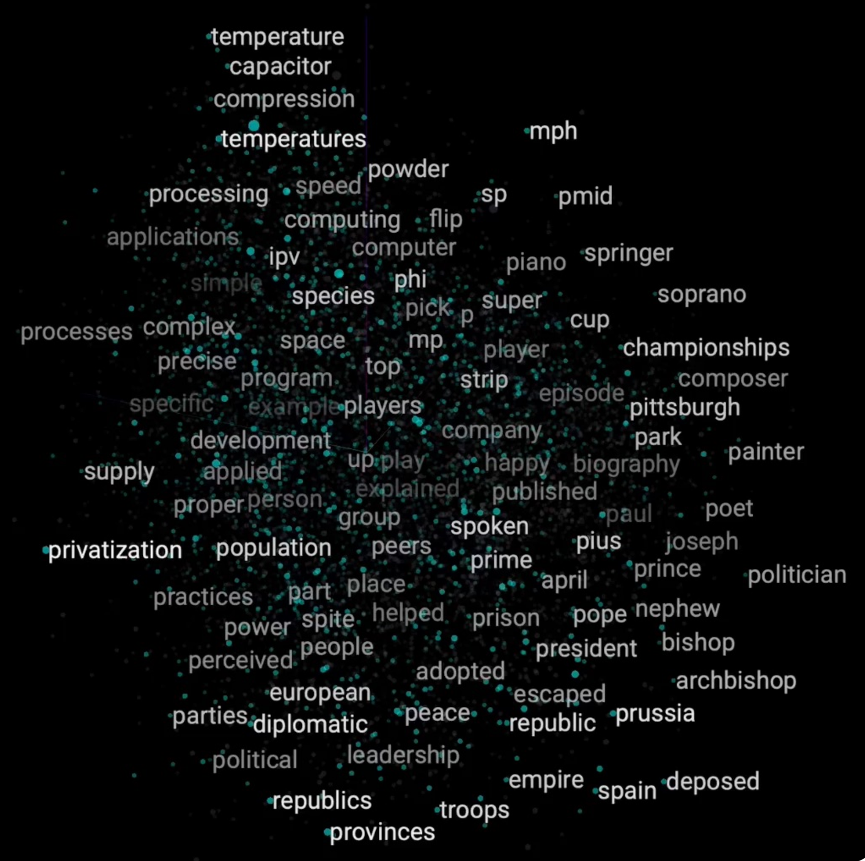
Tracking the Dynamics of Vaccination Sentiment in Large-Scale Social Media Data



Qingqing Chen, Andrew Crooks
University at Buffalo - SUNY

Feb 27, 2022

AAG American Association of Geographers
University at Buffalo The State University of New York

Background

- Diseases outbreak (e.g., HIN1 influenza, the Ebola and Zika viruses) → Have brought tremendous economic losses and deaths
- The COVID-19 virus has caused unprecedented impacts on the life trajectories of millions of people globally (environmental, psychological, social, economic)

New York City, US

Source: https://www.loveexploring.com/galleryextended/95636/the-worlds-cities-before-during-and-after-lockdown?page=1

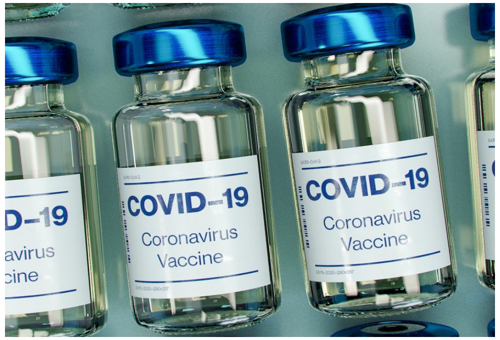


Photo by Daniel Schludi

Lujiazui Traffic Circle in Shanghai's Pudong District

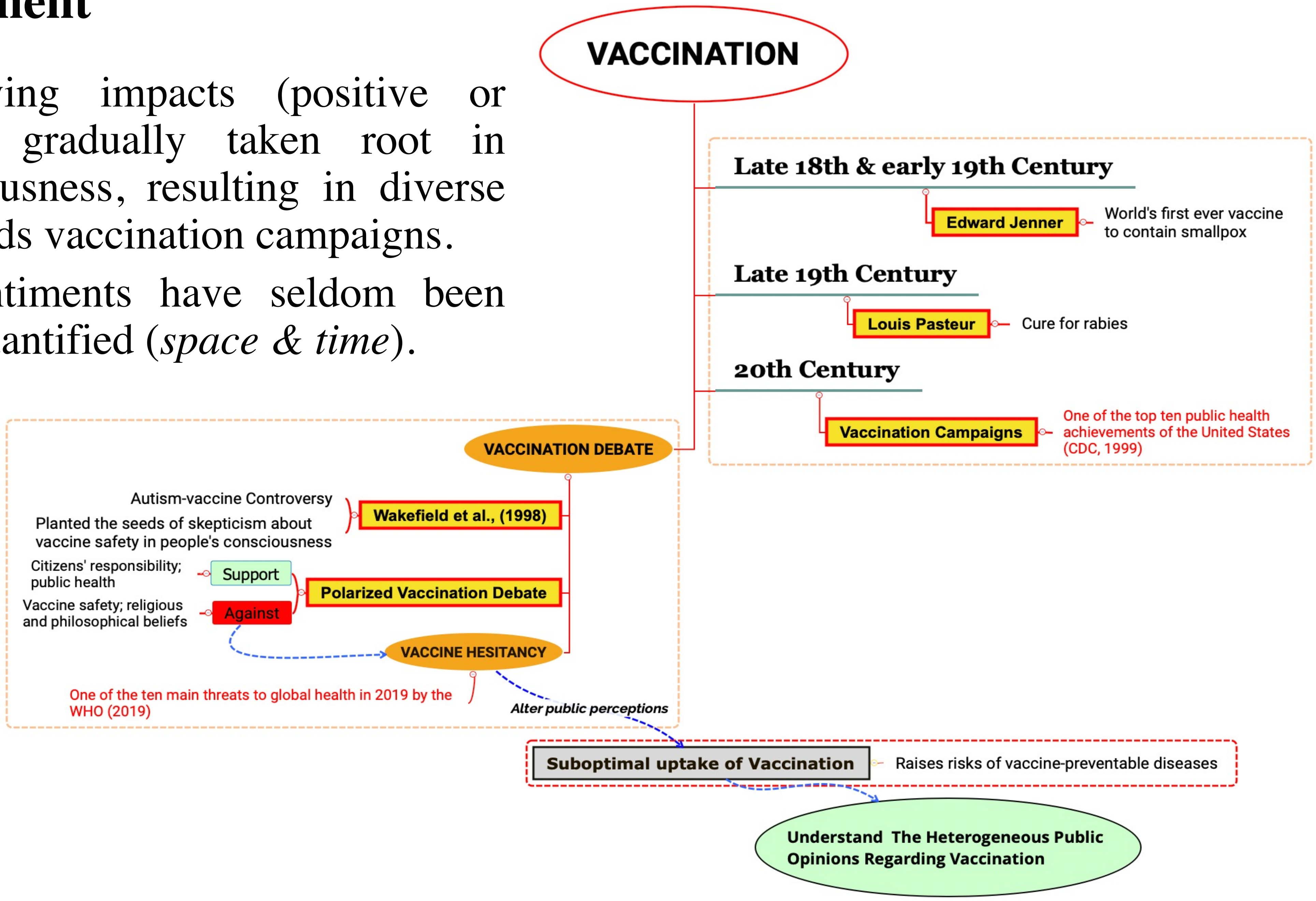
Source::
https://www.abc.net.au/news/2020-02-15/coronavirus-lockdown-in-china-millions-put-into-quarantine/11968114
https://pudong.ca/attractions/lujiazuitrafficcircle.html

