Observing Election Incidents in the United States via Twitter: Does Who Observes Matter?*

Walter R. Mebane, Jr.† Patrick Wu‡ Logan Woods§ Joseph Klaver¶

Alejandro Pineda‖ Blake Miller**

March 27, 2018

*Prepared for presentation at the 2018 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 5–8, 2018. Previous versions were presented at the 2017 Annual Meeting of the Midwest Political Science Association, the 2016 Annual Meeting of the American Political Science Association and the 2017 Annual Meeting of the American Political Science Association. Thanks to Preston Due, Joseph Hansel, Adam Rauh and Barry Snyder for assistance. Work supported by NSF award SES 1523355.

†Professor, Department of Political Science and Department of Statistics, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).

‡Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: pywu@umich.edu).

§Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: ltwoods@umich.edu).

¶Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: jrklav@umich.edu).

‖Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: ajpineda@umich.edu).

**Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: blakeapm@umich.edu).

Abstract

Citizen complaints about election administration can help improve election performance and possibly help establish the legitimacy of an election. We use Twitter and automated machine classification methods to extract hundreds of thousands of observations of election incidents (e.g. long lines or success voting) by individuals all across the United States throughout the 2016 election. Here we focus on the general election. The mix of incidents varies by timing (e.g., weeks before election day versus on election day) and by features of each State (e.g., whether the state allows absentee or early voting), but do individuals with different characteristics report distinct types of incidents? We explore how people who exhibit differing features such as partisanship, ideology and group affiliations tend to report differently. We use information about Twitter friends and followers and about each individuals' Twitter behavior to measure such features as well as features such as the individuals' prominence and position in networks.

1 Introduction

Election forensics is the field devoted to using statistical methods to determine whether the results of an election accurately reflect the intentions of the electors. Most such methods analyze information about voter participation or voters' choices, looking statistically for patterns that suggest frauds occurred (e.g. Myagkov, Ordeshook and Shaikin 2009; Levin, Cohn, Ordeshook and Alvarez 2009; Mebane 2010; Pericchi and Torres 2011; Cantu and Saiegh 2011; Beber and Scacco 2012; Mebane 2014; Montgomery, Olivella, Potter and Crisp 2015; Mebane 2016; Rozenas 2017). It would be useful to draw other information into statistical analysis, both generally to enhance diagnosis of what happened in an election and more specifically to help address the primary challenge for election forensics: trying to tell whether patterns in election results that may appear anomalous in statistical estimates and tests are the results of election frauds or of strategic behavior (Mebane 2013, 2016).

Problems in elections that are not due to frauds may also stem from legal or administrative decisions. Long waiting times or crowded polling place conditions (Stewart and Ansolabehere 2015; Herron and Smith 2016), for example, are themselves concerns and may also produce distortions in turnout or vote choice data. As another example, the deployment of badly designed ballots (Lausen 2007; Quesenbery and Chen 2008) or defective election equipment (Herrnson, Niemi, Hanmer, Bederson, Conrad and Traugott 2008; Jones and Simons 2012) is inherently interesting and may also cause distortions in other election data. Another example is the number of polling stations opened for an election and where the polling stations are located (Brady and McNulty 2011).

Observing how individual people—voters or would-be voters—interact with such conditions is a challenge. In some countries systems for recording citizen complaints or the findings of observers are robust (e.g. Mebane and Wall 2015), but not for instance in the United States (Mebane, Pineda, Woods, Klaver, Wu and Miller 2017). Survey data cannot produce information with sufficient granularity to locate potential problems throughout an entire electoral system—at every polling station throughout the entirety of a multi-day election, for example. Either for further use in election forensics or because of their inherent interest as causes or consequences of political behavior, it can be useful to obtain observations that originate from ordinary individuals of how elections are administered and of how individuals respond to election circumstances.

We use data from Twitter to get information about American election administrative performance from individual observers throughout the country: the beginnings of a "Twitter Election Observatory." We describe election observations—extracted from Twitter—by individuals during the 2016 election from across the United States. While we do not address how to integrate these data in an election forensics analysis, we do show how various observed phenomena—such as individuals waiting in long lines or having difficulties in casting votes—are associated with state-level election procedures and demographic variables. We show that the types of incidents individuals Tweet about are related to attributes of the individuals. Deriving measures of individual positions in the Twitter network from "friend," "follower" and triad data, measures of partisan affiliation from the descriptions users provide of themselves and measures of participation in political movements based on other features of their Twitter data, we find that a variety of individual features are associated with the types of incident Tweets users create. To extract measures of partisan affiliations from user descriptions we use word2vec (Mikolov, Chen, Corrado and Dean 2013) to convert user `description` text into numerical vectors, using patterns of retweets in users' timelines to validate cosine similarities between the description vectors and the vectors associated with words like "republican" and "democrat." We also flag some users who probably are bots.

We describe an observation-focused scheme used with Tweets during the Fall general election. The system involves extracting Tweets using keyword filters, collecting information about election officials' and other leading actors' Twitter accounts, and classifying Tweets for relevance and for type of incident. For the classification tasks we apply active learning techniques with automated machine classification methods to Tweet texts, although both images and text associated with Tweets are important for classification decisions.

For the general election period we show that hundreds of thousands of incident observations can be recovered from Tweets gathered during the election period, observations that get at many different aspects of election performance. Incidents vary over states and over time, and they are associated with election administration features such as how early voting and absentee ballots are handled and with demographic features such as the racial composition and educational attainment of state populations. The distribution of incident Tweets also varies depending on individuals' attributes. For instance, "republican" users Tweet about significantly different types of incidents than "democrat" users do.

Monitors, Observers and Official Complaints in the United States: Another potential source of data to supplement forensic statistics is reports from election observers. Indeed election observation, particularly that performed by international monitors, has become a global norm and some evidence has shown that it can improve the quality of elections (Hyde 2011). Election observation can be conducted either by international or domestic groups (Bjornlund 2004). Such monitoring is far from perfect. There is little in the way of international standards for election observation missions and the nature of this fragmentation can lead to biases in monitoring practices (Kelley 2012). These missions are also frequently limited in scope and can simply displace fraudulent activities (Ichino and Schuendeln 2012).

While most monitoring is performed by international organizations, numerous countries possess domestic institutions that enable citizens or domestic political parties to file formal election disputes, essentially deputizing these groups into the role of informal election observers. Mebane, Klaver and Miller (2016) and Mebane and Wall (2015) use such data, respectively from German citizens and from Mexican parties. In Germany data come from citizen complaints about the federal election filed with a committee of the *Bundestag*, and in Mexico information comes from petitions parties filed to try to nullify the votes counted in particular ballot boxes. In both cases the auxiliary data facilitate seeing that election forensics statistics are responding to strategic behavior or to parties' tactical actions, as well as perhaps to frauds.

For several reasons it is difficult to obtain information about citizens' observations of election incidents in the United States. Election complaint processes in most states are convoluted and characterized by multiple possible channels for disputes, and they usually depend on particular election laws allegedly being violated. These channels may include submitting a complaint or dispute via an online portal, reporting an incident via phone, printing out a particular form and submitting a hard copy, or even simply emailing the relevant election authority. In many cases the process for filing a dispute is itself burdensome, leading to few complaints being submitted. For instance, all complaints submitted in compliance with the Help America Vote Act of 2002 must be notarized. Consequently very few complaints are submitted via this process. Few states make what complaint data exist from official channels publicly available. In Mebane et al. (2017) we detail the unavailability of official data about citizen complaints in the United States.

The impossibility of obtaining citizen observations of election incidents through such means for the United States prompts us to turn to social media. We find that voluminous observations can be gathered from Twitter. The biggest challenges with such data concern whether observations are reliable, whether the location of reported incidents can be determined and whether the observations we are able to collect accurately represent the full set of all incidents that occur.

2 Using Twitter to Capture Election Observations

We use Tweets to build data regarding election observations by individuals in the United States. We focus on the 2016 general election. For the general election period we collect Tweets continuously starting on October 1 and ending on November 8.

2.1 Collecting Twitter Data

For the general election period we used data from officials' timelines[1] along with data from the Twitter Streaming API (Twitter, Inc. 2016*b*,*a*). Keywords we used to select Tweets are shown in

[1]We obtained a list of election official, party and other Twitter accounts ("handles") (see Appendix section 4.1 for details regarding compiling the list). The proportion of county election offices that have an affiliated Twitter account varies greatly across states.

the note.[2] In all during October 1 through November 8, 2016, we downloaded 44,329 Tweets from timelines and 16,221,304 Tweets via the Streaming API. Removing retweets leaves 6,163,890 unique Tweets which contain 4,541,097 unique Tweet texts. Only 598,783 Tweets have `place` and `fullname` information (see Appendix 4.1), which is needed to be able to locate any incident observation reliably in geography, which means to place it in a state, city or neighorhood. Among these Tweets there are 505,112 unique Tweet texts. We drew a sample of 19,789 Tweet texts from this collection of 505,112 Tweets and labeled them by hand as containing an incident observation ($n = 2,610$) or not ($n = 17,179$). This is the initial sample of human-labeled Tweets we use to begin the active learning process described in section 2.2.1.

Table 1 reports the distribution of the initially sampled incident observations over states. "States" includes Puerto Rico (PR) and the United States Virgin Islands (VI). All states are covered, and in general but not always states that are larger in population have more Tweets with incident observations.

*** Table 1 about here ***

2.2 Categorizing Twitter Data

To determine whether a downloaded Tweet includes any relevant observations of the electoral process and then to say what types of incidents are being reported, we augment, clean and classify the Tweets.

We augment the text "content" of each Tweet in two ways. In general we get the resource, if any, located at each URL the content contains. If that resource contains any text, we capture that

[2]Twitter API Keywords: line to vote, long line to vote, wait to vote, absentee voting, early voting, problems voting, voting rights, right to vote, election fraud, corruption, voter fraud, stole election, stolen election, rigged, election stealing, tamper, manipulate, voter id, voter identification, election complaint, election problem, broken voting machine, election officials, electronic voting, election audit, election observer, poll watch, vote protection, election protection, disenfranchised, campaign finance, election system, primary election, general election, voter complaint, polling place, registration database, statevote, votestate, stateelection, vote count, vote tabulation, voter database, voter registration, voter suppression, voting machine, voting machine hacked, vote not counted, vote, US election, American election, not enough ballots, absentee ballot, voter intimidation, voter harassment, mail in ballot, vote by mail, voter hotline, waiting to vote, precinct, precinct officials, precinct captain, replacement ballot, ballot selfie, my ballot, my vote, eleccion, fila para votar, derecho al voto, derecho al votar, fraude electoral, maquina de votacion, funcionarios electorales, colegio electoral, neo-nazi, white supremacist, white nationalist, alt-right.

text and append it to the original content.[3] If that resource contains an image, we capture the image's URL.[4] Human coders examine any images associated with a Tweet when labeling it, but currently the machine learning algorithms we use use only the augmented text. Images often decisively affect human coders' judgments regarding any information Tweets may contain—e.g. an image of a person wearing an "I Voted" sticker or an image of many people in line at a polling place—but the machine classification algorithm currently does not have access to images or descriptions of images.

Cleaning the augmented Tweet content involves removing nonprintable characters, stray HTML codes, internal quotation marks and the '*' character. For the version of the contents used in machine classification and active learning processes, we also removed URLs and made some frequently occurring text strings generic instead of specific to each state. The latter changes replaced some state-specific strings with strings like "#XXvotes," "#XXprimary," "#XXcaucus" and "#XXvoterfraud," where "XX" originally was the postal code abbreviation for a state. We did this to enhance the comparability of Tweets across states for the machine classification algorithm.

