ABSTRACT
The mass shooting at Sandy Hook elementary school on December 14, 2012 catalyzed a year of active debate and legislation on gun control in the United States. Social media hosted an active public discussion where people expressed their support and opposition to a variety of issues surrounding gun legislation. In this paper, we show how a content-based analysis of Twitter data can provide insights and understanding into this debate. We estimate the relative support and opposition to gun control measures, along with a topic analysis of each camp by analyzing over 70 million gun-related tweets from 2013. We focus on spikes in conversation surrounding major events related to guns throughout the year. Our general approach can be applied to other important public health and political issues to analyze the prevalence and nature of public opinion.

1. INTRODUCTION
Gun control in the United States is a major public policy issue that has polarized US society [26]. Although public opinion has been strongly in favor of stricter gun control policies for over two decades [29], federal gun control legislation has been a hotly contested issue meeting little legislative success, where even local restrictions have been met with opposition [31]. Insofar as public opinion affects the bills debated and passed into law, accurately gauging public opinion and salience on the various issues associated with gun control is important to inform the legislative process [6, 18, 21].

Public opinion is typically estimated through written or telephone surveys where subjects are asked to share their level of approval of different policies up for debate [2, 15]. Assuming the population is uniformly sampled and that subjects are able and willing to divulge their true beliefs, these are reliable proxies for public opinion. Gun control polls are often conducted over the phone and ask respondents about their gun ownership, as well as opinions on different forms of gun control legislation (e.g., “Saturday night special” bans, assault weapon bans, national firearm registration, universal background checks) [29]. However, traditional surveys have a number of drawbacks, including limitations on the response types and cost restrictions on producing timely results. These limitations are well known in the public health realm where surveys, a critical data source for a variety of public health topics, are facing increasing feasibility challenges. As a result, researchers have turned to new data sources, such as search queries [10] and social media [7]. Social media has been used to estimate public opinion on a range of topics, including political sentiment [4, 20, 25, 28] and a range of public health topics [3], including gun control [4]. Some work has looked at gun control tweets, but has focused on argument framing and not measuring public opinion [27].

Issues of gun control came to the forefront of national discussion with the mass shooting at Sandy Hook Elementary School in Newtown, Connecticut on December 14, 2012. This tragedy followed six months after another mass shooting in an Aurora, Colorado movie theater, and prompted a concerted effort to pass stronger gun restrictions at the federal level. In April, 2013, a bill to expand background checks was defeated in the senate, ending federal legislative efforts. Failure to pass national gun control legislation led many states, including Colorado and Connecticut, to pass their own gun control bills.

Public opinion played a major role throughout this time period, where discussions of gun control on social media rose in prevalence and prominence. The richness of social media data, where we have both overall prevalence, content and location data, presents new opportunities for analyzing and understanding the nature of public opinion surrounding guns.

We present an analysis of gun-related Twitter data from all of 2013, over 70 million tweets in total. We focus on two main questions: 1) Do Twitter conversations in support of or opposition to gun control reflect public opinion as measured by traditional surveys? 2) What events generate online activity from gun control supporters and opponents.
the tracking of topic proportions in a corpus over time [11]. Topic models have become popular tools for analyzing text data in social science [12], the humanities [13, 14] and health [22, 23], with numerous examples of applications to Twitter data [15, 24, 32].

We sub-sampled 6 million tweets (8.5% of the total collection) to train an LDA model, and then used the learned parameters to infer document specific topic distributions for each tweet. Tweets were tokenized by non-alphanumeric characters into unigrams and filtered using a stopword list specific to Twitter. We retained the 40,000 most frequent word types for learning. We used the LDA implementation in Mallet [16] and tuned model parameters on a held out set of 1 million tweets to maximize model log-likelihood. We swept the number of topics from 25 to 500, and the document-topic Dirichlet prior hyper-parameter α from 0.25 to 10 (with an asymmetric prior.) We used Mallet’s parallel Gibbs sampler with a burn-in of 100 iterations, 500 total iterations, with hyper-parameter optimization every 10 iterations. Our tuned model used an initial α = 1 and 250 topics. The final model was then used to infer topics for the entire corpus using 200 sampling iterations.

We obtained a location for each tweet using Carmen [8], a high-precision geocoder for Twitter based on a user’s profile. Wherever possible, we obtained the US state associated with a tweet. We chose to rely on an automatic geocoder since the proportion of tweets with location information provided by Twitter was small (around 1-2%).

