compared to the earlier months. The key findings of July 12-September 19 are high emotional volume and a huge spike in discussion due to debates around safety restrictions, announcements about vaccine trials, and online misinformation.

Covid 1 September 20 - October 22

In this latter dataset, emotional tones shift once again as vaccine-related news becomes a major topic. Emotions such as trust, anticipation, and joy grow noticeably, indicating rising optimism as vaccine announcements and trial results begin circulating. At the same time, the results show notable amounts of fear, anger, and disgust, implying that political debate and public frustration with vaccine distributions and government responses were still playing a major role during this period. This phase reveals a complex emotional mix: rising optimism by scientific progress, alongside persistent tension and skepticism shaped by political narratives and ongoing uncertainty. Key findings: vaccine related political discussions, hope and anticipation for vaccines, and positive and trust emotions increasing again because of vaccine announcements.

The results of this nrc sentiment shows emotional progression that mirrors the unfolding of the Covid-19 pandemic. Across all datasets, there is a noticeable emotional pattern: early confusion and fear, then intense negative emotions during lockdowns, then a rise in hope accompanied by political controversy as vaccines become a main topic. This analysis provides valuable insights for improving public health strategies, crisis communication, and understanding societal reactions to global emergencies.

Discussion

Overall, the findings show that changes in global COVID-19 death tolls strongly shaped how people talked about the pandemic online. When deaths were low, conversations were casual and focused on personal reactions. As deaths surged, discourse became more polarized, with spikes in conspiracy theories, political blame, and mistrust. Once the situation stabilized and vaccine trials progressed, discussion shifted toward scientific updates and practical concerns, and public sentiment temporarily improved. The most positive period occurred not when deaths declined, but when people felt real progress was being made on vaccines. However, as the crisis eased, sentiment became more divided again, driven by frustration over antivaxxers, distribution concerns, and institutional failures. Together, these patterns suggest that public reactions are influenced less by the objective severity of a crisis and more by uncertainty, trust, and perceived momentum toward solutions.

Future research could explore how different communities respond to pandemics. Since our study looked at aggregated global Twitter data, it captures broad patterns but not local differences. Breaking the data down by region, political identity, or cultural context could reveal how misinformation spreads differently across networks, or how trust in institutions varies across countries. Johns Hopkins reveals some data about how each country is affected by COVID 19 in an article, and it would be interesting to compare the mortality per region and analyze the data as a result (Johns Hopkins University, 2023). Another promising direction would be comparing COVID-19 sentiment patterns to reactions during other natural disasters, such as hurricanes or wildfires, to see whether similar cycles emerge, especially due to stress or other factors these disasters may cause (Keller, 2012). This would help clarify whether the emotional stages observed here are unique to pandemics or part of a broader human response to large-scale crises.

Incorporating more advanced sentiment models, such as transformer-based emotion classifiers, could capture nuance beyond simple positive or negative tone. Adding bot-detection methods would help separate genuine public reaction from automated or coordinated misinformation campaigns. More specific time slicing could reveal faster shifts tied to specific announcements or events. Finally, integrating a qualitative reading of tweets would provide human interpretation to complement the quantitative patterns. These changes would allow for a fuller, more precise understanding of how people process and communicate during global crises.

References

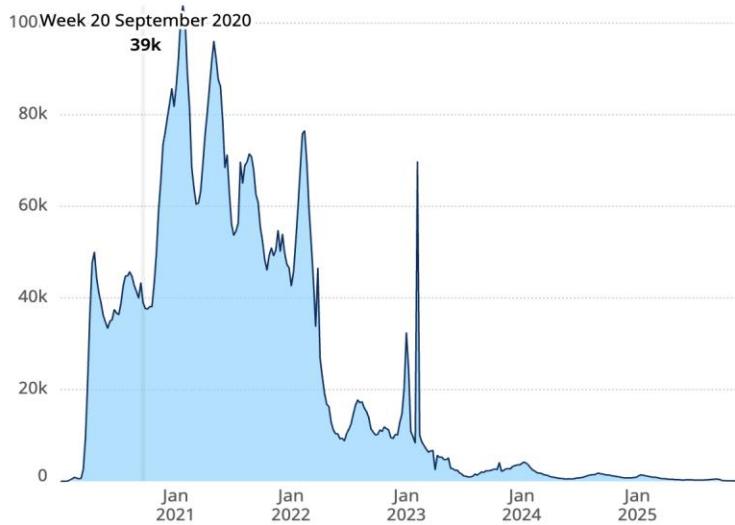
- Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(993-1022), 993–1022.
<https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 168–177. <https://doi.org/10.1145/1014052.1014073>
- John Hopkins University. (2023, March 10). *Mortality Analyses*. Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/data/mortality>
- Keller, A., Litzelman, K., Wisk, L. E., Maddox, T., Cheng, E. R., Creswell, P. D., & Witt, W. P. (2012). Does the perception that stress affects health matter? The association with health and mortality. *Health Psychology*, 31(5), 677–684. APA PsycNet.
<https://doi.org/10.1037/a0026743>
- Levin, A. (2008). Death, Injury Tallies Mask Toll on Families' Mental Health. *Psychiatric News*, 43(21), 17–17. <https://doi.org/10.1176/pn.43.21.0017a>
- Mohammad, S. (2011). *Sentiment and Emotion Lexicons*. Saifmohammad.com.
<http://saifmohammad.com/WebPages/lexicons.html>
- World Health Organization. (2024). *COVID-19 deaths | WHO COVID-19 dashboard*. World Health Organization Data. <https://data.who.int/dashboards/covid19/deaths?n=o>
- Yadav, R. (2020). *COVID-19 Vaccine Tweets*. Kaggle.com.
<https://www.kaggle.com/datasets/ritesh2000/covid19-vaccine-tweets/data>