To determine whether each downloaded Tweet includes relevant observations, we began by using humans[5] to examine the raw Tweets directly. A Tweet that contains relevant observations about electoral processes is coded to be a "hit." Each "hit" was also classified into one or more categories. Classification rules originate with the categories in Mebane, Klaver and Miller (2016) and the Election Incident Reporting System (EIRS) (Verified Voting Foundation 2005; Hall 2005; Johnson 2005) with modifications to refer to all observed incidents without emphasizing "complaint" observations (see Appendix section 4.4).

The procedure we developed for humans to use when making "hit" determinations for the general election data is shown in Figures 1, 2 and 3. The background for these flowcharts is discussed in Appendix section 4.3. The coding rules for categorizing the incidents to which "hits"

[3]Specifically, we capture any text in the `og:description` field in the resource's HTML code. For general election type-of-incident classification we also append to the text the date (month, day and year) and `place$fullname` of the Tweet.

[4]Specifically, we capture any URL in the `og:image` field in the resource's HTML code.

[5]The human coders were subsets of this paper's authors and two undergraduate assistants.

refer are described in Appendix section 4.4 (for general election Tweets).

*** Figures 1, 2 and 3 about here ***

As detailed in Appendix section 4.4, for coding general election Tweets there are twelve main categories: Outside Influence; Disability/Accessibility Issue; Line Length, Waiting Time, Polling Place Crowding; Polling Place Event; Electoral System; Absentee, Mail-In, or Provisional Ballot Issue; Election Official; Voter Identification; Registration; Voter Fraud; Ballot and Voting Technology; Unspecified Other. For most of these categories we also record which "adjective" modifies the incident. For example, for the Line Length, Waiting Time, Polling Place Crowding category adjectives distinguish no lines from short lines from long lines. See Appendix section 4.4 for details regarding the definition of these adjectives. Many adjectives reflect judgments about things working well or poorly, but our coding scheme does not depend on and is not intended to measure any kind of sentiment. For example, many express warm feelings when encountering a very long line to vote: we record the long line and ignore how the person Tweeting said they felt about it.

2.2.1 Active Learning

To produce a training set to use to start active learning with the general election Tweets, we used a sample of 19,179 Tweets from the streaming API. For a description of how the sample was drawn see Appendix section 4.2. The "hits" in this sample were initially produced by several coders but then all were checked by one pair of coders working in tandem.

To grow the initially small training sets we use active learning, an iterative supervised machine learning technique (Settles 2010). Active learning allows us to build training sets with fewer labeled observations and a good balance between classes, which is useful because of the scarcity of the some classes (Miller and Klaver 2016). This framework uses uncertainty sampling to identify observations that we should label by hand to provide the most useful new input to the next iteration of the classifier. At each iteration, we train a support vector machine (SVM) on labeled Tweet texts. We use the distance from the SVM's separating hyperplane to measure model uncertainty. We iteratively label the texts closest to the hyperplane and refit a model until acceptable average precision, recall and F-measure are achieved.

2.2.2 Classification

For both the active learning SVM and the algorithms we use for the final classification step[6] we first preprocess each Tweet's augmented text. This involves removal of all duplicate texts. We do not use stemming nor remove stop words. For classification we use a word n-gram model for the preprocessed text and a character n-gram model for hashtags to convert the Tweet corpus into a document term matrix.[7] Each row of the matrix represents a Tweet and each column represents a unique character or word n-gram. Cell values represent the count of each n-gram in the document. Finally we do a TF-IDF transformation of the raw count matrix (Leopold and Kindermann 2002; Lan, Tan, Low and Sung 2005). Because the feature space is high dimensional, and we want to avoid overfitting, we select features using the coefficients of a linear SVM with ℓ_1 norm penalty. Features with SVM coefficients lower than the mean of all coefficients are discarded (Rakotomamonjy 2003). For the final classification step we use a randomized search to select parameters for the various algorithms.[8]

For general election "hit" labeling humans manually labeled 5,224 Tweet texts with `place` information selected in active learning, for a total of 25,013 human-labeled Tweets. Among the human-labeled texts, 3,689 are "hits" and 21,324 are "not-hits." Over all unique Tweet texts with `place` information we classify 40,687 texts as "hits" and 464,425 as "not-hits." Over all unique Tweet texts with or without `place` information we classify 315,180 texts as "hits" and 4,225,917 as "not-hits."[9] Classification performance measures (Table 2) for the set of Tweets that have

[6] The classification algorithms we use from `sklearn` (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot and Duchesnay 2011) are `linear_model.LogisticRegression`, `naive_bayes.MultinomialNB` and `svm.LinearSVC` as estimators in `ensemble.VotingClassifier`.

[7] We allow up to 5-grams for words and 2-, 3- and 4-grams for characters in hashtags.

[8] To execute the search we use `RandomizedSearchCV` from `sklearn` (Pedregosa et al. 2011).

[9] Before classifying all 4,541,097 Tweet texts regardless of whether a Tweet has `place` information, we use active learning to human-label an additional 100 Tweets from the pooled corpus of all 4,541,097 Tweet texts.

place information show that overall we achieve average precision, recall and F-measure of .88, .89 and .88, respectively.[10]

*** Table 2 about here ***

To determine what type of incident is represented by each of the 40,687 general election Tweets texts with place information that are classified as "hits," we begin by manually labeling a random sample of 1,419 of the texts then augment the initial sample using binarized active learning. While each Tweet may mention several types of incidents, the distribution of individual types of incidents in this initial sample is shown in Table 2. A few types are scant, and some possible "adjectives" do not occur in the initial sample. To try to boost a few of the type frequencies before beginning machine-assisted sampling, we hand-labeled a few Tweets located by doing keyword searches in the set of 40,687 Tweet texts.[11]

*** Table 3 about here ***

For binarized active learning we use the SVM approach we used for "hits" for each type and each type adjective separately. For instance one step of the process includes Tweet texts in the sample for human labeling if they are near the separating hyperplane for the "Polling Place Event" incident versus all other types of incidents. Samples are weighted using the inverse relative frequency of occurrence among the human-labeled texts, so that texts that are uncertain members of less frequent classes are sampled more frequently. Types or adjectives that occur too infrequently are not used to determine sampling, although labels for these too-scarce classes may still be assigned by human coders. Table 4(a) shows F-measure classification performance statistics for each class used to determine sampling, as assessed at the end of the active learning

[10]Results for the larger set of Tweets, which includes 100 more human-labeled Tweets, are nearly the same. Classification performance is assessed as similar when done both without stemming and with stemming. Indeed, every Tweet with place information is classified identically in both cases, even though algorithm parameters vary when stemming is enabled. For instance, without stemming the randomized search finds for words it is best to use up to 3-grams while with stemming it is best to use up to 5-grams.

[11]In particular we searched for the strings "disabl," "handicap," "technology" and "electronic." By this method we added 18 type 2 incidents and ten type 11 incidents, along with a scattering of incidents of other types. We did not label as "not hits" Tweets we located through these keyword searches that did not actually report an incident.

process for the Tweets that have `place` information. By the end of active learning there are 4,018 human-labeled Tweet texts.

*** Table 4 about here ***

For both the set of Tweets that have `place` information and the larger set of Tweets, we use a binarized approach with the ensemble classifier for final classifications.[12] We predict classes only for those classes that have a reasonably large set of human-labeled instances. Table 4(b) shows F-measure classification performance statistics for each such class.

2.3 Observing User Characteristics

We use the Twitter API to collect information about characteristics of the users who posted the Tweets we classify as general election incidents. In all 215,230 distinct users posted the 315,180 Tweets we classify as referring to an incident. From the Tweets originally obtained via the streaming API we extract the texts of self-descriptions users present.[13] For each user who Tweeted an incident we attempted to obtain IDs for up to 10,000 of the user's Twitter "friends" (other users whom the user follows) and "followers" (other users who follow the user).[14] We obtained "friend" IDs for 195,879 users and "follower" IDs for 196,739 users. For each user who Tweeted an incident we also attempted to obtain the user's timeline of Tweets.[15] We obtained timelines for 197,366 users: Figure 4 shows the distribution of counts of Tweets in the timelines. The maximum number of Tweets we obtained in the timeline for a user is 3,399, and over all timelines there are 527,961,969 Tweets. We noted which users enclosed their `name`[16] in three

[12] For details about the classifier see note 6.

[13] For each user we use the user `description`—taken from the `user` object in the original JSON object—associated with the chronologically first Tweet the user posted that we classify as referring to an incident.

[14] These are the IDs returned by using the `get_friends_ids` and `get_followers_ids` functions of Twython (McGrath 2016). We executed this process early in 2018. We started getting friend and follower IDs on January 15, 2018. Follower ID acquisition completed on February 14, 2018, and friend ID acquisition completed on March 5, 2018. Some Twitter accounts were inaccessible.

[15] These are the Tweets returned by using the `get_user_timeline` function of Twython (McGrath 2016). We executed this process early in 2018. We started getting timelines on January 24, 2018, and acquisition completed on February 19, 2018. Some timelines were inaccessible.

[16] `name` taken from the `user` object in the original JSON object.

pairs of parenthesis marks: such a use of “echoes” might indicate a user’s participation in a protest movement against antisemitism (Williams 2016).

*** Figure 4 about here ***

We obtained information about which users may be bots or otherwise artificial from several sources. We checked our incident-tweeting users against a list of Russian bots provided by Democrats on the House Intelligence Committee (Select Committee Democrats 2017): 17 incident-tweeting users are on the list; these users produced 41 Tweets that are classified as referring to an incident. We checked against the bots listed at `probabot` (Quartz 2017; Collins 2017): 22 of the incident-tweeting users are on this list; these users produced 302 Tweets that are classified as referring to an incident. Also we note that 466 incident-tweeting users produced 2,385 incident Tweets that are not formally retweets but instead internal or “quoted” retweets. Last we note that 812 users produced 2,502 incident Tweets that concern entertainment activities and are deemed to be irrelevant.[17]

2.3.1 Network features: triads

To represent aspects of the networks incident-Tweeting users are involved in, we define triads of these users using the friend and follower ID information we obtained. Triads are important features that relate to the persistence and dynamics of social networks and to the propagation of information through networks (Krackhardt and Handcock 2007; Lazer, Rubineau, Chetkovich, Katz and Neblo 2010). Our use of triads barely skims the surface of what they can reveal about networks involving our users.

We restrict attention to relations entirely among incident-Tweeting users. We consider two kinds of triads: user-friend-friend triads and user-follower-follower triads. That is, for the

[17] We noticed these irrelevant Tweets because they include many repetitions of nearly identical text. Two examples are “Vote 4 @TrinityRband & help make the 10/23 Chart! Voting ends 10/22 4pmET https://t.co/cv0J3wcCjB #countrymusic” and “Vote 4 @thejblundell & help make the 10/16 Chart! Voting ends 10/15 4pmET https://t.co/B4FuuHNTnA #countrymusic.” Tweets that included the following hashtags we marked “irrelevant”: ariasbeyonce, ariasjustinbieber, countrymusic, demilovato, emabiggestfansarianagrande, emabiggestfansjustinbieber, emabiggestfansladygaga, emabiggestfansshawnmendes, entertaineroftheyear, mtvmama2016, 2016mama_vote, streamys, heismanhouse.

user-friend-friend triads we take all of a user's "friends" that also sent a Tweet referring to an incident and define triads by adjoining all of the "friends" of those "friends" who also sent an incident-related Tweet. Likewise for "followers." Notwithstanding the importance of closed triads (Lazer et al. 2010), we do not include cyclical triads or check closure: all members of each triad are distinct users, and we do not require that the second "friend" ("follower") have the initial user as a "friend" ("follower"). In the future we will consider additional types of and restrictions on triads.