Paris, France

Source: https://www.loveexploring.com/galleryextended/95636/the-worlds-cities-before-during-and-after-lockdown?page=1

Problem Statement

- The accompanying impacts (positive or negative) have gradually taken root in people’s consciousness, resulting in diverse sentiments towards vaccination campaigns.
- The diverse sentiments have seldom been systematically quantified (*space & time*).



Current State of the Art

Sentiment analysis:

- Machine Learning techniques (*Piedrahita-Valdés et al., 2021; Zhang et al., 2021*)
- Valence Aware Dictionary for Sentiment Reasoning (VADER) (*Hu et al., 2021; Karami et al., 2021*)
- NRC Word-Emotion Association Lexicon (*Dubey et al., 2020, Das et al., 2020*)

Topic modeling:

- Latent Dirichlet Allocation (LDA) (*Ridhwan et al., 2021; Jang et al., 2021*)

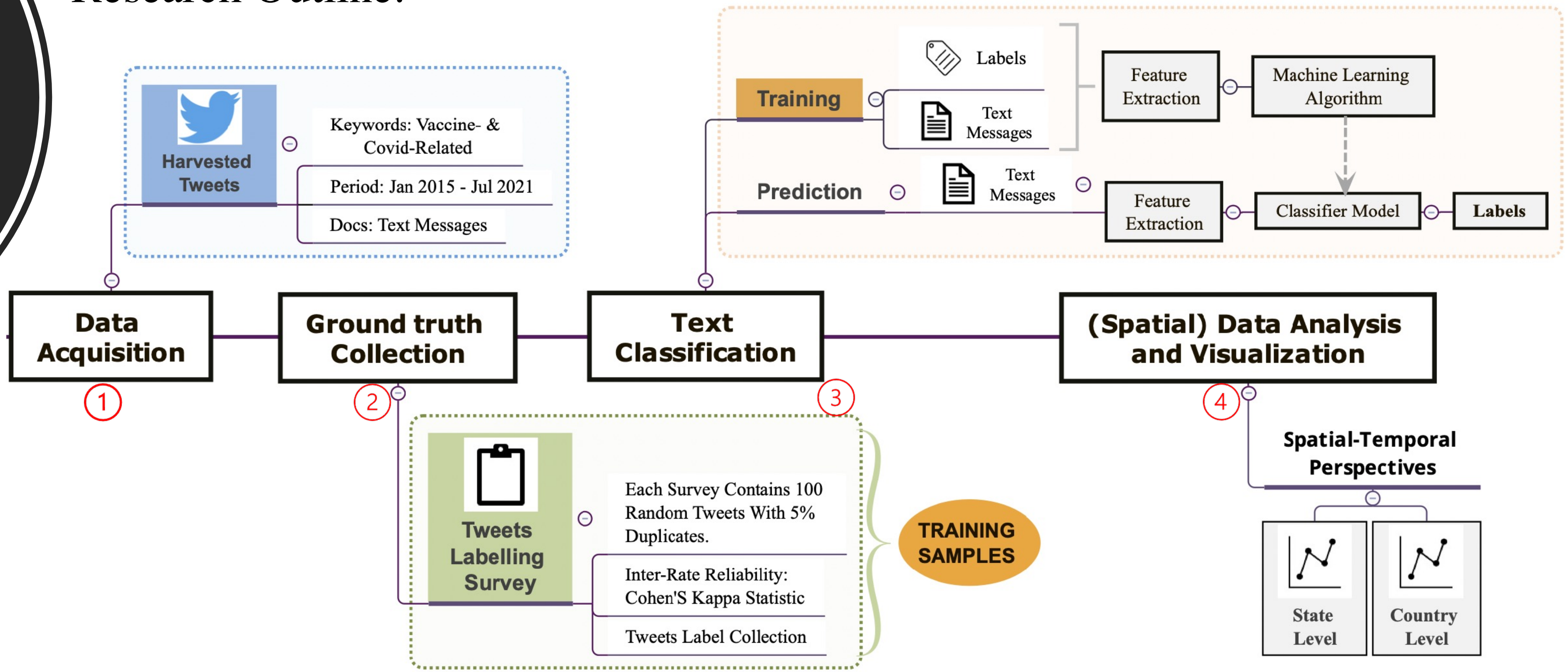
However,

- *Simple random sample/model*
- *Limited/short study period*
- *Lack of the combination of spatial & temporal perspectives*
- *No one has compared the dynamic changes of vaccine sentiment before and after a disease outbreak*

Objectives

Analyze vaccine sentiments in *SPACE* and *TIME* by integrating Word Embedding and Machine Learning techniques based on Social Media Data.