Using the sentiment coded tweets and their inferred topic distributions, we measured the following trends. 1) The overall number of gun-related tweets for each day and week during 2013. 2) The number of Control and Rights tweets for each day and week. Since the overall Twitter volume remained relatively stable in 2013, our counts are not normalized. 3) The most likely topics associated with Control or Rights tweets over the entire corpus, as well as for each week. This gives us a fine-grained look at which topics were discussed by each gun control camp for each week. We compute these trends for both the entire United States and for each US state.

3. RESULTS

3.1 Comparison with Polling Data

We begin by measuring the ability of Twitter to track gun related opinions as compared to results from traditional sur-

| Keyword type | Keywords |
|--------------|----------|
| General      | gun, guns, second amendment, 2nd amendment, firearm, firearms |
| Control      | #guncontrol, #guncontrolnow, #momsdemandaction, #momsdemand, #demandplan, #nowaynra, #gunskillpeople, #gunviolence, #endgunviolence |
| Rights       | #gunrights, #protect2a, #molonlab, #molonlab, #noguncontrol, #progun, #nogunregistry, #voteongunsrights, #firearmrights, #gungrab, #gunfriendly |

Table 1: Keywords used to collect tweets are listed as General keywords, and hashtags suggesting a Control or Rights gun control stance.

and how do the arguments and issues discussed change in response to these events? While there has been significant work addressing our first question in regards to other topics of public opinion [20, 21], the second question gives us a new framing in terms of social media studies; we are concerned with what social media users are saying about gun control, in addition to how many people are saying it.

2. METHODS

Our data set contains 70,514,588 publicly-available tweets collected using the Twitter streaming API based on keywords and phrases associated with guns or gun control in the United States: gun, guns, second amendment, 2nd amendment, firearm, firearms. Our collection covers just over one year, starting on December 16, 2012 (two days after the Sandy Hook shooting) and ending on December 31, 2013.

We identified hashtags indicative of support for (Control) or opposition to (Rights) gun-control as a rough estimate of sentiment towards gun control. These hashtags were strongly associated with either the Control or Rights gun control positions. We obtained this list by examining the most popular hashtags in a subset of our data and selecting those that strongly indicated either one of these positions. Table 1 shows these hashtags: 11 for Control and 11 for Rights. A tweet was labelled as Control gun control if it contained more Control hashtags than Rights, and vice-versa for Rights tweets. A total of 304,142 tweets were labelled as Control and 125,936 as Rights using this method. Although only about 0.6% of tweets were coded with gun control stance, this labelling method resulted in a high precision coding of tweets by gun control stance. Although only about 0.6% of tweets were coded with gun control stance, this labelling method resulted in a high precision coding of tweets by gun control stance. We leave for future work statistical methods that identify gun control sentiment of a larger percentage of our data [14, 30].

Our content analysis of the sentiment coded tweets relies on latent Dirichlet allocation (LDA) [5], a data-driven probabilistic topic model that can identify the major thematic elements in a text corpus. Topic models infer the parameters of a probability distribution with Bayesian priors, producing for each topic a distribution over the words in the corpus. Reviewing the most probable words for each topic is a common technique for establishing a semantically grounded label for the topic. Additionally, the model assigns a distribution over topics to each document (tweet), which enables the tracking of topic proportions in a corpus over time [11].

| State        | Date          | State        | Date          |
|--------------|---------------|--------------|---------------|
| Alaska       | 4/26/2013     | Montana      | 6/23/2013     |
| Arizona      | 4/26/2013     | Nevada       | 4/26/2013     |
| Arkansas     | 5/23/2013     | North Carolina | 5/14/2013 |
| Georgia      | 5/23/2013     | North Carolina | 7/14/2013 |
| Georgia      | 8/5/2013      | Ohio         | 4/26/2013     |
| Iowa         | 6/7/2013      | Ohio         | 8/19/2013     |
| Louisiana    | 5/1/2013      | Tennessee    | 5/23/2013     |
| Louisiana    | 8/19/2013     | Texas        | 7/1/2013      |
| Michigan     | 6/2/2013      | Virginia     | 7/14/2013     |
| Minnesota    | 5/19/2013     | Wyoming      | 7/21/2013     |

Table 2: Description of the states that were polled by Public Policy Polling, and the date they were polled. Dates are the last day the poll was conducted.
survey methods. We obtained US state level polling for 16 states gathered between April 4, 2013 and August 19, 2013 by Public Policy Polling — a total of 20 polls. The state and date of each poll is included in Table 2. Our sentiment coding technique identified 304,142 tweets as Control and 125,936 Rights. Of these tweets, a total of 165,360 (38%) were geocoded with a US state.