The numbers of these kinds of triads among incident-Tweeting users is large. We find 21,243,891,283 user-friend-friend triads and 3,057,780,205 user-follower-follower triads. The distributions of triad counts across users are displayed in Figure 5. Apart from spikes in the count of users who are involved in zero triads—this includes the roughly 20,000 users for whom we failed to obtain "friend" or "follower" IDs—the per user triad counts have approximately log-Normal distributions.

*** Figure 5 about here ***

While there are large numbers of these triads, the number of friends in the sense that a user directed Tweets to them (Huberman, Romero and Wu 2009) is small, at least if we restrict attention to the 315,180 incident Tweets. We identified the instances when one incident-Tweeting user included the `screen_name` (preceded by '@') of another incident-Tweeting user in the text of a Tweet. 159,804 of the 215,230 users never sent such a Tweet, another 43,538 directed Tweets to exactly one other user and 11,764 users directed Tweets to between two and ten users. A scant 124 users directed Tweets to more than ten users, up to the one user who directed Tweets to 329 other users. "Friends" in the sense defined by Huberman, Romero and Wu (2009) are sparse in our set of 315,180 incident Tweets. The large number of triads we identify among Twitter "friends" and "followers" should not be interpreted as suggesting our data include many "actual friends" (Huberman, Romero and Wu 2009, 6).

2.3.2 User features from user descriptions and timelines

To estimate political stances for our users, we focus on their user description, which is a short biography that is (optionally) provided by each Twitter user.[18] The idea is that some user descriptions will describe their political stance explicitly (e.g. Trump supporter, Clinton supporter, "lifelong Democrat", etc.), while other user descriptions will provide incomplete or no political stance information. Both users that explicitly describe their political stance and users that do not, however, may also list other characteristics such as occupation, city, hobbies, etc. The idea is to recognize patterns in the words surrounding explicit political stances and to use that to estimate each user's political stance, even if the user does not explicitly describe their political stance.

We do this by employing Word2vec, a methodology that produces word embeddings (Mikolov et al. 2013). In technical terms, Word2vec is a shallow two-layer neural network that is trained on a "fake" task of reconstructing the linguistic contexts of inputted words (Mikolov et al. 2013). Word2vec assumes that the meaning of a word is best determined by its linguistic context; in other words, it assumes that two words that come from very similar linguistic contexts mean similar things. The word embeddings are drawn from the hidden layer of the neural network. Each word is a p-dimensional vector positioned such that word embeddings that are close to each other, in a cosine similarity sense, come from similar linguistic contexts (and are thus assumed to mean similar things). Likewise word embeddings that are far apart from each other, again in a cosine similarity sense, come from dissimilar linguistic contexts (and are thus assumed to mean different things).

To estimate some political stances for our users, we employ a novel methodology of estimating partisan affiliation based on Word2vec word embeddings. We first obtained word

[18] We considered attempting to match voter files to Twitter accounts, but decided this was not feasible for numerous reasons. First Twitter accounts do not include the identifying information such as birth dates and registration address that facilitates matching between files. Without that identifying information it is impossible to match Twitter accounts to voter files. For example if a Twitter user listed her name as "Elaine Smith" and her location as "Cleveland," we have no way of identifying which of the numerous "Elaine Smiths" appearing in the Cuyahoga County voter file is the correct match. A match using only names that appear once in the voter file would not be a random subset of Twitter users or voters (the data would likely be biased towards counties with low populations). Finally many individuals on Twitter do not use their legal name as their profile or user name, and so even attempting matches for those individuals is futile.

embeddings across all words in all user descriptions among users who made some observation about the election. Of these 215,247 users, 195,943 users filled out something in their user description. We then calculated a weighted average of all the word embeddings corresponding with each word in a user's description. Each term's frequency-inverse document frequency statistic was used as the weight in the averaging process. We then compared each user's description, which is a weighted average of its word embeddings, with specific word embeddings by calculating its cosine similarity, which is defined between vectors $A$and B as

$$\cos(A, B) = \frac{A \cdot B}{||A||||B||}. \tag{1}$$

Cosine similarity is a value between -1 and 1, with 1 meaning that the vectors line up exactly, 0 meaning the vectors are orthogonal, and -1 meaning that the vectors are facing exactly away from each other.

We calculated cosine similarities between the user's descriptions and the word embeddings for "trump", "clinton", "donald", "hillari" (the stemmed version of "Hillary"), "democrat", "republican", "realdonaldtrump" (Donald Trump's Twitter handle), "hillaryclinton" (Hillary Clinton's Twitter handle), "maga" (Donald Trump's campaign slogan), and "strongertogeth" (Hillary Clinton's campaign slogan, stemmed). These cosine similarities estimate how similar a user's description is to these specific word embeddings, the idea being that a user who is more Democrat will have higher positive cosine similarity with Democratic-associated word embeddings such as "clinton", "hillari", "democrat", etc and have negative cosine similarity with Republican-associated word embeddings. Thus, each user has ten cosine similarity values that estimate how closely their user `description` aligns with these political terms of interest.

To verify that these cosine similarities are measuring relevant political stances, we use the connections users exhibit to external political communication networks to validate that these cosine similarity coefficients relate to partisan affiliation. We show that the cosine similarities are associated with differences in users' tendencies to retweet Tweets from prominent "ideological"

Twitter accounts.

For a sample of 10,000 incident-Tweeting users, we count how often the timelines of each user contain retweets of posts from the "influential" liberal and conservative Twitter acounts listed by three sources (Joyce 2016*a*,*b*; Faris, Roberts, Etling, Bourassa, Zuckerman and Benkler 2017).[19] Figure 6 shows the distribution of these retweet counts, on a logarithmic scale.[20] The spikes of values at zero reflect that most users do not retweet the "influential" accounts: 4,757 users do not retweet a liberal account, and 9,762 users do not retweet a conservative account. The highest number of liberal retweets in the sample of 10,000 users is 804, and the highest number of conservative retweets is 1,089.

*** Figure 6 about here ***

Figures 7, 8 and 9 convey a graphical impression of the distribution of cosine similarity coefficients and of their association with the retweet counts. The displayed coefficients are those based on the words "democrat" and "republican."[21] Each dot in the scatterplot corresponds to a user in the sample of 10,000 users. Dots are colored to indicate user retweeting behavior as assessed from user timelines: red, more conservative retweets; blue, more liberal retweets; black, no conservative or liberal retweets. Clearly the cosine similarities for "democrat" are positively associated with the cosine similarity coefficients for "republican." The retweet-indicating colors show more red dots near the top of the distribution and blue dots near the bottom of the distribution in the upper-right portion of the scatterplot. Inspecting the full descriptions for the users—examples of these descriptions appear in the three plots—shows that users with positive values of the cosine similarity coefficients provide more partisan and politically engaged descriptions while users with negative values provide nonpartisan and even nonpolitical descriptions. Similar patterns appear when the pairs "clinton-trump," "hillari-donald, "hillaryclinton-realdonaldtrump" and "strongertogeth-maga" are considered.

[19] We use 78 conservative accounts and 55 liberal accounts.

[20] Logarithms are natural logarithms.

[21] We impute coefficients values of zero for users for whom the user `description` is empty so that the cosine similarity cannot be computed.

*** Figures 7, 8 and 9 about here ***

Analysis using zero-inflated negative binomial regression models (Zeileis, Kleiber and Jackman 2008) for the 10,000 sampled users shows that the cosine similarities are consistently and significantly associated with ideologically tinged retweeting behavior. Tables 5 and 6 each show results for five models in which the count part of the model includes pairs of the cosine similarity variables as regressors, while the zero-inflation part of the model includes these variables plus variables that use logarithms of the user-friend-friend and user-follower-follower triad counts and of the `friends_counts` and `followers_counts` taken directly from the `user` object in the initial incident-referring Tweet for each user (Zeileis, Kleiber and Jackman 2008; Jackman 2017). The zero-inflation parts of the models show that most of the cosine similarity, triad, `friends_counts` and `followers_counts` variables are associated with the propensity to retweet "influential" accounts, but more directly relevant for the validity of the cosine similarities as measures of political stances is the fact that the similarity coefficients for "republican," "trump," "donald, "realdonaldtrump" and "maga" are negatively associated with liberal retweets and positively associated with conservative retweets, while "democrat," "clinton," "hillari, "hillaryclinton and "strongertogeth" have the opposite pattern of association.

*** Tables 5 and 6 about here ***

On the reasonable assumption that most frequent retweeters are sympathetic to those they retweet, such associations testify to the cosine similarities being good measures of partisan affiliations. Such affiliations are manifest in users' Twitter descriptions, and our cosine similarity method using Word2vec helps to discover them.

2.4 Characteristics of General Election Tweet Contents and Incidents

Incidents occur in every state in the general election period. As Table 7(a) shows, among the Tweets that have `place` information, the highest count of Tweet texts that are labeled or classified as incident observations occur in California, Texas, Florida and New York and the smallest in

Wyoming, North Dakota, South Dakota and Montana.[22] Table 7(b) shows these same states have the largest and smallest counts of incidents among the larger set of all Tweets:[23] Hawaii has fewer incident-observing Tweet texts than does Montana.

*** Table 7 about here ***

The rate of incidents in the sense of incidents per person is not the same across states. To adjust the counts of "hits" for the populations of the various states, Table 8 shows the distributions in terms of observations per million persons in each state. In both the set of Tweets that have `place` information and in the larger set of Tweets, on a per capita basis the District of Columbia stands out with the highest rate followed by Nevada and North Carolina. Wyoming is lowest.

*** Table 8 about here ***

Plotting incident observations by day shows that the most observations occur on election day. Figure 10(a) uses the 40,678 Tweets that have `place` information and Figure 10(b) uses all 315,180 Tweets either with or without `place` information to display histograms of the number of classified "hit" Tweets on each day during October 1 through November 8, 2016.[24] Both histograms show the same pattern of variation over days. The similarity between the histograms provides some evidence that the set of incidents is similar regardless of whether the `place` identifying option had been enabled by the Twitter user.

*** Figure 10 about here ***

[22] For 255 of the Tweets with `place` information that information neither allowed the state to be identified nor indicated the Tweet did not originate in the United States. For all but 65 of these Tweets we used `location` information to identify the state. The `location` information places six of these 65 Tweets outside the United States, eight in "United States," two in one of three states (e.g., "DC MD VA #DMV"), and the rest have information that is geographically uninterpretable.

[23] For Tweets that lack `place` we attempted to recover state locations from `location` information. The `location` information describes the user and is written by the user, so the entries are idiosyncratic. Even if the `location` describes a real geographic location, that location is not necessarily the place from which the Tweet was sent.

[24] The last bar on the right in the histograms in Figure 10 corresponds to November 9, which is the date associated with some Tweets due to our expressing all times in Eastern Standard Time units.

2.4.1 Types of Incidents

A breakdown of all 315,180 incident Tweets by type of incident—see Table 9—shows that "Absentee, Mail-In, or Provisional Ballot Issue" is the most frequent type of incident of interest, followed by "Polling Place Event." The second most frequently occurring category is actually "Not an Incident," but the frequency of such occurrences is of technical interest only: the classifier does better weeding out these genuine non-incidents when sorting among "hits" than when trying to find "hits" among all the other Tweets. "Electoral System" and "Registration" type incidents are next most frequent, and least frequent[25] is "Line Length, Waiting Time, Polling Place Crowding." Considering the adjectives we see that Tweets about success vastly outnumber Tweets about problems for the "Absentee, Mail-In, or Provisional Ballot Issue" and "Polling Place Event" types, while for "Registration" Tweets mentioning problems outnumber Tweets mentioning successes. For "Line Length, Waiting Time, Polling Place Crowding" Tweets about long lines outnumber Tweets about short lines or no lines. A high proportion of the "Absentee, Mail-In, or Provisional Ballot Issue" and "Polling Place Event" success Tweets are proclamations that "I Voted!": many of these Tweets are accompanied by images of people proudly displaying stickers that say variations of "I voted" or "I voted early."