Research Outline:

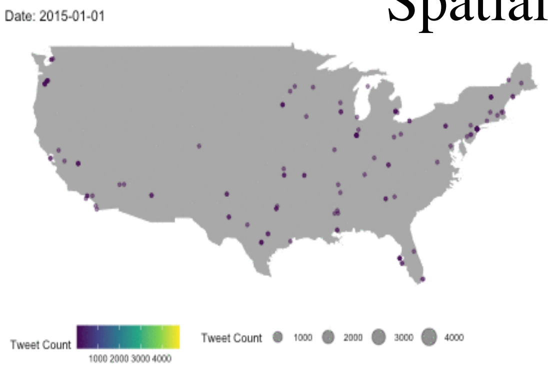


Twitter Data

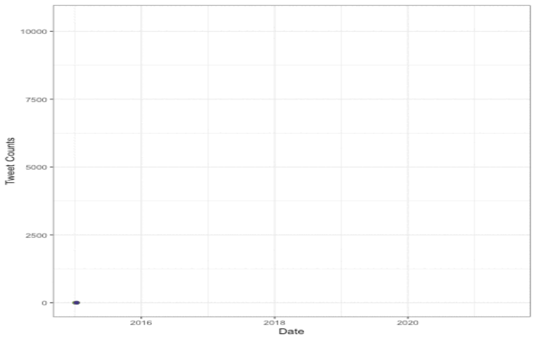
- Keywords: Vaccination and their derivatives (e.g., “vaccine”, “vaxx”, “vaxine”)
- Jan 2015 – July 2021 (~ 12 million tweets sent by 2 million users)