While our sentiment coding of tweets was for a coarse Control/Rights position on gun control, the polls do not directly ask this question. Therefore, as a proxy we selected the following question which appeared in all polls: “Would you support or oppose requiring background checks for all gun sales, including gun shows and the Internet?”.

We used the proportion of “yes” answers from each poll as the value for each US state. For states that had two polls, we used each poll as a separate data point in our correlation. For Twitter, we measured the proportion of Control tweets over the number of both Control and Rights tweets for each US state over our entire collection. Due to data sparsity, we did not limit the tweets to consider only those from the time period the poll was taken.

We obtained a Pearson correlation coefficient of 0.51 between our two variables: proportion “yes” in state polls and proportion of Control tweets. Figure 1 displays the least-squares fit between these two variables, with an $R^2$ value of 0.22. This is a reasonably strong relationship between the variables, demonstrating that relative proportion of buzz in gun conversations on Twitter are reflective of opinions of the actual population.

This reasonably strong relationship was obtained even with several important limitations on our method. First, public opinion varied over time [9], yet these state polls capture just a single point in time, and different time periods at that. The time period of our Twitter data was mismatched to these polls, in that we used tweets from the entire corpus instead of restricting them to the time when the poll was conducted. Doing so would have yielded too few tweets, though future work that expanded our sentiment classifier method could address this problem. Second, we were only able to obtain polls and sufficient tweets for some US states, which reduces our ability to validate this method over the entire United States. Even though this was a major issue in US politics for a sustained period of time, polls were not conducted for every state. Third, gun control opinions can be complex, yet we are measuring only a coarse level of sentiment. The complex opinions expressed on Twitter may not map directly to our selected question. Fourth, we counted tweets, not the number of accounts tweeting. A single prolific account could bias our estimates. Finally, additional errors could be introduced by the accuracy of the geocoder, Twitter’s representativeness of the US population, and the biases and sampling errors inherent in surveys. Despite these limitations, the obtained correlation is a strong indicator of the value of Twitter data for opinion analysis.

### 3.2 Opinions Surrounding Events

We next contextualize opinions as expressed on Twitter within the context of major gun control related events during 2013. We identified significant spikes in activity using the weekly aggregated statistics of Control and Rights tweets. For each spike, we used a historical news collection to identify major gun related events corresponding to the spike. Figure 2 displays Twitter traffic per week by Control and Rights with large spikes annotated with co-occurring events. Events of note are:

- President Obama promises stronger national gun control legislation – Control tweets spike (December 19, 2012)
- The first gun control senate hearing featuring appearances from Gabrielle Giffords and Wayne LaPierre –

![Figure 1: Proportion of Control gun control tweets, over all Control/Rights tweets from that state, against the percent polled in that state supporting universal background checks.](http://projects.fivethirtyeight.com/pollster-ratings/)

![Figure 2: Number of Control and Rights tweets over time. Events of interest are annotated above spikes in activity.](http://nyti.ms/1GfO4jo)
In our corpus, on average, Control tweets were much more common. Rights tweets eclipsed those of Control when gun control legislation failed to pass in April.

### 3.3 Major Topics of Discussion

Beyond detecting the overall sentiment surrounding each event, we characterized the content of each side by examining the topics discovered by the topic model. We selected the ten most likely topics for both the Control and Rights tweets over the entire time period: 304,142 Control tweets and 2,802,636 tokens and 125,936 Rights tweets (1,265,765 tokens). Tables 3 and 4 show these topics, their likelihood, the most likely words, and our assigned label.

These topics and their relative order summarize the main thrusts of the conversation for both the Control and Rights groups. For example, topic 237 centers around the group “Moms Demand Action for Gun Sense in America” and topic 246 around universal background checks – both topics prevalent in Control tweets. Topic 6 discusses armed robbery (presumably as an argument against new gun restrictions that would prevent citizens from protecting themselves) and topic 5 contains language indicative of political conservatives and second amendment rights advocates, in general.

We next contextualized these topics within the events described above. We computed the distribution over the 10 topics for the Control tweets and the 10 topics for the Rights tweets in the week around the event. By comparing how the usage of these topics change for each event, we can compute the dominant topics of conversation around each event. Figure 6 shows the relative proportion of the top 10 topics, overall for the set of Control and Rights tweets independently, as well as their proportion during each of the events. Topics are ordered by the their relative proportion over all Control or Rights tweets.