*** Table 9 about here ***

Figures 11 and 12 show the distributions over time of incident observations by type. Figures 11(b) and 12(b) combine the "Absentee, Mail-In, or Provisional Ballot Issue" and "Polling Place Event" types: a Tweet about success via either modality counts as success for the combined type. The highest single day of voting incidents is election day, although as Table 9 indicates there are cumulatively more voting incident Tweets during the earlier period. Long lines or waiting times to vote are most frequent on election day, although again more such incident Tweets occur during the earlier period. Hundreds on election day also observe that lines or waiting times are not very long (Figures 11(a) and 12(a)). Reports of success with voter registration are slightly more

[25] That is, least frequent among incident types that occurred sufficiently frequently in our active learning sample that we attempted binarized classification.

frequent than reports of problems with voter registration in early October, a pattern that is reversed by election day (more Figure 12(d) than Figure 11(d)). For most of the period after October 1 praise of aspects of the election system is more frequent than reports of problems, although by election day the number of problems mentioned is nearly on par with the number of mentions of correct electoral system functioning (Figures 11(c) and 12(c)).

*** Figures 11 and 12 about here ***

2.4.2 User Attributes

We'd like to know whether people with different attributes report different types of incidents. To capture differences among people we have some measures that are individualized, like the "friend," "follower" and triad counts and the cosine similarities, and some measures that relate to collections of users, such as features of U.S. states. We build to a specification that considers all these types of measures together.

States: We start by illustrating that features of states can be important. Bivariate regression analyses show the type of incident observations depend on several variables, including variables that describe aspects of election administration in each state: whether a state requires some form of photo or non-photo identification ("Voter ID"); whether a state allows no excuse absentee voting ("No Excuse Absentee"); whether a state allows early voting or in-person absentee voting ("EV+In-person Abs."); whether a state has a complaint process outside of Help America Vote Act (HAVA); and whether there is at least one way (HAVA, non-HAVA, online portal) for voters in a state to submit complaints online. The type of incident also depends on a state's general-election turnout—measured in terms of the voting-eligible-population (VEP). State demographic variables such as race, ethnicity and educational attainment also relate to the type of incident.

Table 10 reports regressions that illustrate a few of these associations. Outcome variables are formed from the adjectives that describe three types of incidents: Line Length, Waiting Time; Polling Place Event (denoted "Voting"); and Absentee or Early Voting Issue. Levels of each adjective are associated with the numbers 0, 1 and 2: the value 2 represents a very long line (for Line Length), successful polling place operations or voting (Voting), or successful absentee or early voting operations (Absentee). In the regressions each type-of-incident variable is divided by the state population, so relationships concern the rate of incident reporting.[26] The table shows three models that include the Voter ID variable in interaction with three process variables: whether a state allows early voting ("Early Voting"); EV+In-person Abs.; and No Excuse Absentee. In all three cases the coefficients for Voter ID and for the other process variable have significant positive signs while the interaction has a significant negative sign. The fourth model in Table 10 includes the proportion White and the proportion with at least a bachelor's degree plus the interaction between these two variables. The proportion White and the proportion with at least a bachelor's degree each has a positive coefficient and the interaction has a negative coefficient. This means that, for instance, lines/wait-times are said to be shorter in states with high proportions of both whites and college graduates but otherwise longer.

*** Table 10 about here ***

Associations like these are hard to interpret, not least because states differ in so many more ways than those measured by these variables. Also Tweeting behavior is diverse among individual users within states.

"Friends," "Followers" and Triads: We measure counts of "friends, of "followers" and of user-friend-friend and user-follower-follower triads to try to get a sense of how involved users are in Twitter social networks. Distributions of Tweets by type of incident are strongly associated with the "friends," "followers" and triads counts.

Figure 5 shows that the distributions of triad counts are roughly log-Normal with sets of zero counts added. Figures 13(a,b) show that when the sets of "friends" and "followers" counts are restricted to the users who sent an incident-related Tweet, their distributions are similarly log-Normal plus sets of zero counts. We also consider triads-per-friend and triads-per-follower

[26] Most covariates also relate to the unadjusted counts.

ratios: let $T_{\mathrm{fr}i}$ denote the count of user-friend-friend triads for user i, let $T_{\mathrm{fo}i}$ denote the count of user-follower-follower triads for user i, and let $M_{\mathrm{fr}i}$ and $M_{\mathrm{fo}i}$ respectively denote the numbers of incident-Tweeting "friends" and "followers" user i has; the friend-triad ratio is

$$R_{\mathrm{fr}i} = \log\left(\frac{1 + T_{\mathrm{fr}i}}{1 + M_{\mathrm{fr}i}}\right)$$

and the follower-triad ratio is

$$R_{\mathrm{fo}i} = \log\left(\frac{1 + T_{\mathrm{fo}i}}{1 + M_{\mathrm{fo}i}}\right).$$

These so-to-speak triad density measures should discriminate users. The distribution of the friend-triad ratio $R_{\mathrm{fr}i}$ (Figure 13(c)) is bimodal while the distribution of the follower-triad ratio $R_{\mathrm{fo}i}$ (Figure 13(d)) is unimodal.

*** Figure 13 about here ***

Nonparametric regressions (Bowman and Azzalini 1997, 2003) of Tweet-type dummy variables on logarithms of the triad and "friend" or "follower" counts shows complex patterns of conditional association. Figures 14 and 15 each shows contour plots of regressions for six incident-type outcome variables: (a) long line[27]; (b) no or short line[28]; (c) problem voting[29]; (d) neutral or successful voting[30]; (e) problem registering[31]; (f) neutral or successful registration[32]. Figure 14 has regressors $\log(1 + T_{\mathrm{fr}i})$ and $\log(1 + M_{\mathrm{fr}i})$, and Figure15 has regressors $\log(1 + T_{\mathrm{fo}i})$ and $\log(1 + M_{\mathrm{fo}i})$. All of the regressions exhibit significant associations, and suffice it to say that none of the regression relationships are linear.

*** Figures 14 and 15 about here ***

[27] "Line Length, Waiting Time, Polling Place Crowding" adjective 2.

[28] "Line Length, Waiting Time, Polling Place Crowding" adjectives 0 and 1.

[29] Combining "Absentee, Mail-In, or Provisional Ballot Issue" and "Polling Place Event" adjective 0 or either and not adjectives 1 or 2 on either.

[30] Combining "Absentee, Mail-In, or Provisional Ballot Issue" and "Polling Place Event" adjective 1 or 2 on either.

[31] "Registration" adjective 0.

[32] "Registration" adjective 1 or 2.

We do not attempt to interpret the relationships shown in Figures 14 and 15, but we try to take them into account when assessing the relationships between incident-type variables and other attributes of individual users. To do so we approximate the nonparametric specifications with polynomial surfaces of the form

$$S_{\text{fr}}(p) \equiv \text{poly}(\log(1 + M_{\text{fr}i}), p) * \text{poly}(R_{\text{fr}i}, p) \tag{2}$$

where poly(X, p) denotes a set of orthogonal polynomials for variable X of orders zero to p, and $*$ denotes an interaction that additively includes the two base terms plus every possible crossproduct term. We use analogous polynomials $S_{\text{fo}}(p)$ defined using "follower" variables. Each term in the polynomial object has a coefficient that is estimated. In logit regressions for the type indicator variables we use AIC to choose the value for p for each variable, considering values of $p \in \{1, \ldots, 7\}$.[33] The same values chosen for p for each type-indicator variable are used, respectively for $S_{\text{fr}}(p)$ and $S_{\text{fo}}(p)$, regardless of what other regressors are included in the model. Visual inspection of the surface of predicted values shows that the polynomial models are a good match for the nonparametric results shown in Figures 14 and 15.

User Descriptions: We know that some features of users' descriptions are associated with differences in aspects of the distributions of incident types. For instance, we can flag all users whose descriptions include the keywords (ignoring case) "trump," "pence," "maga" or "make america great again" (excluding "NeverTrump") as "Trump" users and all users whose descriptions include "Hillary," "Clinton," "Kaine," "m with her" or "ImWithHer" (excluding "NeverHillary") as "Clinton" users. Figure 16 shows the effect that considering only Tweets sent by such users has on the distributions over time of line length and voting types of incidents. Figures 16(a,b) can be compared to Figure 12(a), and Figures 16(c,d) can be compared to Figure 12(b). The ratios between the count of incidents on election day and the counts of earlier incidents

[33] To estimate the logit regression models we use glm(·, family="binomial") in **R** (R Development Core Team 2011). poly() is the `poly()` function that is built into **R**, and $*$ is the interaction operator defined as part of **R**'s formula language.

are much smaller for both "Trump" and "Clinton" users (Figure 16) than they are for users overall (Figure 12). Moreover the distributions for "Trump" users in Figure 16 appear visually to differ from the distributions for "Clinton" users in Figure 16. Unfortunately the numbers of these "Trump" or "Clinton" users as identified directly using keywords in their descriptions is small. Analysis not shown here shows that significant differences can nonetheless be identified between such "Trump" and "Clinton" users—as well as between keyword-defined "Republican" and "Democrat" users or between keyword-defined "left" and "right" users—but it is better to consider the attributes measured using the cosine similarities to chosen words. The cosine similarities, which are derived from user descriptions, are meaningfully defined for all users.

*** Figure 16 about here ***

Using logit regression we regress six Tweet-type indicator variables on pairs of the cosine similarity variables, on dummy variables for the U.S. states,[34] on polynomial functions $S_{\text{fr}}(p)$ and $S_{\text{fo}}(p)$, and on a few other variables. The additional variables are (a) `echo parentheses`, which indicates whether a user's `name` is enclosed in triple parentheses, (b) `looks-like-a-bot`, which indicates whether a user is on one of the lists of bots or bot-like activities and (c) `irrelevent`, which indicates that a Tweet includes one of the hashtags that relates to entertainment activities. Tables 11 and 12 reports regression coefficients for five sets of regressions, one set for each of five pairs of cosine similarities. Table 11 has regressions for the pairs (a) `republican` and `democrat`, (b) `trump` and `clinton`, and (c) `donald` and `hillary`,[35] while Table 12 has regressions for pairs (d) `realdonaldtrump` and `hillaryclinton` and (e) `maga` and `strongertogether`.[36] In Tables 11 and 12 coefficients for the state dummy variables and for the polynomial terms are not shown.

*** Tables 11 and 12 about here ***

The regression results show that users with high values for the `republican`, `trump`, `donald`, `realdonaldtrump` or `maga` cosine similarities Tweet about significantly different types of

[34]See section 4.5 in the Appendix.
[35]The `hillary` variable corresponds to the "hillari" word for the Word2vec procedure.
[36]The `strongertogether` variable corresponds to the "strongertogeth" word for the Word2vec procedure.

incidents than do users who have high similarities for `democrat`, `clinton`, `hillary`, `hillaryclinton` or `strongertogether`. In Table 11, for instance, `republican` is negatively associated with Tweeting about a long line, about no line or a short line, about neutral or successful voting and about neutral or successful registration, while `democrat` is positively associated with Tweeting about a long line, about neutral or successful voting and about neutral or successful registration. `democrat` is not associated significantly with Tweeting about no line or a short line. `republican` is positively associated with Tweeting about a problem voting and about a problem registering, while `democrat` is negatively associated with those two types. For all of the six types of incidents, `republican` and `democrat` differ significantly in their propensity to Tweet about that type of incident. Significant differences also occur between the variables in each of the other pairs of cosine similarity variables, although not necessarily for all the types of incidents examined here. The exceptional instances where there is no significant difference are `trump` and `clinton` for problem registering, `donald` and `hillary` for problem registering and for neutral or successful registration, and `maga` and `strongertogether` for no or short line, for neutral or successful voting and for problem registering. In most instances when the types of users exhibit significant differences, the coefficient point estimates in the regression model have opposite signs.

