

## Abstract

### Purpose:

- Vaccines are critical technology to control disease spread, but vaccine refusal compromises efforts towards this end.
- Recent studies focus on the social rationales underlying vaccine refusal [1], but effective techniques to address these rationales remain unclear.
- We use big data to analyze these sociotechnical interactions by studying the distribution of the public's rationales regarding vaccine technology in the US.
- The journal *Vaccine* indicated with a special issue that social media influence patients' behaviors related to vaccines [5]
- And, Twitter has proven effective in disease surveillance [2]
- So we use Twitter to collect and track social trends about vaccine refusal

### Data Source and Methods:

- To obtain social information about vaccine technology, we employ machine learning processes to filter Twitter data described in previous work [3]. This yields
  - a) if a tweet is relevant to vaccination or not;
  - b) if a vaccine tweet is of neutral or non-neutral sentiment, and
  - c) if a non-neutral vaccine tweet is of positive or negative sentiment.
- We also ensure tweets are geotagged using machine learning [4].

### Analyses:

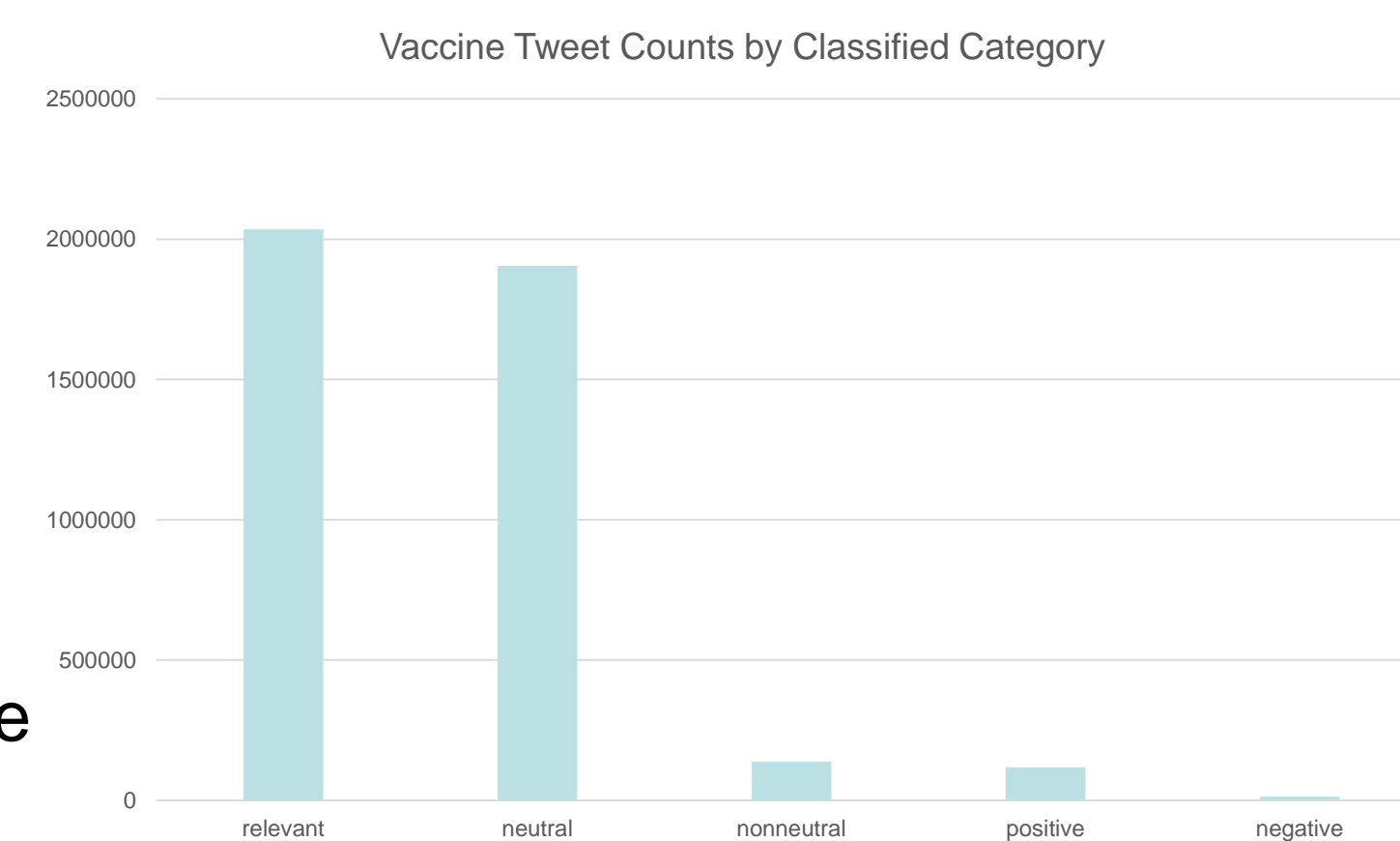
- Temporal, spatial, and social:
  - Use machine learning to detect topics automatically
    - Specific vaccination sub-conversations within various communities
  - Maps of the US detailing in which places topics are discussed
  - Automatic identification of social media users' demographics
    - e.g. age, gender, parent status, income, education [6]
  - By user: maps of the US detailing who discusses vaccination and where
- Combining this information will allow us to pinpoint how vaccine refusal rationales vary across and within communities, in addition to when and where they vary, change, or appear.