When President Obama initially promised federal gun control legislation, gun control advocates tweeted much more frequently about it, but this was not as prevalent during most other events, or overall. Universal background checks and models of more restrictive gun control policy are also mentioned much more frequently during the first Senate hearing on gun control.

---

Table 3: Top 10 topics ranked by \( \text{Prob(topic|Control)} \).

| #  | prob  | Representative tokens | Label |
|----|-------|-----------------------|-------|
| 237 | 0.222 | violence | “Common sense” demand more | “Common sense” demand more |
| 136 | 0.129 | nra | NRA | NRA |
| 7 | 0.108 | violence, barack obama | National gun legislation | National gun legislation |
| 212 | 0.077 | nra | Mix of hashtags | Mix of hashtags |
| 57 | 0.051 | americans died | Domestic violence > foreign violence | Domestic violence > foreign violence |
| 246 | 0.040 | background checks | Universal background checks | Universal background checks |
| 120 | 0.039 | crime, laws, rate | Model gun control policy | Model gun control policy |
| 48 | 0.033 | americans, newtown, america | Domestic violence | Domestic violence |
| 211 | 0.020 | nra, demand, control | Boycott | Boycott |
| 216 | 0.016 | make, people, aware, from side | Safety | Safety |

Table 4: Top 10 topics ranked by \( \text{Prob(topic|Rights)} \).

| #  | prob  | Representative tokens | Label |
|----|-------|-----------------------|-------|
| 69 | 0.258 | tweets | Conservative hashtags/gun registry | Conservative hashtags/gun registry |
| 121 | 0.164 | state law, tex, nra | State gun laws | State gun laws |
| 170 | 0.163 | nra, tweets | Conservative hashtags, mic. | Conservative hashtags, mic. |

---

http://www.huffingtonpost.com/2013/03/30/gun-control-hearing_s_3908651.html
http://www.huffingtonpost.com/2013/04/04/controlgunsrights_nesskorea-19921.html
http://www.huffingtonpost.com/2013/01/30/gun-control-hearing_s_2580691.html
http://www.huffingtonpost.com/2013/04/04/connecticut-gun-control-sandy-hook-law_n_225170.html
http://www.huffingtonpost.com/2013/04/04/gun-control-sandy-hook-law_n_225170.html
http://www.huffingtonpost.com/2013/07/01/gun-control-colorado_n_3033097.html

---

http://www.nytimes.com/2013/03/30/us/obamas-promised-gun-control.html
http://www.nytimes.com/2013/04/04/nytcodex/obama-promised-gun-control.html
http://www.nytimes.com/2013/03/30/us/obamas-promised-gun-control.html
http://www.nytimes.com/2013/04/04/nytcodex/obama-promised-gun-control.html
When federal gun control legislation was first promised, gun rights tweets centered mostly around self-defense applications and state laws permitting carrying guns. During the first senate hearing on gun control, discussion also focused more on restrictions on assault weapons. As time progressed, former secret service agent and Republican political candidate, Dan Bongino became more vocal about gun rights. This is reflected in a greater proportion of tweets mentioning him.

4. DISCUSSION

By analyzing a year’s worth of tweets on guns in the United States, we find variation in each side’s reaction to gun-related events, as well as variation in the arguments cited by each group during events of interest. Control advocates are very vocal early on in the debate when national legislation is still a possibility, but die down later on. From Figure 3, it is clear that a large proportion of this chatter was about national gun control legislation (Topic 7). Rights advocates became more vocal once the national legislation for universal background checks failed in congress, and much of their subsequent discourse focused on an assault weapons ban (Topic 75), the senate filibuster (Topic 227), and political candidate and gun rights advocate Dan Bongino (Topic 99).

We believe that this style of social media analysis is a complement to traditional polling techniques, which typically gauge opinion on a small set of issues. By fitting a topic model to the entire collection of gun-related tweets in 2013, we are able to identify salient issues and arguments for both camps, which researchers may not have identified as relevant, a priori. Most importantly, other than the keywords we searched for to collect this dataset and hashtags we used to label Control and Rights gun control tweets, there was no tailoring of our analysis to the gun control domain. This method of social media analysis can be applied to a wide range of salient public policy issues.