Data Acquisition

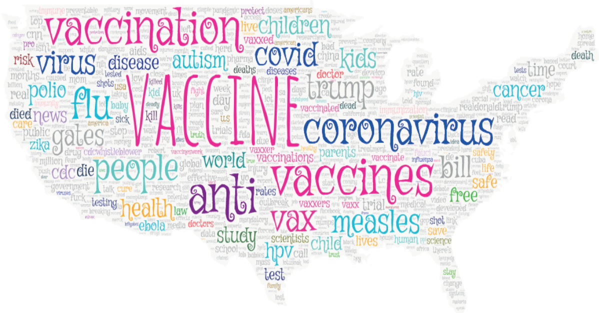
Spatial



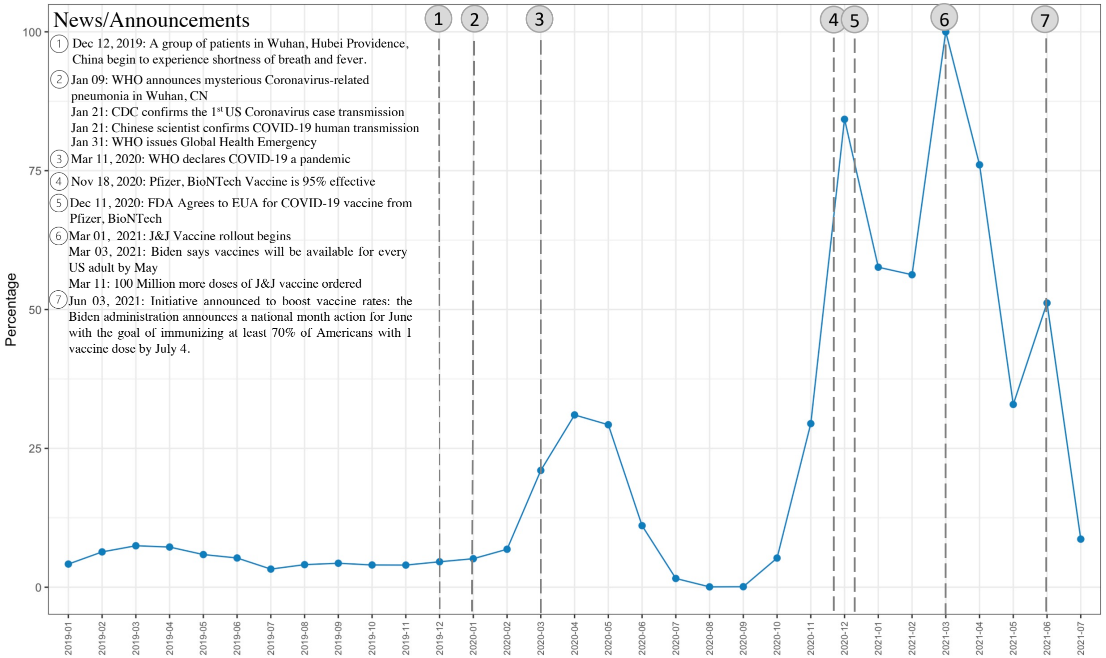
Temporal

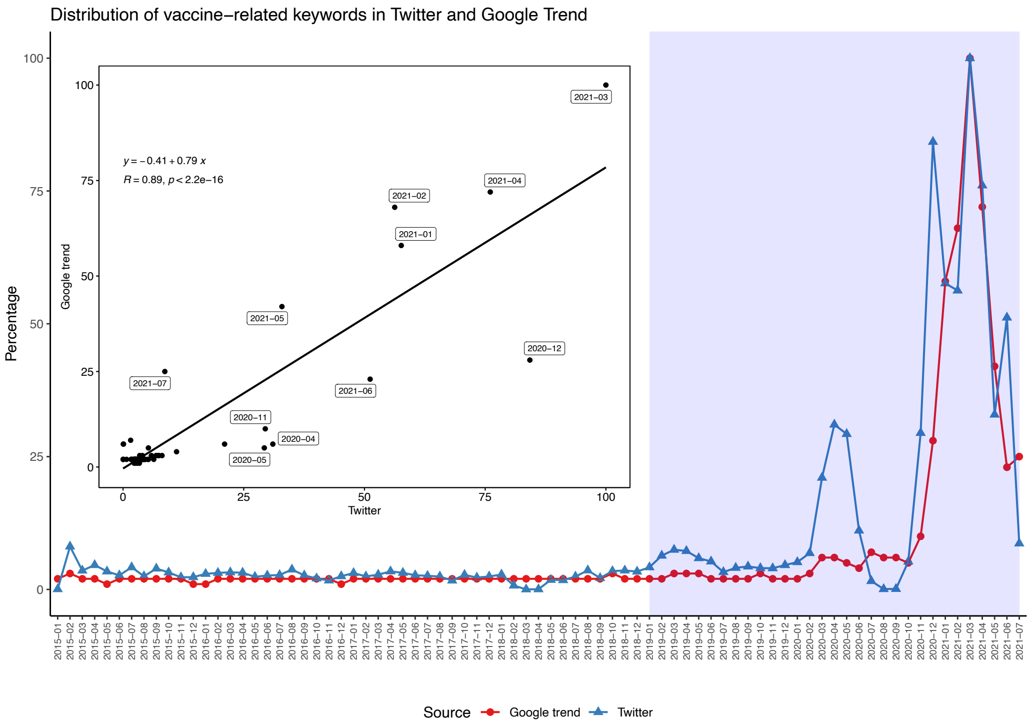


Textual



Sanity Check





Ground Truth Collection

Hand-labeled Tweets: foundation for developing the classifier model

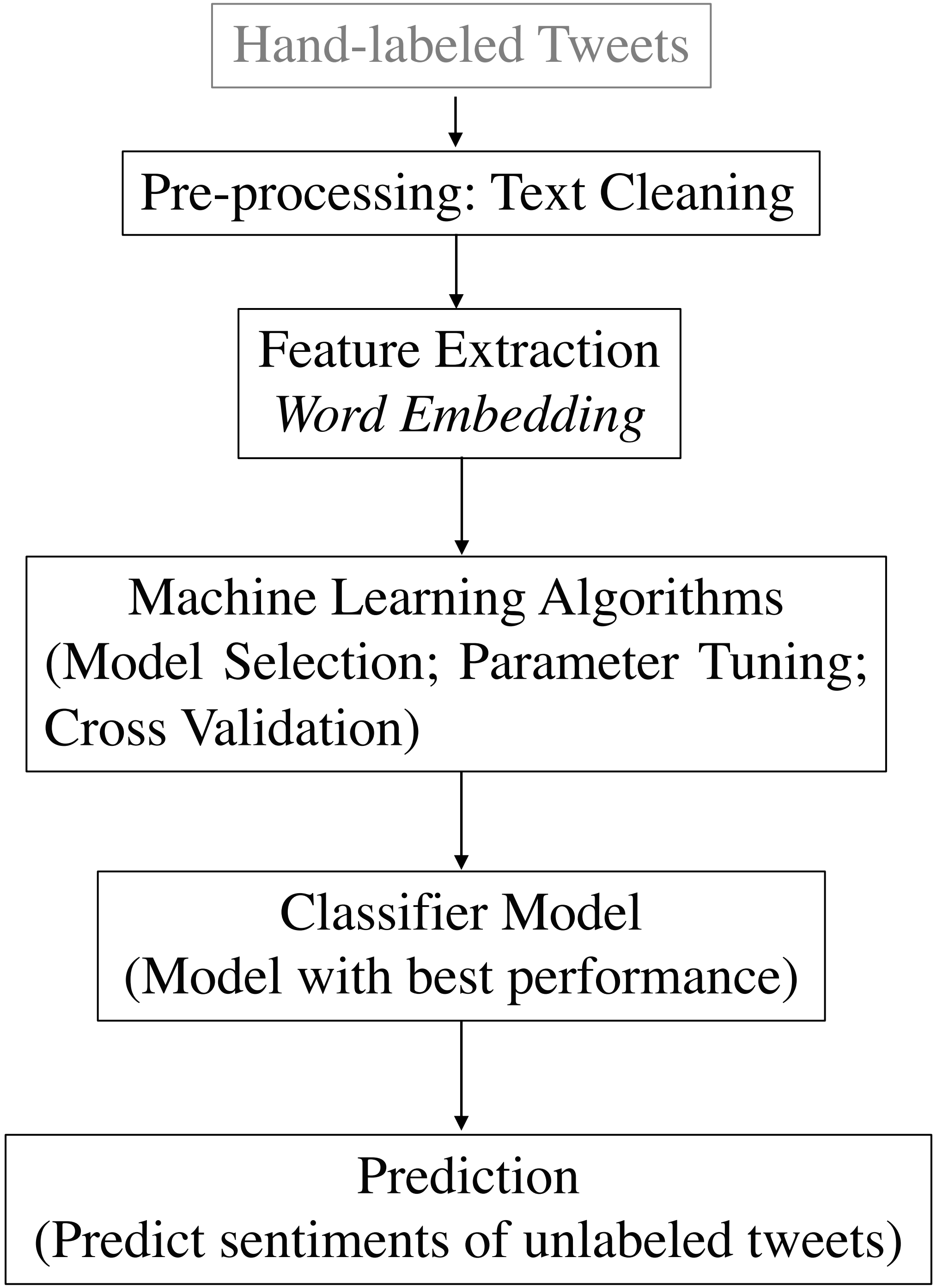
- Dynamic survey (each survey has 100 randomly generated tweets)
- Three sentiment options: (*Pro-vaccine, Neutral, Anti-vaccine*)
- Each tweet should be assigned one sentiment
- Randomly send surveys to volunteers

Sentiment Labels	Examples
Pro-vaccine	Measles can be serious for young children. Make sure your child is up to date on MMR vaccine.
Neutral	State records show Blair County kindergarten measles vaccination rate is among lowest in the state.
Anti-vaccine	The CDC states the 0 have died in the past 10 years of measles while the vaccine has killed 108 according to VAERS.