The other marker for participation in a kind of sociopolitical activism also is associated with significant differences in propensities to Tweet incidents. Based on the estimated regression coefficients, users who have names enclosed in "echo" parentheses are significantly more likely than others to Tweet about long lines and less likely to Tweet about neutral or successful registration.

3 Discussion

Every indication is that Twitter can be used to develop data about individuals' observations of how American elections are conducted, data that cover the entire country with extensive and intensive local detail. Observations for each day can be gathered, and observations can be even more finely resolved in time: times can be resolved to the millisecond using the timestamps on Tweets. The frequency and likely the diversity of observations may vary depending on how many people care about an election and want to participate in it, observe it and comment on it.

Different types of users Tweet different kinds of observations. Users whose descriptions have word vectors that are cosine similar to “republican” are less likely than users whose descriptions are similar to “democrat” to Tweet about vote waiting times—whether about long lines or about short lines—and are more likely to Tweet about voting or registration problems and less likely to Tweet about voting or registration successes. Users whose descriptions are similar to various words that refer to either the Clinton or the Trump campaigns differ similarly. Those who signal affinity for an anti-antisemitic movement by modifying the format of their names Tweet about different types of incidents than do others. It is easy to imagine that other signals of political positions we might have identified would also have been associated with differences in the kinds of incidents users Tweeted about.

Differences in positions in the Twitter social network are also associated with differences in incident reporting, at least to the extent these network positions are captured by the numbers of “friends,” “followers” and triads (user-friend-friend and user-follower-follower triads) users are connected to. In light of regressions like the nonparametric regressions shown in Figures 14 and 15, all we feel confident to say here is that these associations are complicated. Because the differences in user descriptions also are associated with different relationships to networks and media outside of Twitter—as is suggested by the associations between the descriptions and tendencies to retweet influential ideological accounts—social and political positions more generally matter for propensities to Tweet about different types of incidents.

Features of particular localities also relate to what users Tweet about. Features of states such as state laws and state demographics are associated with the distribution of incidents. Whether such differences trace to mere perceptions or biases in reporting or to differences in what actually occurs needs to be investigated further. But findings such as that states like Oregon and Colorado that have eliminated voting in person present significantly and substantially fewer Tweets about long lines to vote suggest that reality is an important determinant of incident Tweets' content.

An important immediate step for development is to try better to exploit geolocation information. `place` information is available for some Tweets obtained via the Twitter API. Here we have illustrated how for such Tweets geography can be reliably resolved to states, but in fact in many cases resolution is possible to the city, neighborhood or even building. In some instances we found information to identify specific locations from metadata associated with Instagram images Tweets contain. We are currently exploring how much help we get from databases that provide "alternative names" for places. We can use such geographic identifications to place Tweets in particular election districts. Ideally we would like to associate Tweets with particular polling places, but for most Tweets that will not be possible. Some Tweets do contain exact information about the polling location in the Tweet text (or image), and we plan to investigate how to organize such information.

`place` information is not available for most Tweets from the Twitter API. When `place` information is missing we often have `location` information from user profiles. Such "location" data usually reflects the location associated generally with (and chosen by) the sender of the Tweet, not necessarily the place whence the Tweet originated. Perhaps in cases where voting happens in person, we can rely on selected locations to correspond both to where the sender lives and to the place where the sender is trying to vote—but clearly such is not a generally reliable assumption. Perhaps geolocation data can be used to develop models to estimate the likelihood that Tweets that do not have reliable geolocation information actually come from the place the "location" indicates. "Location" information is also often vague, which makes it challenging to associate incidents with particular election districts. That presents a challenge for the goal to combine such information with information about votes.

Another important development will be to add capabilities for machine classifiers jointly to use text and image information. Classifier performance for incidents such as line lengths and success at voting is good, but we expect that it would improve significantly if the classifier algorithms were able to interpret both images and text. Many Tweets that humans label for such instances have text that boils down to “look at this!” with an image clearly displaying a polling place, a long line or a smiling person wearing an “I voted” sticker. In fact we’re a bit surprised at how well the classifiers perform given that human judgments so frequently depend on images to which the classifiers have no access.

Even though we have strong evidence that different kinds of users Tweet about different types of incidents, we don’t know how much such differences are due to observational or reporting biases as opposed to differences in real experiences. We don’t know whether other observational biases—call them sampling biases—also affect the set of incidents observed using Twitter data. An obvious bias is that Tweets come only from individuals with a smartphone who use Twitter, and such individuals may not be be as frequently present at every place from which we would like to observe election incidents. Privacy settings in Twitter also limit the number of tweets we see, and incidence of (for us) adverse privacy settings may vary across time and space. When we rely on Tweets at election officials we may be biasing our data to include more observations from states with high degrees of professionalization in their county governments.

In general we cannot know whether purported incidents actually occurred, although in a few cases incidents alleged in Tweets can be verified by information obtained from other channels such as news reports or official reports. Many other questions will arise regarding observations derived from Twitter, but at this point it seems better to get the data so they can be critically appraised rather than not obtain the data at all.

From the Twitter API, using some keywords and a few specially targeted users, we extracted hundreds of thousands of observations of election incidents. Given that the streaming API supplies a one-percent sample of all Tweets, the full population of incident Tweets that exist to be extracted from all Tweets is likely in the tens of millions. Many of the types of incidents that are too infrequent for classifiers to resolve in our data are likely to be resolvable in a larger data collection. The distribution of such incidents is affected by differences across users such as we have examined, but knowing the biases we can adjust for them to try to obtain an unbiased view of what is happening—with great specificity and detail—in elections across the country. The differences across users are themselves interesting objects for study. Astronomers have extensive experience using filters to correct for distortions in images they get from telescopes: gravitational lensing is a thing, and even the Hubble telescope has glasses. The Twitter Election Observatory we are working to construct is no different.

4 Appendices

4.1 Twitter API Data

To access the Twitter API (Twitter, Inc. 2016*b*), we registered an application with Twitter.com, giving us the security tokens necessary to query data from Twitter's database.[37] In order to collect Tweets to and from election officials on and around the respective Election Days, we first had to find the Twitter accounts for those election officials.

These Twitter accounts were found in two ways: first, the Election Assistance Commission has collected information regarding the social media accounts of election officials at both the state and county levels across the United States, with varying degrees of completeness of data across states.[38] The second way these Twitter accounts were obtained was by manually searching Twitter for terms associated with the office of election officials, such as "election official," "county clerk," "department of elections" and "county auditor." Along with manually searching for election officials, user-created lists of election officials were searched for previously not-found election officials.[39] We used similar methods to find the Twitter accounts of state-level Republican and Democratic Parties, state-level Leagues of Women Voters, and state-level ACLUs. In order to facilitate these searches, we created a Twitter account affiliated with this research project.[40]

Our goal was to pull entire timelines from 493 accounts (for perspective, one California

[37] We used a combination of Python modules, mainly Twython and Tweepy. Code was adapted from (Bonzanini 2015; Moujahid 2014; Saxton 2014; Dolinar 2015)

[38] The list of resources can be found at `http://www.eac.gov/voter_resources/state_and_local_election_office_social_media_list.aspx`.

[39] An example of one of these user-created lists can be found at `https://twitter.com/EACgov/lists/us-election-officials/members`.

[40] The Twitter user name for this account can be found at `https://twitter.com/election_ballot`.

account had over eleven thousand Tweets in their timeline). A few challenges arose in querying that much data. First, user timelines are not static: a user can post Tweets while our application queries the data, which would effect the results; we had to recursively pull Tweets twenty at a time, starting with the user's most recent Tweet and ending with the first Tweet posted (in some cases dating back to 2007). Second, the sheer size of the query would occasionally break the script, so we had to pull timelines in batches; that is, we could not pull all 493 accounts at once, but rather, pull them fifty at a time. For perspective, a single batch would return hundreds of thousands of Tweets. Finally, Twitter places rate limits[41] on applications that query data from the API, so we had to design the script to pause in between requests. This way, we would not exceed rate limits, and the script could complete each query.

Part of the data collection was to identify tweets by their unique identification number, allowing us to quickly identify and omit duplicate tweets from our final dataset. The data returned are formatted in JSON[42], so we had to identify the specfic fields of interest (in this case, the unique identification number of each Tweet, its content, its timestamp, the name and location of each user, and the place whence the Tweet was sent, which was missing in most Tweets) and write them to a .csv file. Additionally, we were interested in obtaining geo-location data from each Tweet (returned in the form of coordinates) but Twitter's privacy settings are such that, this kind of data is not readily available for most users.

The bulk of the content was from outside of our time range, so it was not used. For the primary/caucus period we made sure that the data used from the Twitter API were from the same time frame as the data obtained via Sysomos. Part of the data collection was to identify Tweets by their unique identification number, allowing us to quickly identify and omit duplicate Tweets from our final dataset.

The second phase of data collection started in October 2016, corresponding with the

[41] Enforced on a "per access token" basis, Twitter limits users to fifteen requests per fifteen-minute window, although this number varies with the object being called; for more information on Twitter rate limits, see `https://dev.twitter.com/rest/public/rate-limiting`.

[42] JavaScript Object Notation, a data format represented by simple text, used to transfer data objects that consist of atrribute-value pairs; for more information on the format of Twitter data, see: `https://dev.twitter.com/rest/reference/get/search/tweets`.

beginning of early voting in the general election. Because we were now streaming data, we could use keywords as filters to capture tweets of interest. These keywords signaled issues with voting—voter complaints, registration issues, long lines, broken machines, etc. To supplement the absence of geo-location data, we pulled data from the `place` object. This object is part of the JSON metadata, but is associated with individual tweets rather than with a users' profile.[43].

The `fullname` field is used to do a reverse lookup of the state. Our code uses the GeoPy module in Python to access the Nominatim search tool used by OpenStreetMap. The tool itself allows for non-standard search of places and returns a standard dictionary of addresses and latitude/longitude coordinates.[44] The GeoPy module also offers the use of Google Maps, Bing Maps, or Yahoo BOSS, but the Nominatim geolocation service has the advantage of breaking down addresses into key-attribute pairs (Python dictionaries), whereas the other services rely solely on comma separated values. As addresses are not standardized, this can be problematic because the `state` field will not be in the same location for every query. Search results were checked by the authors to ensure the states returned matched the addresses provided in the Twitter metadata.

4.2 Sampling for Tweets in Training Set

The sample used for the initial general election training set was drawn in stages. The population is the Tweets from the streaming API that have `place` information. Initially we drew a simple random sample of 2,000 Tweets. Then from the remaining Tweets we added another sample of 10,000 Tweets. "Hits" being sparse—there were only 247 "hits"—we used a nearest-neighbor algorithm (trying to match the "hits") to select an additional 2,969 Tweets from 482,485 unique Tweet texts. Then from the remaining Tweets we added first a sample of 5,000 Tweets then a

[43]`place` is specified at the time a user posts a tweet: users are asked if Twitter can access location information, and if they respond yes, the object is attached to the tweet: "*Places* are specific, named locations with corresponding geo coordinates. They can be attached to Tweets by specifying a *place id* when tweeting. Tweets associated with places are not necessarily issued from that location but could also potentially be about that location" (Twitter 2017)

[44]OpenStreetMap is an open source, collaborative project that seeks to produce geographical data provided by users. Companies that use OpenStreetMap data include: Apple Inc., Flickr, MapQuest, and Craigslist (OpenStreetMap 2017).

sample of 10,000 Tweets stratified to include 5,000 Tweets during Oct 1-Nov 3 and 5,000 Tweets from Nov 4—Nov 8. Finally based on all the "hits" found in the previous samples, we added another 4,140 nearest neighbors. Dropping duplicate Tweet texts and Tweets whose `uniqueid` values had become corrupted we labeled in all 20,139 Tweet texts as "hits" or "not hits." Eliminating Tweets for which the `place` object exists but the `place$placename` value is missing leaves us with an initial training set of 19,179 Tweets.