### End goal:

- Such analyses will contribute towards understanding the interactions between the spread of a disease, the associated social dynamics, and how technology may be used to reduce overall disease prevalence.

## Data Source: Social Media

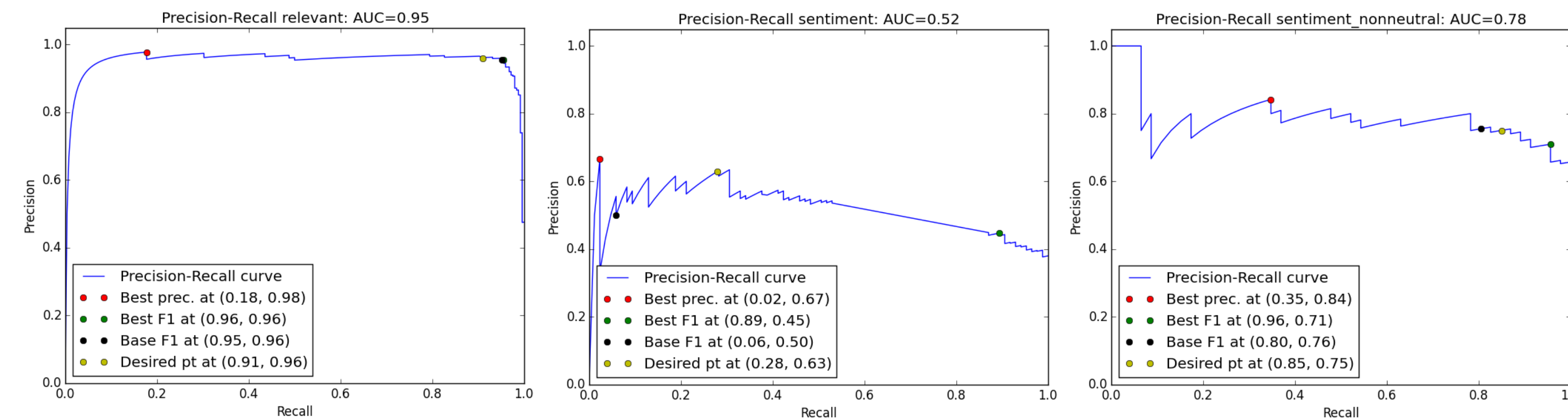
- Data collected beginning from November 2014
- Measles outbreak in Disneyland prominent example of vaccination debate
- Cut off at March 2015 for preliminary analyses
- 2M+ vaccine-relevant tweets
  - Many neutral, but more negative than positive