Appendix

Weekly death tolls

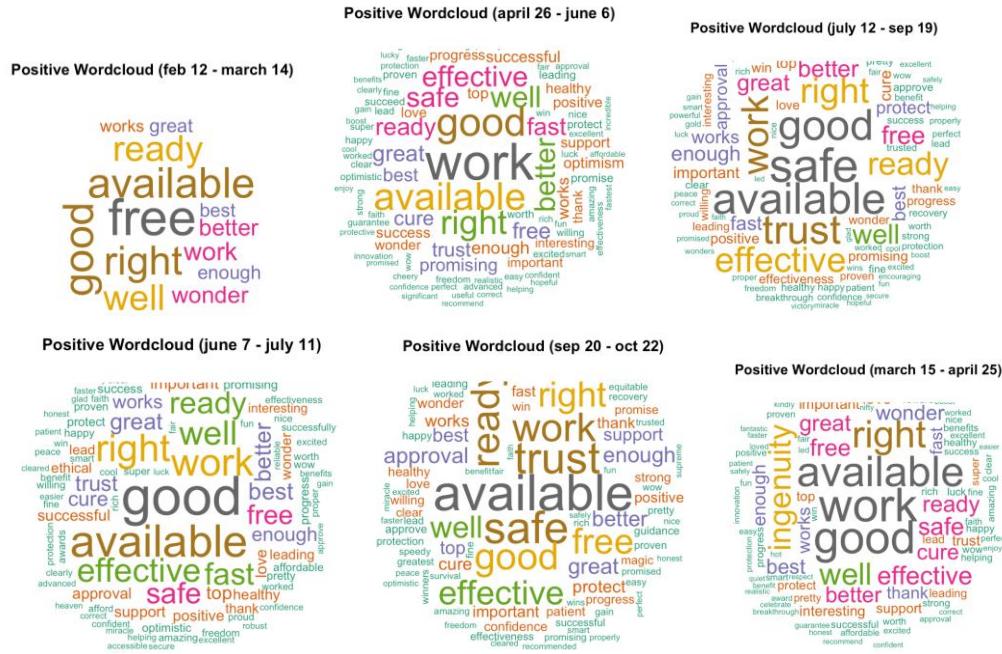
Total COVID-19 deaths reported to WHO (weekly)

World, January 2020 - present



Source: World Health Organization

Sentiment wordclouds (bing)



Negative Wordcloud (july 12 - sep 19)



Negative Wordcloud (march 15 - april 25)



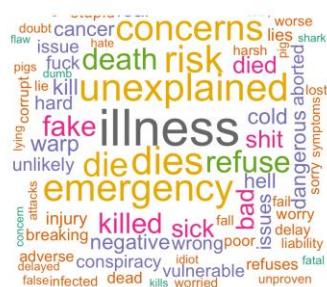
Negative Wordcloud (april 26 - june 6)



Negative Wordcloud (feb 12 - march 14)



Negative Wordcloud (sep 20 - oct 22)



Negative Wordcloud (june 7 - july 11)



Sentiment top words (bing)