4.3 Flowchart Development

A primary/caucus "hits" flowchart (not shown) was developed over the course of several individual handcoding sessions. Tweets with three or more agreements as "hits" (among five coders) were designated core Tweets; a random sample of Tweets with two or fewer agreements as "hits" were reviewed and collectively discussed. After the discussion, we used both the core Tweets and the discussion of the marginal Tweets to create what we call the "hits flowchart." The flowchart was developed to standardize "hits" classification among the authors and avoided a simple definitional basis for classifying "hits." The first half of the "hits" flowchart lays out what a "hit" is *not* (for instance, a "hit" is *not* an endorsement of a candidate); the second half of the "hits" flowchart engages with the substantive content of the Tweet and classifies the Tweet as a "hit" or not. This flowchart was used to create the training set, and coders currently use the flowchart to engage with the Tweets given by the active learning framework.

The general election "hits" flowchart (Figures 1, 2 and 3) reflects modifications to refer to all observed incidents without emphasizing "complaint" observations.

4.4 Coding Scheme for General Election Tweets

Instructions
After you have decided whether the Tweet in question is a "hit" according to the flowchart, use the categories and subcategories listed below to classify that "hit." These categories and definitions also may help you decide whether or not a Tweet is a "hit," if you are having trouble. A Tweet can be appropriately classified into multiple categories. For example, a Tweet that reads

“For some reason there was a problem with my voter registration, but the workers at my polling place were very helpful!” would fall within the “Election Official” and “Registration ” categories, with Adjectives 2 and 0, respectively.

Once you have placed a Tweet in its appropriate category(s), you will also note which adjective applies to the Tweet. A Tweet stating “The line at my polling place was long” would be coded as a 2. So your task is both to place the Tweet in its appropriate categories as well as to choose the appropriate adjective that more specifically describes the content of the Tweet. These adjectives are either dichotomous (0 or 2) or trichotomous (0, 1, or 2).

Importantly, at this time we are not concerned with any sentiment or emotion contained within the Tweet. We are concerned with some statements that are evaluative or normative. We are concerned with describing the factual (or purported factual) event or item to which the Tweet refers.

Coding Scheme for Categorization

0 or blank: Tweet does not fit within this category
1: Tweet fits within this category

4.4.1 Categories for Coding

1. Outside Influence

This category includes cases where the complainant encountered improper campaign advertising, such as advertising too close to a polling place. This category also includes observations alleging the campaign practices of candidates, parties, or outside entities such as PACs violate the spirit or letter of the law. Also included in this category are allegations of police misconduct relating to the administration or outcome of the election, as well as complaints or incidents regarding the media. For example, an individual might complain that the media called the election while people were still in line to vote, or reporters may be improperly interviewing voters. This category is in part analogous to the “Improper Outside Influence” EIRS category.

Adjective: N/A.

2. Disability/Accessibility Issue

Tweets that fall under this category would include observations about some aspect of the election that is accessible or not accessible for those with disabilities—for example, a polling place not offering special ballots or assistance to voters who are blind, or a polling place not being wheelchair accessible. This relates to the lack of disability access EIRS category.

Adjective:

0: The aspect of the election was inaccessible
1: Neutral mention of disability/accessibilty concerns.
2: The aspect of the election was accessible.

3. Line Length, Waiting Time, Polling Place Crowding

This category refers to the length of a line or time to wait to vote or register, or to a crowded or empty polling place, including statements about the polling place being too small. Other examples of this category include a person referencing how long they have had to wait to vote, or reporting that their caucus has been moved outside due to crowding. This relates to the "polling place chaos and crowding" EIRS category.

Adjective:

0: There is no crowd or line at the polling place;

1: There was a small crowd or short line or wait;

2: The polling place was crowded or there was a long line or wait (20 minutes or longer).

4. Polling Place Event

This category includes incidents related to the polling place, such as the set-up of the voting booths and other election structures. Another example of a problem that would fit in this category is the presence of security cameras observing how individuals vote. Furthermore, this category includes a voter being told a correct or incorrect location for their polling place or precinct's line, or not knowing where to go to vote. Statements about the convenience of a polling place are included in this category. "I voted" statements referring to actions on election day or by in-person early voting are included in this category. Finally, this category includes complaints or reports that allege intimidation by persons other than polling place officials that occurred while the relevant person was casting his or her ballot, approaching the polling place, or in the polling place. This category does not include corruption, malfeasance, impropriety or other comments regarding poll workers. It partially relates to the "Incorrect polling place/precinct information" and "Voter Intimidation" EIRS categories.

Adjective:

0: The polling place did not function as expected or information is incorrect

1: The Tweet describes the polling place without noting whether it or an aspect functioned correctly or incorrectly

2: The polling place did function correctly or information is correct

5. Electoral System

This includes observations relating to specific aspects of the American electoral system, such as voluntary participation, the necessity to register to vote (e.g., registration deadlines), the first-past-the-post system, top-two electoral systems, caucuses, open/closed primary elections or non-proportional representation. This also includes comments about improper district boundaries and gerrymandering. Finally, comments about the integrity of the voting process due to hacking or hacking concerns are included here.

Adjective:

0: the electoral system did not function appropriately

1: the Tweet makes a neutral statement about the electoral system without an indication of if it functioned appropriately

2: the electoral system functioned appropriately

6. Absentee, Mail-In, or Provisional Ballot Issue

This category relates to features of absentee or mail-in ballots, including ballots being received or not being received by the voter, ballots being mailed or ballots not being counted. Early voting incidents are also included: "I voted" statements referring to actions during early voting are included in this category.. This category also applies to incidents relating to provisional ballots, such as a voter having to vote provisionally (or not being allowed to). This category relates to the "provisional ballot abuse" and "Non-receipt of requested absentee ballots" EIRS category.

Adjective:

0: the absentee, mail-in, or provisional ballot system did not function appropriately

1: the Tweet makes a neutral observation or statement about the absentee, mail-in, or provisional ballot system without noting it having functioned correctly or incorrectly

2: the absentee, mail-in, or provisional ballot system functioned correctly

7. Election Official

Comments that accuse governmental, election workers (including poll workers), or election officials of corruption, malfeasance, ignorance, being unhelpful or non-responsive, being rude, or some other complaint. This includes allegations of mismanaging the election. This category includes reports that allege intimidation by polling place officials that occurred while the relevant person was attempting to register, casting his or her ballot, approaching the polling place, or in the polling place. A Tweet that falls in this category might instead note that a pollworker or election official was helpful, or the staff managed the polling place well. This category is analogous to the EIRS categories for "pollworker malfeasance/ineptitude" as well as "other election worker problem."

Adjective:

0: The Tweet notes that the election officials did not perform their duties

1: the Tweet makes a neutral observation about election officials without noting them having performed or not performed their duties

2: the Tweet notes that the election officials performed their duties

8. Voter Identification

The voter or prospective voter had issues relating to voter identification requirements. This might include an election official improperly asking for identification, problems or no problems acquiring or using identification, or being rejected at the polls due to lack (or accused lack) of necessary identification. This relates to the "Improper ID requirements" EIRS category.

Adjective:

0: the Tweet notes that there were problems with the voter identification process or application

1: the Tweet makes a neutral observation about voter identification policies

2: the Tweet indicates that the voter identification process or application functioned appropriately

9. Registration

Voters or prospective voters encountered difficulty registering to vote, had problems registering with their preferred party or registered without difficulty. It could also include instances of registration records being incorrect, or positive or neutral statements about the registration process. This also includes an individual noting that he or she has been able to register. Also included is information about registration deadlines or processes. This relates to "Incorrect registration lists/non-receipt of registration cards" EIRS category.

Adjective:

0: The Tweet indicates that an individual was not able to register to vote

1: the Tweet makes a neutral observation about the voter registration process without noting if the individual in question registered or not

2: the Tweet notes that the individual was able to register to vote

10. Voter Fraud

This category refers to instances or alleged instances of voter fraud, including a voter being told that he or she has already voted. This also includes an individual noting that another individual has voted twice or is impersonating another eligible voter. This category is analogous to EIRS category "Voter fraud." Need to update this language–will look at previous categories (EIRS, Germany)

Adjective:

0: The Tweet indicates that some form of voter fraud did occur

1: the Tweet makes an unspecific assertion about voter fraud.

2: the Tweet indicates that some form of voter fraud did not occur

11. Ballot and Voting Technology

This category includes complaints or incidents regarding the design of the ballot, including layout and foldability, or the design or operation of voting technology. The category includes voting technologies working well or being inoperable as well as clear or unclear instructions regarding how to use the voting technology. Also included are observations about the security of the technology. Examples could include machines misreading scanned ballots, not printing receipts, or machines being difficult to use. This category also applies to individuals being given the incorrect ballot, as well as a voter's preferred candidate or party not appearing on the ballot. Also situations involving electronic pollbooks. This category relates to the "other ballot problems" and "Machine malfunction/usage problem" EIRS categories.

Adjective:

0: the ballot or voting technology was confusing or defective

1: the Tweet makes a neutral observation or statement about the ballot or voting technology without noting it having functioned correctly or incorrectly

2: the ballot or voting technology was well-designed or functioned correctly

12. Unspecified Other

Includes complaints of which the nature is unclear as well as non-sequitur complaints. Analogous to the EIRS category "Other."

Adjective: N/A

4.5 Geodata

Utilizing location information from tweets is helpful for understanding voting patterns. Such location data can be found in a number of potential fields associated with a particular Tweet. The `fullname` field, as previously discussed, serves as one option. Users also have the option of selecting a pre-existing location name from a list of Twitter generated suggestions or writing a custom location name, and this forms the `location` field in the `user` object. Additional information can be recovered from tweets that link to an Instagram post, where the user has tagged the photo with a location. Instagram locations must be associated with an actual location, even if the user creates a custom name, so latitude and longitude coordinates are easily obtainable.[45] In order to try to resolve more locations to city-level detail, a basic search process was developed to find locations in a data base of over two million locations from GeoNames, including cities, buildings, and other natural features, like bodies of water.[46] This database contains a variety of information about each location, including population numbers and latitude/longitude coordinates. Consequently, tweets with locations that can be matched to an entry in the data base can therefore be associated with specific coordinates on a globe.

The search procedure was designed to look for locations that were formatted in the following ways:[47]

1. <city>, <state name>
2. <city>, <state postal abbreviation>
3. <city>, <state name>, <other information (most likely "USA" or "US")>
4. <city>, <state postal abbreviation>, <other information (most likely "USA" or "US")>

Both city names in the database as well as the location strings from tweets were lowercased, with special characters and spaces also removed. For each tweet with a location that was formatted in one of the 4 ways described above, the GeoNames data base was queried for locations that matched the "city" string exactly. When multiple results were returned, state information was used as a filter to obtain a unique location name, latitude, and longitude. These coordinates can then be used to determine other useful information, such as census tract, from the US Census Geocoder API.

[45] See https://help.instagram.com/408972995943225 and https://smartphones.gadgethacks.com/how-to/instagram-101-create-custom-location-instagram-0178307/.
[46] See http://www.geonames.org/about.html.
[47] There were some special cases handled as well. For example, New York, New York was mapped to New York City.

References

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20(2):211–234.

Bjornlund, Eric C. 2004. *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington, DC: Woodrow Wilson Center Press.

Bonzanini, Marco. 2015. "Mining Twitter Data with Python (Part 1: Collecting data)." `https://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/`. Accessed: 2016-06-15.