5. REFERENCES

[1] J. W. Ayers, B. M. Althouse, E. C. Leas, T. Alcorn, and M. Dredze. Big media data can inform gun violence prevention. In Bloomberg Data for Good Exchange, 2016.
[2] C. L. Barry, E. E. McGinty, J. S. Vernick, and D. W. Webster. After Newtown – public opinion on gun policy and mental illness. New England journal of medicine, 368(12):1077–1081, 2013.
[3] A. Benton, M. J. Paul, B. Hancock, and M. Dredze. Collective supervision of topic models for predicting surveys with social media. In Association for the Advancement of Artificial Intelligence (AAAI), 2016.
[4] A. Bermingham and A. F. Smeaton. On using twitter to monitor political sentiment and predict election results. In IJCNLP Workshop on Sentiment Analysis where AI meets Psychology, 2011.
[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. J. Mach. Learn. Res., 3:993–1022, Mar. 2003.
[6] P. Burstein. The impact of public opinion on public policy: A review and an agenda. Political Research Quarterly, 56(1):29–40, 2003.
[7] M. Dredze, D. A. Broniatowski, M. Smith, and K. M. Hilyard. Understanding vaccine refusal: Why we need social media now. American Journal of Preventive Medicine, 2015.
[8] M. Dredze, M. J. Paul, S. Bergsma, and H. Tran. Carmen: A twitter geolocation system with applications to public health. In AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI), 2013.
[9] A. M. Gallup. The Gallup Poll: Public Opinion, 2000. Rowman & Littlefield, 2001.

[10] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[11] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

[12] J. Grimmer. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35, 2010.

[13] L. Hong and B. D. Davison. Empirical study of topic modeling in twitter. In *KDD Workshop on Social Media Analytics*, pages 80–88, 2010.

[14] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent twitter sentiment classification. In *Association for Computational Linguistics (ACL)*, 2011.

[15] G. Kleck, M. Gertz, and J. Bratton. Why do people support gun control?: Alternative explanations of support for handgun bans. *Journal of Criminal Justice*, 37(5):496–504, 2009.

[16] A. K. McCallum. Mallet: A machine learning for language toolkit. http://www.cs.umass.edu/~mccallum/mallet, 2002.

[17] D. M. Mimno. Computational historiography: Data mining in a century of classics journals. *JOCCH*, 5(1):3, 2012.

[18] A. D. Monroe. Public opinion and public policy, 1980-1993. *Public Opinion Quarterly*, pages 6–28, 1998.

[19] D. J. Newman and S. Block. Probabilistic topic decomposition of an eighteenth-century american newspaper. *J. Am. Soc. Inf. Sci. Technol.*, 57(6):753–767, Apr. 2006.

[20] B. O’Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[21] B. I. Page and R. Y. Shapiro. Effects of public opinion on policy. *The American Political Science Review*, pages 175–190, 1983.

[22] M. J. Paul and M. Dredze. You are what you tweet: Analyzing twitter for public health. In *International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[23] M. J. Paul and M. Dredze. Experimenting with drugs (and topic models): Multi-dimensional exploration of recreational drug discussions. In *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, 2012.

[24] D. Ramage, S. T. Dumais, and D. J. Liebling. Characterizing microblogs with topic models. In *International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[25] E. T. K. Sang and J. Bos. Predicting the 2011 dutch senate election results with twitter. In *EACL Workshop on Semantic Analysis in Social Media*, pages 53–60, 2012.

[26] J. Schildkraut and T. C. Hernandez. Laws that bit the bullet: A review of legislative responses to school shootings. *American Journal of Criminal Justice*, 39(2):358–374, 2014.

[27] M. A. Stefanone, G. D. Saxton, M. J. Egnoto, W. Wei, and Y. Fu. Image attributes and diffusion via twitter: The case of guncontrol. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pages 1788–1797. IEEE, 2015.

[28] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[29] J. S. Vernick, S. P. Teret, K. A. Howard, M. D. Teret, and G. J. Wintemute. Public opinion polling on gun policy. *Health Affairs*, 12(4):198–208, 1993.

[30] P. Vijayaraghavan, I. Sysoev, S. Vosoughi, and D. Roy. Deepstance at semeval-2016 task 6: Detecting stance in tweets using character and word-level cnns. arXiv preprint arXiv:1606.05694, 2016.

[31] W. J. Vizzard. The current and future state of gun policy in the united states. *Journal of Criminal Law and Criminology*, 104(4):879 – 904, 2015.

[32] W. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, and V. Mudoch, editors, *Advances in Information Retrieval*, volume 6611 of *Lecture Notes in Computer Science*, pages 338–349. Springer Berlin Heidelberg, 2011.