word	sentiment	n Dataset	word	sentiment	n Dataset
illness	negative	684 (Sep 20 - Oct 22)	refuse	negative	1231 (July 12 - Sep)
dies	negative	512 (Sep 20 - Oct 22)	risk	negative	1056 (July 12 - Sep)
unexplained	negative	499 (Sep 20 - Oct 22)	die	negative	865 (July 12 - Sep)
emergency	negative	493 (Sep 20 - Oct 22)	bad	negative	816 (July 12 - Sep)
risk	negative	486 (Sep 20 - Oct 22)	shit	negative	800 (July 12 - Sep)
available	positive	1323 (Sep 20 - Oct 22)	safe	positive	3271 (July 12 - Sep)
safe	positive	1153 (Sep 20 - Oct 22)	available	positive	2997 (July 12 - Sep)
trust	positive	1127 (Sep 20 - Oct 22)	good	positive	2922 (July 12 - Sep)
work	positive	1000 (Sep 20 - Oct 22)	trust	positive	2709 (July 12 - Sep)
ready	positive	998 (Sep 20 - Oct 22)	work	positive	2506 (July 12 - Sep)
word	sentiment	n Dataset	word	sentiment	n Dataset
refuse	negative	212 (Jun 7 - Jul 11)	refuse	negative	270 (Apr 26 - Jun 6)
shit	negative	189 (Jun 7 - Jul 11)	die	negative	226 (Apr 26 - Jun 6)
risk	negative	176 (Jun 7 - Jul 11)	risk	negative	217 (Apr 26 - Jun 6)
death	negative	165 (Jun 7 - Jul 11)	doubt	negative	207 (Apr 26 - Jun 6)
untested	negative	165 (Jun 7 - Jul 11)	conspiracy	negative	154 (Apr 26 - Jun 6)
good	positive	607 (Jun 7 - Jul 11)	work	positive	672 (Apr 26 - Jun 6)
available	positive	489 (Jun 7 - Jul 11)	good	positive	508 (Apr 26 - Jun 6)
work	positive	416 (Jun 7 - Jul 11)	available	positive	480 (Apr 26 - Jun 6)
right	positive	404 (Jun 7 - Jul 11)	right	positive	415 (Apr 26 - Jun 6)
well	positive	346 (Jun 7 - Jul 11)	well	positive	349 (Apr 26 - Jun 6)
word	sentiment	n Dataset	word	sentiment	n Dataset
refuse	negative	134 (Mar 15 - Apr 25)	strain	negative	10 (Feb 12 - Mar 14)
shit	negative	107 (Mar 15 - Apr 25)	bad	negative	9 (Feb 12 - Mar 14)
die	negative	96 (Mar 15 - Apr 25)	outbreak	negative	7 (Feb 12 - Mar 14)
conspiracy	negative	78 (Mar 15 - Apr 25)	shit	negative	5 (Feb 12 - Mar 14)
bad	negative	68 (Mar 15 - Apr 25)	worry	negative	5 (Feb 12 - Mar 14)
work	positive	290 (Mar 15 - Apr 25)	free	positive	21 (Feb 12 - Mar 14)
good	positive	285 (Mar 15 - Apr 25)	available	positive	17 (Feb 12 - Mar 14)
available	positive	259 (Mar 15 - Apr 25)	good	positive	17 (Feb 12 - Mar 14)
right	positive	228 (Mar 15 - Apr 25)	right	positive	16 (Feb 12 - Mar 14)
ingenuity	positive	183 (Mar 15 - Apr 25)	ready	positive	14 (Feb 12 - Mar 14)

Sentiment top words (nrc)

February 22- March 14

```
> covid6_nrc
      sentiment   n   dataset
1       positive 941 Dataset 6
2       negative 425 Dataset 6
3           fear 306 Dataset 6
4          trust 296 Dataset 6
5 anticipation 256 Dataset 6
6       sadness 168 Dataset 6
7       anger 160 Dataset 6
8     surprise 154 Dataset 6
9         joy 140 Dataset 6
10      disgust 107 Dataset 6
```

March 15-April 25

```
> covid5_nrc
      sentiment   n   dataset
1       positive 19637 Dataset 5
2       negative 7703 Dataset 5
3          trust 6593 Dataset 5
4 anticipation 5098 Dataset 5
5           fear 4956 Dataset 5
6           joy 3142 Dataset 5
7       sadness 2959 Dataset 5
8       anger 2661 Dataset 5
9     surprise 2479 Dataset 5
10      disgust 1974 Dataset 5
```

April 26-June 6

```
> covid4_nrc
      sentiment   n   dataset
1       positive 41311 Dataset 4
2       negative 15391 Dataset 4
3          trust 13528 Dataset 4
4 anticipation 10833 Dataset 4
5           fear 10546 Dataset 4
6           joy 6521 Dataset 4
7       sadness 6514 Dataset 4
8       anger 5639 Dataset 4
9     surprise 5221 Dataset 4
10      disgust 3672 Dataset 4
```

June 7-July 11

```
> covid3_nrc
      sentiment     n   dataset
1      positive 41013 Dataset 3
2      negative 14355 Dataset 3
3          trust 13098 Dataset 3
4 anticipation 11072 Dataset 3
5          fear  9390 Dataset 3
6          joy   6230 Dataset 3
7      sadness  5815 Dataset 3
8      anger   5383 Dataset 3
9    surprise  4785 Dataset 3
10     disgust  3543 Dataset 3
```

July 12- September 19

```
> covid2_nrc
      sentiment     n   dataset
1      positive 239537 Dataset 2
2      negative  83747 Dataset 2
3          trust  79148 Dataset 2
4 anticipation  59622 Dataset 2
5          fear  57283 Dataset 2
6          joy   35970 Dataset 2
7      sadness  35079 Dataset 2
8    surprise  33812 Dataset 2
9      anger   31267 Dataset 2
10     disgust  20834 Dataset 2
```

September 20-October 22

```
> covid1_nrc
      sentiment      n   dataset
1      positive 99886 Dataset 1
2      negative 38135 Dataset 1
3       trust 33347 Dataset 1
4 anticipation 26943 Dataset 1
5       fear 26329 Dataset 1
6     sadness 17246 Dataset 1
7      anger 14754 Dataset 1
8    surprise 14629 Dataset 1
9       joy 14138 Dataset 1
10     disgust  8785 Dataset 1
```