## Methods and Analyses

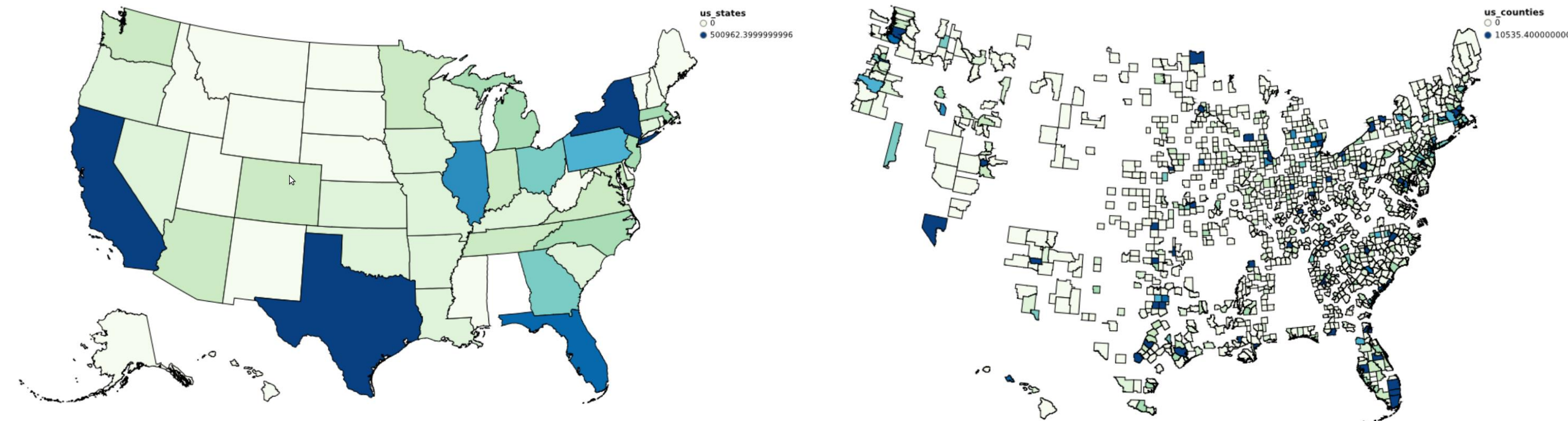
### Relevance and Sentiment:

- Support Vector Machines, Python + sklearn, modified parameters on precision-recall curves
  - Favored precision over recall when necessary (e.g. neutral vs non-neutral sentiment)



### Location augmentation:

- Carmen system infers locations using geocoding tools and automatic and manual alias resolution [3]
- Initial foray yielded 44% data with location info (limitations section for improvements)
  - Where available: lat-lon, city, county, state, country