Bowman, Adrian W. and Adelchi Azzalini. 1997. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford: Clarendon Press.

Bowman, Adrian W. and Adelchi Azzalini. 2003. "Computational Aspects of Nonparametric Smoothing with Illustrations from the 'sm' Library." *Computational Statistics and Data Analysis* 42(4):545–560.

Brady, Henry E. and John E. McNulty. 2011. "Turning Out to Vote: The Costs of Finding and Getting to the Polling Place." *American Political Science Review* 105(1):115–134.

Cantu, Francisco and Sebastián M. Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning." *Political Analysis* 19(4):409–433.

Collins, Keith. 2017. "probabot: This new Twitter account hunts for bots that push political opinions." Quartz Media, October 25, 2017. URL: `https://qz.com/1110481/this-new-twitter-account-hunts-for-bots-that-push-political-opinions/`.

Dolinar, Sean. 2015. "Collecting Twitter Data: Using A Python Stream Listener." `http://stats.seandolinar.com/collecting-twitter-data-using-a-python-stream-listener/`. Accessed: 2016-06-15.

Faris, Rob, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman and Yochai Benkler. 2017. "Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election." Berkman Klein Center for Internet & Society at Harvard University, August

16, 2017. URL: https://cyber.harvard.edu/publications/2017/08/mediacloud.

Hall, Joseph Lorenzo. 2005. "Towards Work-Practice Modeling of Elections and the Election Incident Reporting System (EIRS)." URL https://josephhall.org/papers/eirs_brahms_mims.pdf (accessed July 5, 2006).

Herrnson, Paul S., Richard G. Niemi, Michael J. Hanmer, Benjamin B. Bederson, Frederick G. Conrad and Michael W. Traugott. 2008. *Voting Technology: The Not-So-Simple Act of Casting a Ballot*. Washington: Brookings.

Herron, Michael C. and Daniel A. Smith. 2016. "Precinct Resources and Voter Wait Times." *Electoral Studies* 42:249–263.

Huberman, Bernardo A., Daniel M. Romero and Fang Wu. 2009. "Social Networks That Matter: Twitter Under the Microscope." *First Monday* 14(1).
URL: *http://firstmonday.org/article/view/2317/2063*

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma*. New York: Cornell University Press.

Ichino, Nahomi and Matthias Schuendeln. 2012. "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana." *Journal of Politics* 74(1):292–307.

Jackman, Simon. 2017. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. Sydney, New South Wales, Australia: United States Studies Centre, University of Sydney. R package version 1.5.2.
URL: *https://github.com/atahk/pscl/*

Johnson, Jeff. 2005. "The EIRS 2004 and 2005 Voting Problem Incident Data: A Valuable Public Resource." URL http://triton.towson.edu/~jlazar/EIRS_factsheet.doc (accessed July 10, 2014).

Jones, Douglas W. and Barbara Simons. 2012. *Broken Ballots: Will Your Vote Count?* Chicago: University of Chicago Press.

Joyce, Gemma. 2016*a*. "Ranked: The Most Influential Democrats on Twitter." Brandwatch Blog, September 23, 2016. URL: https://www.brandwatch.com/blog/

react-the-most-influential-democrats-on-twitter/.

Joyce, Gemma. 2016*b*. "Ranked: The Most Influential Republicans on Twitter." Brandwatch Blog, September 13, 2016. URL: https://www.brandwatch.com/blog/react-ranked-influential-republicans-twitter/.

Kelley, Judith. 2012. *Monitoring Democracy: When International Election Observation Works and Why it Often Fails*. Princeton: Princeton University Press.

Krackhardt, David and Mark S. Handcock. 2007. "Heider vs. Simmel: Emergent Features in Dynamic Structures." *Lecture Notes in Computer Science* 4503:14–27.

Lan, Man, Chew-Lim Tan, Hwee-Boon Low and Sam-Yuan Sung. 2005. A Comprehensive Comparative Study on Term Weighting Schemes for Text Categorization with Support Vector Machines. In *Special interest tracks and posters of the 14th international conference on World Wide Web*. ACM pp. 1032–1033.

Lausen, Marcia. 2007. *Design for Democracy: Ballot and Election Design*. Chicago: University of Chicago Press.

Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz and Michael Neblo. 2010. "The Coevolution of Networks and Political Attitudes." *Political Communication* 27(3):248–274.

Leopold, Edda and Jörg Kindermann. 2002. "Text Categorization with Support Vector Machines. How to Represent Texts in Input Space?" *Machine Learning* 46:423–444.

Levin, Inès, Gabe A. Cohn, Peter C. Ordeshook and R. Michael Alvarez. 2009. "Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela." 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections, Montreal, August 10–11.

McGrath, Ryan. 2016. "twython." Version: 3.4.0. URL: https://twython.readthedocs.io/en/latest/.

Mebane, Jr., Walter R. 2010. "Fraud in the 2009 Presidential Election in Iran?" *Chance* 23:6–15.

Mebane, Jr., Walter R. 2013. "Election Forensics: The Meanings of Precinct Vote Counts' Second Digits." Paper presented at the 2013 Summer Meeting of the Political Methodology Society, University of Virginia, July 18–20, 2013.

Mebane, Jr., Walter R. 2014. Can Votes Counts' Digits and Benford's Law Diagnose Elections? In *The Theory and Applications of Benford's Law*, ed. Steven J. Miller. Princeton: Princeton University Press pp. 206–216.

Mebane, Jr., Walter R. 2016. "Election Forensics: Frauds Tests and Observation-level Frauds Probabilities." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2016.

Mebane, Jr., Walter R., Alejandro Pineda, Logan Woods, Joseph Klaver, Patrick Wu and Blake Miller. 2017. "Using Twitter to Observe Election Incidents in the United States." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 6–9, 2017.

Mebane, Jr., Walter R. and Jonathan Wall. 2015. "Election Frauds, Postelection Legal Challenges and Geography in Mexico." Paper presented at the 2015 Annual Meeting of the American Political Science Association, San Francisco, CA, September 3–6.

Mebane, Jr., Walter R., Joseph Klaver and Blake Miller. 2016. "Frauds, Strategies and Complaints in Germany." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2016.

Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." Proceedings of the International Conference on Learning Representations. URL: `http://arxiv.org/abs/1301.3781`.

Miller, Blake and Joseph Klaver. 2016. "Active Learning for Costly Text Classification Problems and Exploratory Data Analysis." Unpublished paper draft.

Montgomery, Jacob M., Santiago Olivella, Joshua D. Potter and Brian F. Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* forthcoming.

Moujahid, Adil. 2014. "An Introduction to Text Mining using Streaming API and Python." `http://adilmoujahid.com/posts/2014/07/twitter-analytics/`. Accessed: 2016-06-15.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitry Shaikin. 2009. *The Forensics of Election*

Fraud: With Applications to Russia and Ukraine. New York: Cambridge University Press.

OpenStreetMap. 2017. "Nominatim." http://wiki.openstreetmap.org/wiki/Nominatim. Accessed: 2017-01-15.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Pericchi, Luis Raúl and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26(4):502–516.

Quartz. 2017. "probably_bots: List members." List as of December 23, 2017. URL: https://twitter.com/probabot_/lists/probably-bots/members.

Quesenbery, Whitney and Margaret Chen. 2008. "Better Ballots." Brennan Center for Justice. URL: https://www.brennancenter.org/publication/better-ballots.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

Rakotomamonjy, Alain. 2003. "Variable Selection Using SVM-based Criteria." *Journal of Machine Learning Research* 3(Mar):1357–1370.

Rozenas, Arturas. 2017. "Detecting Election Fraud from Irregularities in Vote-Share Distributions." *Political Analysis* 25(1):41–56.

Saxton, Gregory. 2014. "Using Python to Grab Twitter User Data." http://social-metrics.org/twitter-user-data/. Accessed: 2016-06-15.

Select Committee Democrats. 2017. "HPSCI Minority Exhibit B." U.S. House of Representatives Permanent Select Committee on Intelligence: Democrats, HPSCI 11 1, https://democrats-intelligence.house.gov/hpsci-11-1/hpsci-minority-exhibit-b.htm. URL: https://democrats-intelligence.house.gov/uploadedfiles/exhibit_ b.pdf.

Settles, Burr. 2010. "Active learning literature survey." from Computer Sciences Technical Report 1648 http://burrsettles.com/pub/settles.activelearning.pdf.

Stewart, III, Charles and Stephen Ansolabehere. 2015. "Waiting to Vote." *Election Law Journal* 14(1):47–53.

Twitter, Inc. 2016*a*. "Documentation." URL https://dev.twitter.com/overview/documentation.

Twitter, Inc. 2016*b*. "REST APIs." URL https://dev.twitter.com/rest/public.

Twitter, Inc. 2017. "Twitter Development Documentation: Places." https://dev.twitter.com/overview/api/places. Accessed: 2017-01-15.

United States Census Bureau. 2016. "State Population Totals Tables: 2010-2016." URL https://www.census.gov/data/tables/2016/demo/popest/state-total.html, accessed March 12, 2016.

Verified Voting Foundation. 2005. "Election Incident Reporting System." URL https://voteprotect.org/index.php?display=EIRMapNation&tab=ED04 (accessed July 5, 2006).

Williams, Zoe. 2016. "(((Echoes))): Beating the Far Right, Two Triple Brackets at a Time." June 212. URL https://www.theguardian.com/technology/shortcuts/2016/jun/12/echoes-beating-the-far-right-two-triple-brackets-at-a-time.

Zeileis, Achim, Christian Kleiber and Simon Jackman. 2008. "Regression Models for Count Data in R." *Journal of Statistical Software* 27(8).
URL: *http://www.jstatsoft.org/v27/i08/*

Table 1: Number of General Election Incident Observations in Training Sample by State

State	Count	State	Count	State	Count
Alabama	23	Kentucky	19	North Dakota	4
Alaska	4	Louisiana	42	Ohio	116
Arizona	62	Maine	9	Oklahoma	29
Arkansas	18	Maryland	71	Oregon	25
California	245	Massachusetts	92	Pennsylvania	72
Colorado	28	Michigan	44	Rhode Island	5
Connecticut	13	Minnesota	42	South Carolina	38
Delaware	6	Mississippi	10	South Dakota	2
District of Columbia	64	Missouri	48	Tennessee	65
Florida	157	Montana	5	Texas	296
Georgia	99	Nebraska	12	Utah	19
Hawaii	7	Nevada	58	Vermont	1
Idaho	4	New Hampshire	7	Virginia	61
Illinois	85	New Jersey	34	Washington	43
Indiana	69	New Mexico	5	West Virginia	12
Iowa	15	New York	188	Wisconsin	25
Kansas	18	North Carolina	175	Wyoming	2
Puerto Rico	1	Virgin Islands	1		

Note: Number of Tweets observing incidents in initial human-labeled training sample ($n = 2,610$) by State obtained via Twitter Streaming API during October 1–November 8, 2016.

Table 2: General Election Machine "Hit" Classifier Performance

	Without Stemming			With Stemming			
Class	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Support
Not a hit	.92	.93	.92	.91	.92	.92	3300
Hit	.78	.76	.77	.81	.72	.76	1127
Average/Total	.88	.89	.88	.88	.89	.88	4427

Table 3: Incident Type Frequency in Initial Sample of General Election "Hits"

Incident Category		Raw	Adjectives		
Description	Number	Count	0	1	2
Outside Influence	1	1	—	—	—
Disability/Accessibility Issue	2	2	1	0	1
Line Length, Waiting Time	3	752	61	77	614
Polling Place Event	4	700	51	327	322
Electoral System	5	199	46	137	16
Absentee or Early Voting Issue	6	733	70	269	394
Election Official	7	29	9	1	19
Voter Identification	8	14	2	7	5
Registration	9	156	27	107	22
Voter Fraud	10	5	1	4	0
Ballot and Voting Technology	11	20	15	4	1
Unspecified Other	12	3	—	—	—
Not an Incident		42	—	—	—

Note: Manual type classifications for 1,149 Tweet texts sampled from the 40,687 Tweet texts classified as "hits," using the coding scheme described in Appendix Section 4.4. Dashes indicate subtypes ("adjectives") that are not defined.