Hand-labeled Tweets → Reliability check:
- Internal consistency
- Inter-rater reliability

→ Robust training samples

Text Classification



Parameter settings that generates the best performance used in the XGBoost classifier:

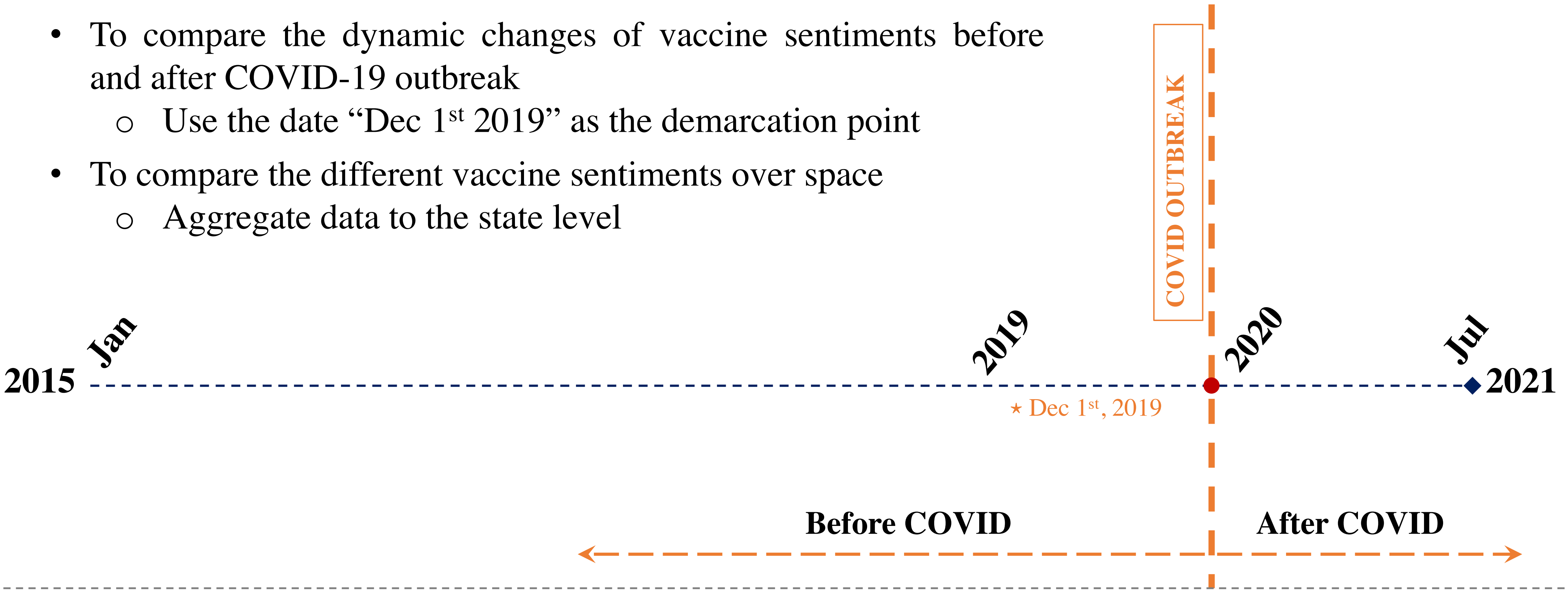
Hyperparameter settings	
Parameters	Values
Number of trees	500
Maximum depth of a tree	8
Learning rate	0.2
γ	0.0

Performance metrics of the XGBoost classifier:

Performance metrics			
	Precision	Recall	F1 Score
Anti-vaccine	0.78	0.61	0.69
Neutral	0.71	0.59	0.65
Pro-vaccine	0.74	0.87	0.80
Macro average	0.75	0.69	0.71
Weighted average	0.74	0.74	0.74
Accuracy		0.74	

Spatial-Temporal Analysis

- To compare the dynamic changes of vaccine sentiments before and after COVID-19 outbreak
  - Use the date "Dec 1st 2019" as the demarcation point
- To compare the different vaccine sentiments over space
  - Aggregate data to the state level