### Topics:

- Latent Dirichlet Allocation run over all vaccine tweets for 50 topics

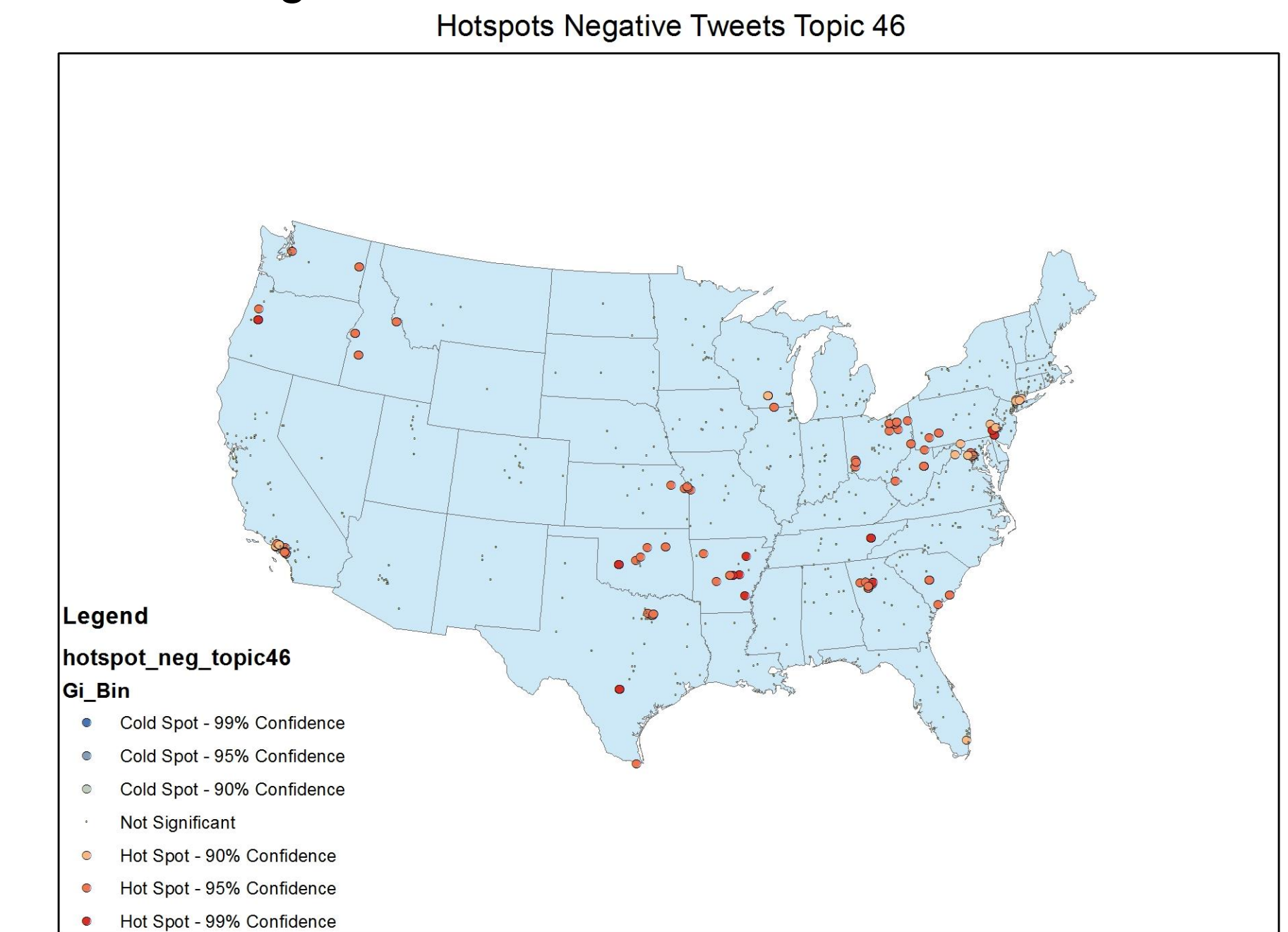
Topic ID	Parameter	Category	Descriptions / samples of keywords from topic
30	0.00521		vaccine, kid, AMP, flu, good, debate, vaccination, anti, children, love, jenny, mccarthy, bad, poisoning, <expletives>, shot, research, sick, baby,
16	0.00501	pro	herd immunity important = anti, people, measles, autism, science, vaxxers, pro, good, movement, stop, stupid, outbreak, bad, vaccineswork, herd, important, die, immune
39	0.00325	pro	vaccine, science, truth, cdwhistleblower, autism, media, facts, tannersdad, debate, doctors, health, pro, trump, pharma, research, reality, vaccineswork
8	0.00292		vaccinated, measles, mmr, cdcgov, preventable, polio, risk, recently unvaccinated
24	0.00263	pro	vaccine, ebola, trial, monkey, effective, hiv, highly, promise, success, show, world, experimental, hope, development
20	0.00236	anti	cdwhistleblower, vaccine, autism, cdcfraud, autismwarriorny, ageofautism, blacklivesmatter
46	0.0023	anti	vaccines, bill, california, mandatory, exemptions, children, parents, medical, school, religious, forced, mandate, personal, senate, freedom, require, laws, tyranny, eliminate
9	0.00215		vaccinate, kids, parents, school, doctor, refuse, unvaccinated, slippery, slope, disneyland, politics, <expletives>, love, diseases, child, catch, admit, eliminated, choose, absolutely, thetweetofgod= satirical
34	0.00209	anti	vaccine, measles, autism, children, adverse, reactions, cdwhistleblower, mexico, risk, infant, mmr, death, bad, kill, die, vaccinateyourkids

Acknowledgement: Research supported by a grant from the National Institutes of Health-National Institute of General Medical Sciences - 5R01GM114771-02

