Research Plan

Jaime Lara Levin: Categorizing geocoded anti-vaccination tweets in urban areas using Latent Dirichlet Allocation (LDA) and dictionary based modeling

In recent years, a highly feminized anti-vaccination movement has taken hold of social media platforms, predominantly Twitter. (Larson et al., 2014; Chen & DeStefano, 1998; Chen & Hibbs, 1998; Smailbegovic et al., 2003) Twitter's features, including the ability to "retweet" like-minded messages, contribute to ideological overlap on the platform and the creation of online echo chambers, which lead to an amplification and reinforcement of similar vaccine misinformation (Witteman & Zikmund-Fisher, 2012; Eady et al., 2019). The detrimental nature of vaccine misinformation online has led researchers to study the content and geographic origins of anti-vaccination tweets. Past studies have assessed the most frequent words that appear in the context of anti-vaccination discourse on Twitter (Radzikowski et al., 2016), while others have examined prominent geographic locations of Twitter users who are anti-vaccination proponents (Brooks, 2014; Lieu et al., 2015).

While both content analysis and geographic clustering have been studied separately among anti-vaccination users on Twitter, the geographic base of each subcategory of anti-vaccination tweets and retweets is unknown; the reasons for anti-vaccination attitudes as expressed on Twitter in distinct geographic regions of the United States are unclear. This study will aim to test whether there is a significant difference in Twitter users' arguments against vaccination in different geographic regions of the United States. The study intends to analyze users' arguments against vaccination in urban areas on opposite coasts, in particular, where there is a high prevalence of anti-vaccine rhetoric.

This study will utilize R's Twitter streamer, rtweet, to collect anti-vaccination tweets surrounding four urban cities on opposite coasts of the United States: Rochester, New York City, San Francisco, and Los Angeles. After streaming anti-vaccination tweets surrounding these four locations, this study plans to utilize Latent Dirichlet Allocation (LDA) to create categories or "topics" for different types of anti-vaccination tweets in both California and New York. The study will then create a topic density proportion to measure the relative density of particular arguments against vaccination within cities in these two states. Understanding where anti-vaccination sentiment is most prevalent and what the sentiment discusses is crucial to combating the outbreak and spreading of deadly diseases across the country.

Research Plan

References

Brooks, B. Using Twitter data to identify geographic clustering of anti-vaccination sentiments (2014). Accessed at https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/26187/Brooks_washington_0250O_13777.pdf?sequence=1&isAllowed=y

Chen, R. T. & Destefano, F. Vaccine adverse events: causal or coincidental? *The Lancet* **351,** 611–612 (1998).

Chen, R. T. & Hibbs, B. Vaccine Safety: Current and Future Challenges. *Pediatric Annals* **27,** 445–455 (1998).

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. *SAGE Open*, (2019).

Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. & Paterson, P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine* **32,** 2150–2159 (2014).

Lieu, T. A. (2015). Geographic Clusters in Underimmunization and Vaccine Refusal. *Pediatrics*, *135*(2). doi: 10.1542/peds.2014-2715d.

Radzikowski, J, Stefenidis, A, Jacobsen, KH, Croitoru, A, Crooks, A, Delamater, PL. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. *JMIR Public Health and Surveillance* **2** (2016).

Smailbegovic, M. S., Laing, G. J. & Bedford, H. Why do parents decide against immunization? The effect of health beliefs and health professionals. *Child: Care, Health and Development* **29,** 303–311 (2003).

Witteman, H. O. & Zikmund-Fisher, B. J. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. *Vaccine* **30,** 3734–3740 (2012).