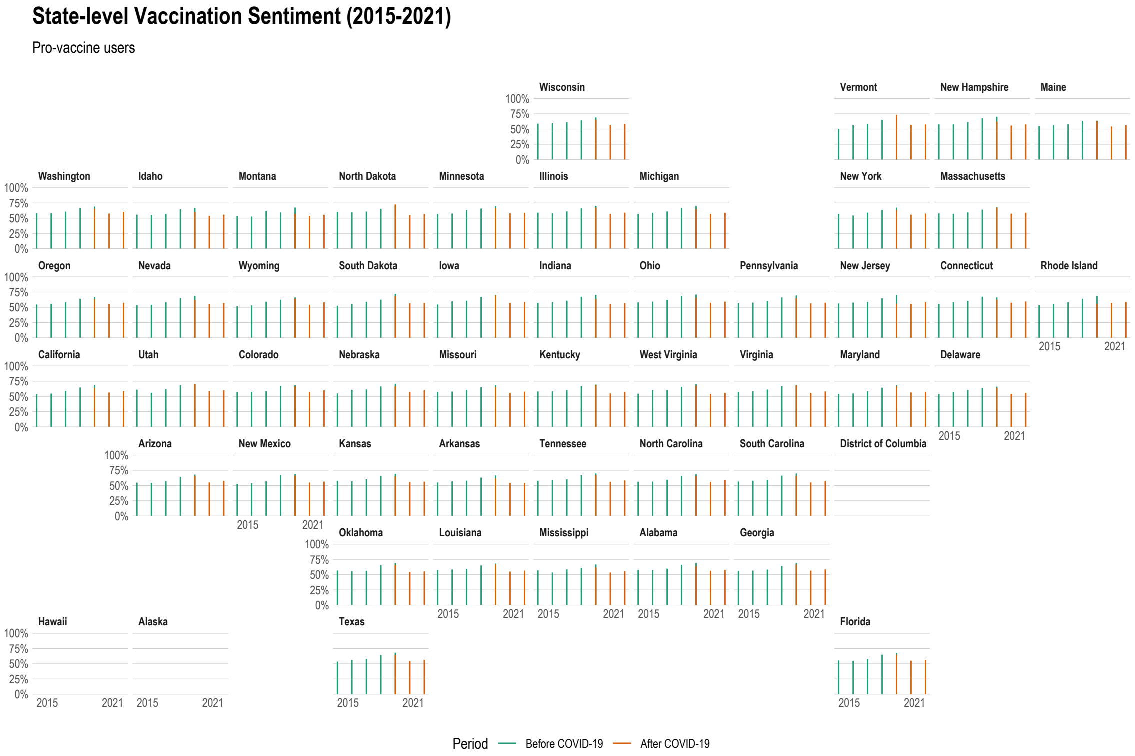
⋆ *Dec 1st, 2019: the date of symptom onset of the 1st patient in Wuhan, China (Huang et al., 2020)*

Results

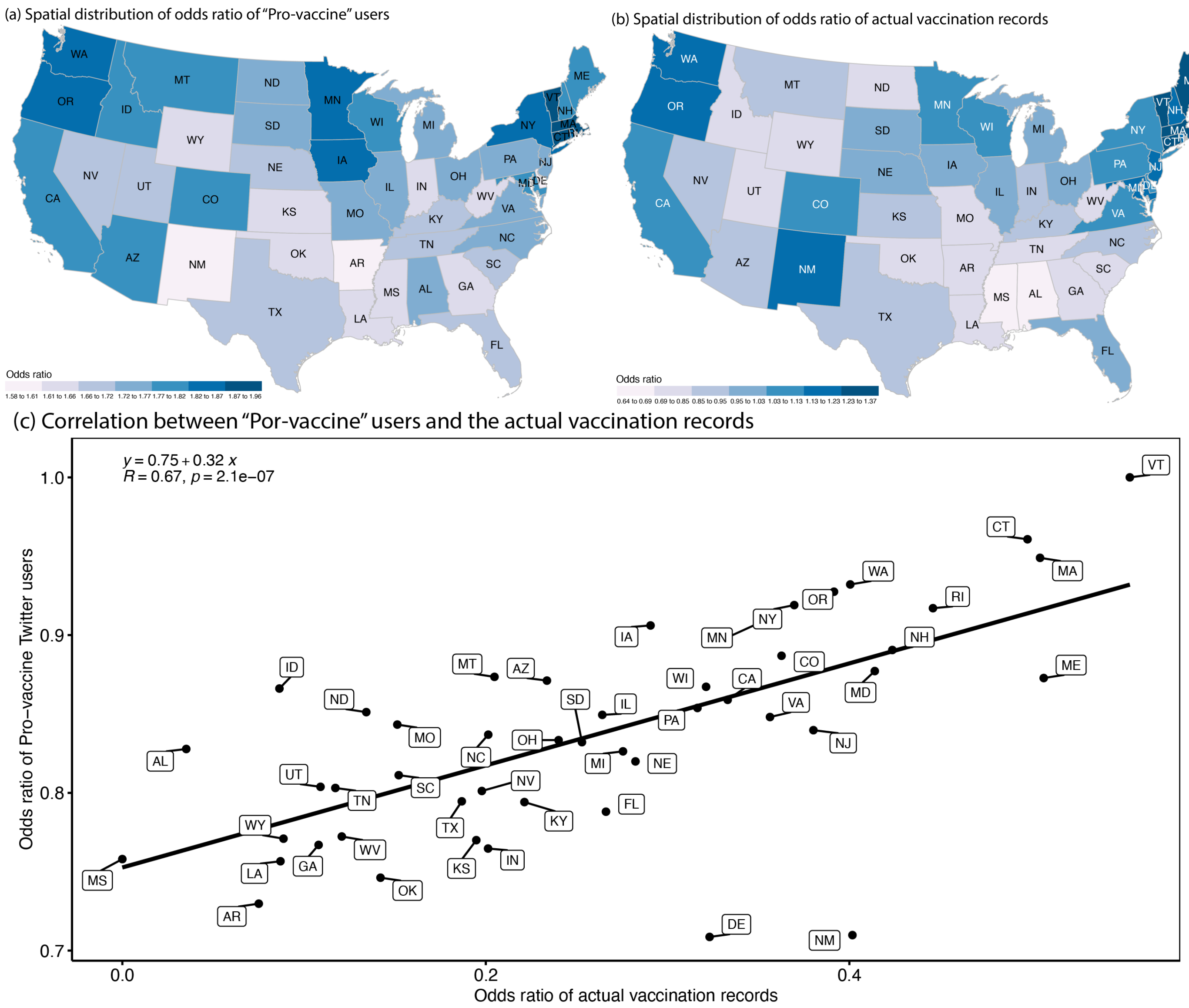
Sentiments	Number of Users	Number of Tweets
Entire period		
Pro-vaccine	2,055,959 (56.97%)	7,190,846 (61.58%)
Neutral	729,868 (20.22%)	2,158,271 (18.48%)
Anti-vaccine	822,926 (22.8%)	2,327,495 (19.93%)
Before COVID-19		
Pro-vaccine	544,365 (61.56%)	1,655,642 (60.56%)
Neutral	161,609 (18.28%)	457,925 (16.75%)
Anti-vaccine	178,339 (20.17%)	620,103 (22.68%)
After COVID-19		
Pro-vaccine	1,631,444 (56.2%)	5,535,204 (61.89%)
Neutral	595,655 (20.52%)	1,700,346 (19.01%)
Anti-vaccine	675,706 (23.28%)	1,707,392 (19.09%)

- The positive vaccine sentiment was the dominant opinion
- The rate of “Pro-vaccine” users decreased after the outbreak (61.56% → 56.20%)
- The percentage of “Anti-vaccine” users revealed a modest increment after the outbreak (20.17% → 23.28%).
- The outbreak indeed moderately shifted public attitudes towards vaccination.