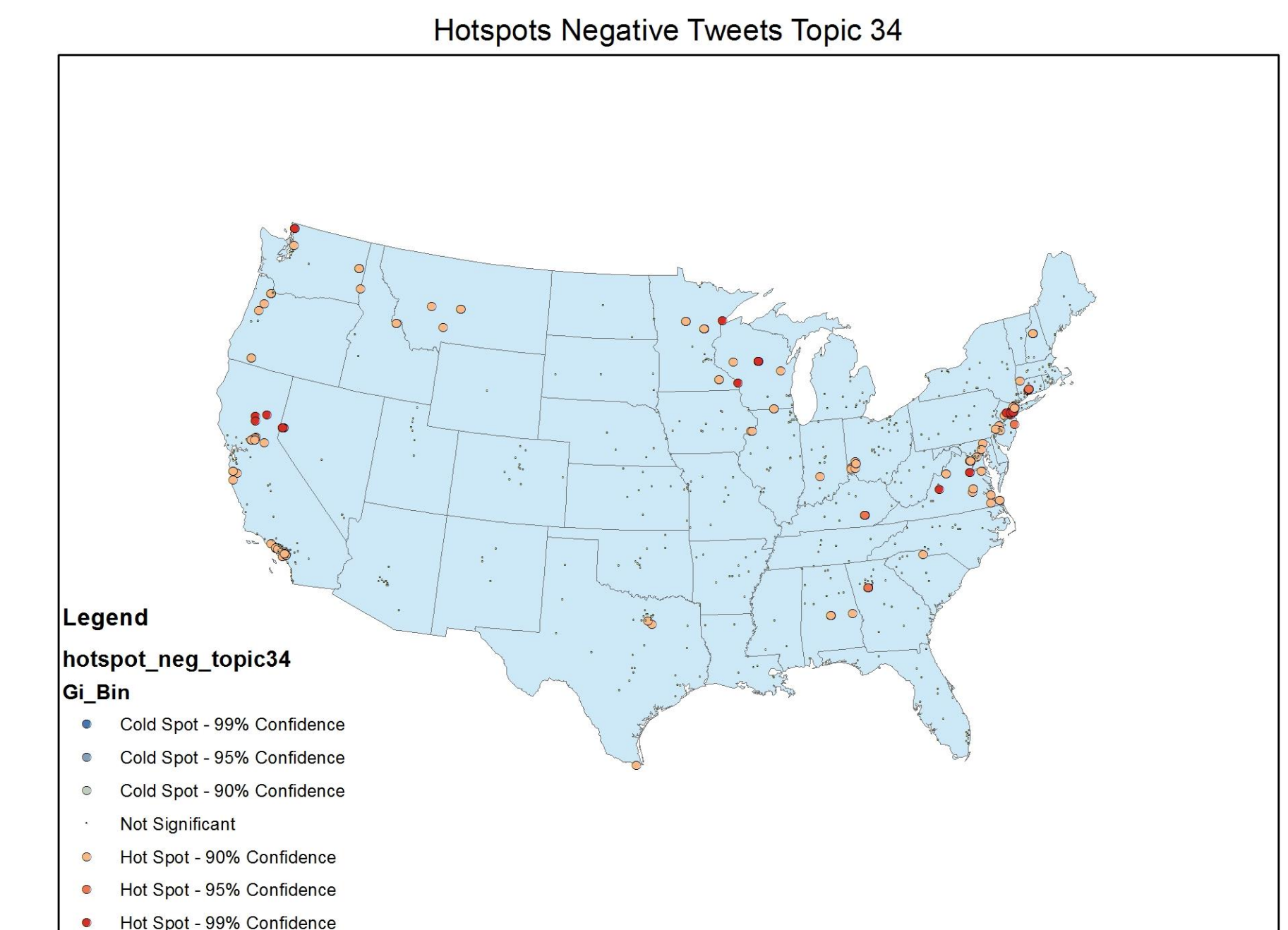
## Analyses – Cont'd

### Spatial:

- ArcGIS 10.2, Getis-Ord Gi\* (statistical hotspot analysis) [6]
  - Topic 46 – anti: gov't mandate



- Topic 34 – anti: autism and adverse reactions in kids



### Limitations / Improvements:

- Group and analyze by available granularity of location
- Inspect and deal with mistaken or mislabeled lat-lons -> improve Carmen
- Check for bot accounts reverberating information
- Improve topics on subsets of tweets, 'de-trend' the general debate
  - Supervised topic models
- Incorporate demographics

## References

- Downs, Julie S., Wändi Bruine de Bruin, and Baruch Fischhoff. 2008. "Parents' Vaccination Comprehension and Decisions." *Vaccine* 26 (12): 1595–1607.
- Broniatowski, David A., Michael J. Paul, and Mark Dredze. 2013. "National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic." *PLOS ONE* 8 (12): e83672.
- Dredze, Mark, David A. Broniatowski, Michael C. Smith, and Karen M. Hilyard. 2016. "Understanding Vaccine Refusal: Why We Need Social Media Now." *American Journal of Preventive Medicine* 50 (4): 550–52.
- Dredze, Mark, Michael J Paul, Shane Bergsma, and Hieu Tran. 2013. "Carmen: A Twitter Geolocation System with Applications to Public Health." In *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, 20–24. Citeseer.
- Betsch, C, and K Sachse. 2012. "Special Issue: The Role of Internet Use in Vaccination Decisions." *Vaccine* 30 (25): 3723–3818.
- Getis, Arthur, and J. K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24 (3): 189–206.
- Culotta, Aron, and Nirmal Kumar Ravi, and Jennifer Cutler. 2016. "Predicting Twitter User Demographics Using Distant Supervision from Website Traffic Data." *Journal of Artificial Intelligence Research* 55: 389–408.