- The rate of “Anti-vaccine” users approached the highest point in 2020, then slightly shrank in 2021;
- The uptake of the coronavirus vaccine(s) in some cases is accompanied by various side effects;
- Coincided with Yousefinaghani et al.’s (2021) results



- A similar trend was observed when disaggregating the sentiments into the state level.
- For most states, the rate of "Anti-vaccine" users increased in 2020 compared to 2019 and showed a minimal drop in 2021, while the changes in the rate of "Pro-vaccine" users over time are the opposite.



The proposed approach for identifying positive vaccine sentiments online can be used as an indicator for evaluating offline vaccination rates.

Online v.s. Offline

Online: *"Pro-vaccine" users online*
Offline: *Actual vaccination rate*
[Our World in Data (Ritchie et al., 2020)]

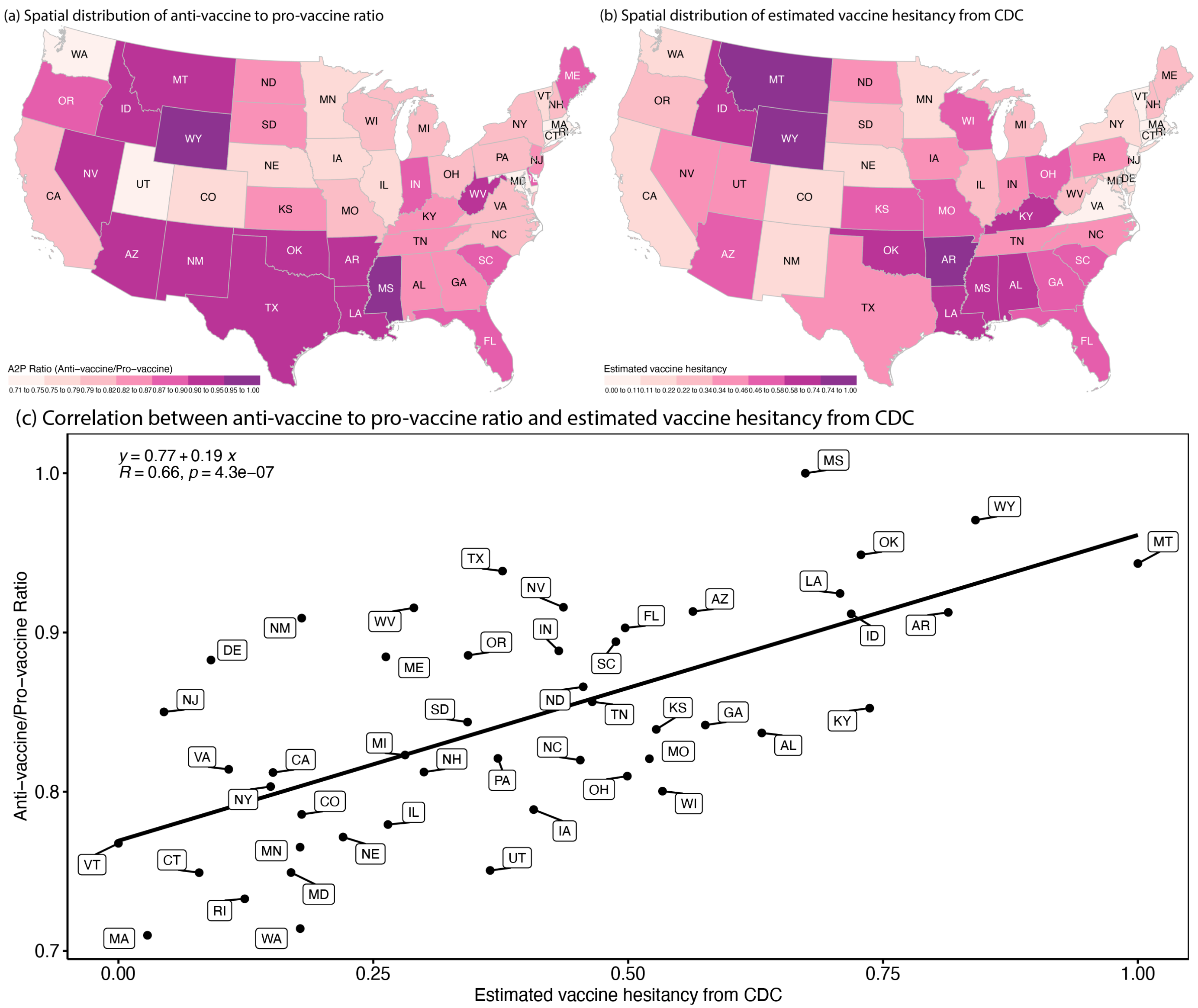
- Odds ratio: to alleviate size-related issues;
- Geographic difference in Pro-vaccine sentiment on Twitter: MA, CT, VT, CO, WA, NY had relatively higher Pro-vaccine odds than other states → relatively complete health system;
- Follow a similar trend to that of the actual vaccination rate;
- A positive correlation (R = 0.67)

Online v.s. Offline

Online: A2P Ratio $R_{A2P} = \frac{Odds\ Ratio_{Anti-vaccine}}{Odds\ Ratio_{Pro-vaccine}}$

Offline: estimated vaccine hesitancy rate from CDC (based on U.S. Census Bureau's Household Pulse Survey (HPS))

- Relatively higher A2P ratios and estimated vaccine hesitancy are mostly entrenched in states in the *__West__* and *__South__* (e.g., WY, AR, FL, LA, NV);
- WY stood out from the other states in both maps, appearing as the most vaccine-hesitant state in the country, (consistent with the result in Douthit et al.'s (2015) study) → inequality in resources allocation and distribution;
- A positive correlation (R = 0.66).



The proposed A2P ratio has the ability to capture a comparable pattern as the estimated vaccine hesitancy obtained from a survey research.

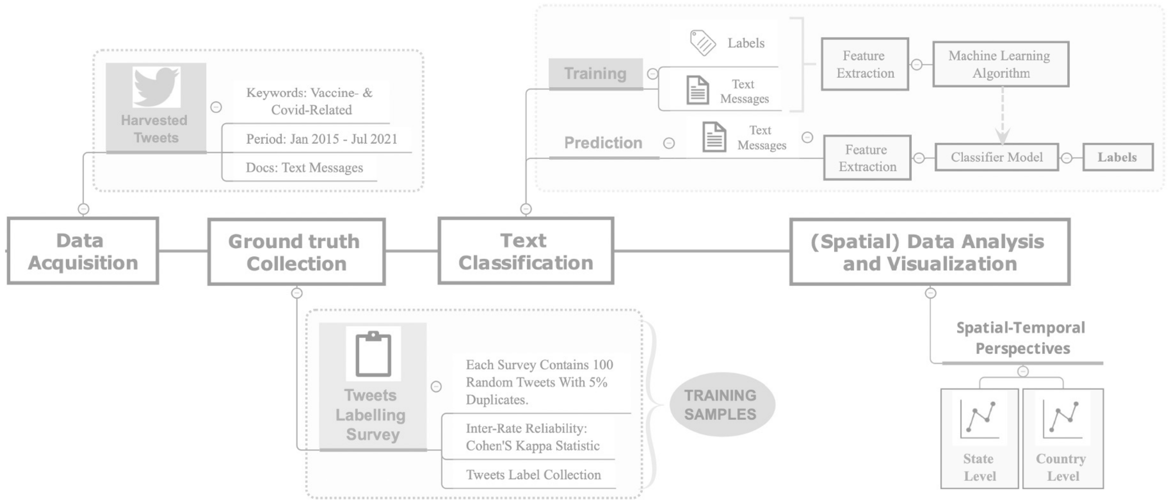
Discussion & Conclusion

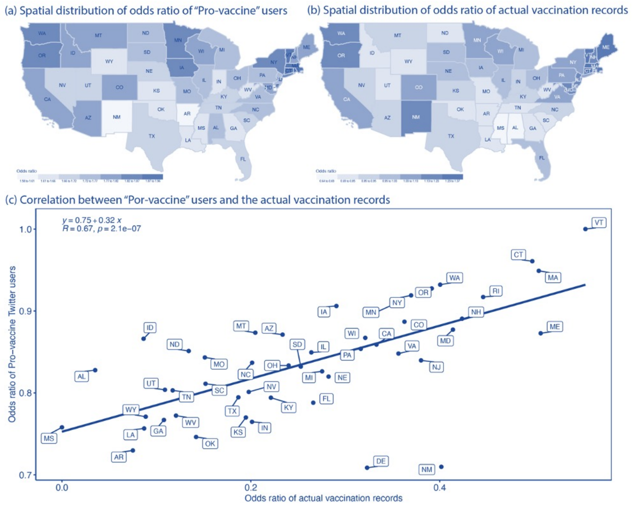
Utilized	Utilized the analytical latitude offered by social media data to compare changes in vaccine sentiments before and after a disease outbreak (SPACE & TIME)
Demonstrated	Demonstrated the effectiveness and practicality of integrating Word Embeding and ML techniques for sentiment analysis
Identified	Identified the positive correlation between online vaccine discussion and offline vaccination rates
Proposed	Proposed a A2P ratio to effectively monitor vaccine hesitancy in near real-time, which complements the limitation of vaccine hesitancy study based on survey research

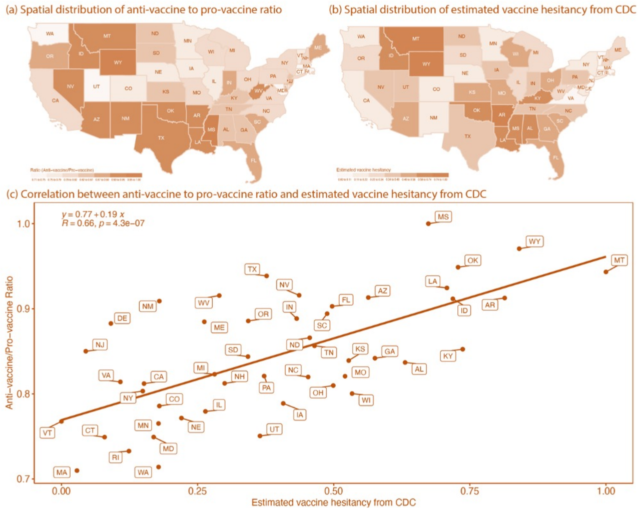
Future Work

- Applies the method to other social media data (e.g., Facebook, Weibo)
- Calibrates or evaluates sampling bias
- Explores key topics/themes discussed online about vaccination

The study offers a scan of the changing pulse of the public's perceptions towards vaccination, which is in light of vaccination campaigns and policy decision-making facing the current COVID-19 and future public health challenges.







THANK YOU FOR YOUR ATTENTION

Qingqing Chen <qchen47@buffalo.edu>
Crooks, Andrew <atcrooks@buffalo.edu>