Table 4: General Election Binarized Classifier Performance

(a) SVM Classifier:

	Raw		Adjective					
			0		1		2	
Class	F	*M*	F	*M*	F	*M*	F	*M*
Outside Influence	—	2	*		*		*	
Disability/Accessibility Issue	1.00	21	—	9	—	0	—	12
Line Length, Waiting Time	.88	996	.26	82	.26	85	.82	829
Polling Place Event	.76	1411	.27	83	.52	449	.61	879
Electoral System	.62	552	.08	81	.67	438	.15	33
Absentee or Early Voting Issue	.84	1614	.33	116	.60	565	.66	933
Election Official	.07	50	—	17	—	7	.22	26
Voter Identification	.63	31	—	8	—	16	—	7
Registration	.86	440	.24	58	.78	308	.41	74
Voter Fraud	—	15	—	7	—	8	—	0
Ballot and Voting Technology	.29	48	.27	31	—	15	—	2
Unspecified Other	—	3	*		*		*	
Not an incident	.63	1044	*		*		*	

(b) Ensemble Classifier:

	Raw		Adjective					
			0		1		2	
Class	F	*M*	F	*M*	F	*M*	F	*M*
Line Length, Waiting Time	.91	996	.61	82	.21	85	.84	829
Polling Place Event	.78	1411	.08	83	.47	449	.63	879
Electoral System	.63	552	.05	81	.65	438	—	
Absentee or Early Voting Issue	.87	1614	.34	116	.60	565	.70	933
Registration	.85	440	.17	58	.84	308	.50	74
Not an incident	.66	1044	*		*		*	

Note: Overall number of labeled Tweets: 4,018. "F" is F-measure and *M* is support (the number of instances of the class). A dash indicates a class that is not used to determine (a) active-learning sampling or (b) a final classification. An asterisk indicates a class that is not defined. (a) binarized SVM performance at the end of the process of human labeling of types of incidents guided by active learning. (b) binarized ensemble classifier performance.

Table 5: Liberal Retweet Counts Regressed on Word2vec Cosine Similarities

	Count Models				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	2.43 (.03)	2.33 (.02)	2.42 (.02)	2.39 (.02)	2.37 (.02)
trump	−0.18 (.15)				
clinton	1.37 (.14)				
donald		−2.12 (.17)			
hillary		3.19 (.16)			
realdonaldtrump			−0.18 (.11)		
hillaryclinton			1.48 (.11)		
maga				−0.45 (.12)	
strongertogether				1.81 (.12)	
republican					−1.52 (.18)
democrat					2.89 (.17)
log(θ)	−1.32 (.03)	−1.31 (.02)	−1.30 (.03)	−1.31 (.03)	−1.28 (.03)
	Zero-inflation Models				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	2.49 (.38)	2.69 (.39)	2.56 (.37)	2.44 (.38)	2.61 (.38)
trump	1.17 (.40)				
clinton	0.69 (.42)				
donald		1.67 (.52)			
hillary		−0.68 (.56)			
realdonaldtrump			1.01 (.33)		
hillaryclinton			1.30 (.32)		
maga				1.95 (.38)	
strongertogether				−1.27 (.37)	
republican					1.88 (.52)
democrat					0.16 (.50)
log(1 + friend triads)	−0.25 (.02)	−0.28 (.02)	−0.24 (.02)	−0.26 (.02)	−0.24 (.02)
log(1 + friend_counts)	−0.45 (.07)	−0.44 (.07)	−0.47 (.07)	−0.42 (.07)	−0.47 (.07)
log(1 + follower triads)	−0.33 (.02)	−0.34 (.03)	−0.33 (.02)	−0.32 (.02)	−0.32 (.02)
log(1 + follower_counts)	0.34 (.05)	0.29 (.06)	0.34 (.05)	0.31 (.05)	0.33 (.05)
log likelihood	−24960	−24830	−24920	−24890	−24830

Note: zero-inflated negative binomial regression of counts of retweets of liberal influential accounts on cosine similarity coefficients derived from word2vec scores for users' `description` strings. Estimated using a random sample of incident-Tweeting users (n = 10000). Table reports coefficient estimates (standard errors in parentheses).

Table 6: Conservative Retweet Counts Regressed on Word2vec Cosine Similarities

	Count Models				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	2.13 (0.30)	3.32 (0.40)	2.61 (0.31)	2.38 (0.29)	3.67 (0.47)
trump	7.35 (1.22)				
clinton	−4.91 (1.42)				
donald		5.85 (1.39)			
hillary		−4.27 (1.66)			
realdonaldtrump			4.22 (0.71)		
hillaryclinton			−3.37 (1.00)		
maga				4.69 (0.65)	
strongertogether				−3.62 (0.87)	
republican					6.90 (1.02)
democrat					−6.98 (1.56)
log(θ)	−3.02 (0.13)	−3.16 (0.14)	−3.10 (0.14)	−3.04 (0.13)	−2.17 (0.30)
	Zero-inflation Models				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	6.86 (0.66)	6.82 (0.68)	6.54 (0.64)	6.60 (0.67)	7.75 (0.54)
trump	−4.69 (0.82)				
clinton	−2.02 (1.02)				
donald		4.22 (1.03)			
hillary		−11.36 (1.41)			
realdonaldtrump			−4.32 (0.66)		
hillaryclinton			−1.84 (0.65)		
maga				−4.08 (0.16)	
strongertogether				−3.80 (0.76)	
republican					−2.77 (0.67)
democrat					−5.21 (0.95)
log(1 + friend triads)	0.03 (0.04)	−0.02 (0.04)	0.03 (0.04)	0.03 (0.04)	−0.12 (0.03)
log(1 + friend_counts)	−0.45 (0.15)	−0.44 (0.15)	−0.51 (0.15)	−0.45 (0.15)	−0.45 (0.12)
log(1 + follower triads)	−0.58 (0.07)	−0.54 (0.07)	−0.59 (0.07)	−0.55 (0.07)	
log(1 + follower_counts)	0.52 (0.12)	0.56 (0.13)	0.63 (0.12)	0.55 (0.13)	0.16 (0.09)
log likelihood	−2006	−2020	−2034	−2002	−1975

Note: zero-inflated negative binomial regression of counts of retweets of conservative influential accounts on cosine similarity coefficients derived from word2vec scores for users' `description` strings. Estimated using a random sample of incident-Tweeting users (n = 10000). Table reports coefficient estimates (standard errors in parentheses).

Table 7: General Election Incident Observation Tweets by State

(a) Tweets with `place` information:

Unique Tweet Texts

State	count	State	count	State	count	State	count
AK	52	ID	80	MT	45	RI	88
AL	401	IL	1,279	NC	2,079	SC	402
AR	289	IN	848	ND	42	SD	45
AZ	823	KS	268	NE	151	TN	893
CA	4,522	KY	386	NH	106	TX	4,395
CO	590	LA	526	NJ	737	UT	239
CT	272	MA	1,043	NM	141	VA	1,185
DC	787	MD	978	NV	707	VT	59
DE	85	ME	132	NY	2,773	WA	745
FL	3,249	MI	792	OH	1,673	WI	401
GA	1,532	MN	479	OK	351	WV	154
HI	122	MO	556	OR	383	WY	16
IA	247	MS	198	PA	1,358		
PR	19	VI	2				

(b) Tweets with or without `place` information:

Unique Tweet Texts:

State	count	State	count	State	count	State	count
AK	133	ID	510	MT	289	RI	454
AL	1,775	IL	5,381	NC	8,274	SC	1,626
AR	1,277	IN	3,202	ND	175	SD	244
AZ	1,907	KS	1,559	NE	871	TN	3,872
CA	20,546	KY	1,373	NH	459	TX	19,922
CO	3,121	LA	2,348	NJ	2,806	UT	1,259
CT	1,274	MA	5,433	NM	722	VA	4,625
DC	6,047	MD	3,299	NV	2,489	VT	346
DE	380	ME	627	NY	15,182	WA	4,095
FL	12,552	MI	3,455	OH	6,342	WI	2,224
GA	6,069	MN	2,658	OK	1,855	WV	546
HI	273	MO	2,720	OR	2,317	WY	111
IA	1,220	MS	674	PA	5,372		
PR	121	VI	57				

Note: Number of unique Tweet texts classified as "hits" by State. (a) Counts using the 39,726 (of 40,678) Tweets for which a state could be identified from `place` or `location` information. (b) Counts using the 176,468 (of 315,180) Tweets for which a state could be identified from `place` or `location` information.

Table 8: Per Capita General Election Incident Observations by State

(a) Tweets with `place` information:

Unique Tweet Texts

State	rate	State	rate	State	rate	State	rate
AK	70.1	ID	47.5	MT	43.2	RI	83.3
AL	82.5	IL	99.9	NC	204.9	SC	81.0
AR	96.7	IN	127.8	ND	55.4	SD	52.0
AZ	118.7	KS	92.2	NE	79.2	TN	134.3
CA	115.2	KY	87.0	NH	79.4	TX	157.7
CO	106.5	LA	112.4	NJ	82.4	UT	78.3
CT	76.1	MA	153.1	NM	67.8	VA	140.9
DC	1155.4	MD	162.6	NV	240.5	VT	94.5
DE	89.3	ME	99.1	NY	140.4	WA	102.2
FL	157.6	MI	79.8	OH	144.0	WI	69.4
GA	148.6	MN	86.8	OK	89.5	WV	84.1
HI	85.4	MO	91.3	OR	93.6	WY	27.3
IA	78.8	MS	66.2	PA	106.2		

(b) Tweets with or without `place` information:

Unique Tweet Texts

State	rate	State	rate	State	rate	State	rate
AK	179.3	ID	303.0	MT	277.2	RI	429.8
AL	365.0	IL	420.3	NC	815.4	SC	327.7
AR	427.3	IN	482.7	ND	230.9	SD	281.9
AZ	275.1	KS	536.2	NE	456.7	TN	582.2
CA	523.5	KY	309.4	NH	343.9	TX	715.0
CO	563.3	LA	501.5	NJ	313.7	UT	412.6
CT	356.2	MA	797.6	NM	346.9	VA	549.8
DC	8877.4	MD	548.3	NV	846.6	VT	554.0
DE	399.1	ME	470.9	NY	768.9	WA	561.9
FL	609.0	MI	348.0	OH	546.0	WI	384.9
GA	588.6	MN	481.5	OK	472.8	WV	298.2
HI	191.1	MO	446.4	OR	566.0	WY	189.6
IA	389.2	MS	225.5	PA	420.2		

Note: Number per million persons (per state) of Tweets observing incidents based on (a) 39,726 and (b) 176,468 classified "hits" obtained via Twitter APIs for the general election period. 2016 state population data source: United States Census Bureau (2016).

Table 9: General Election Types of Incidents

Class	Raw	Adjective 0	Adjective 1	Adjective 2
Line Length, Waiting Time	27,167	2,159	1,060	23,869
Polling Place Event	58,871	1,946	15,445	49,561
Electoral System	49,359	10,378	38,831	—
Absentee or Early Voting Issue	105,577	9,127	31,816	65,168
Registration	49,020	17,578	32,160	6,325
Not an incident	89,917	*	*	*

Note: Overall $n = 315,180$ incident Tweets. A dash indicates a class that is not used to determine active-learning sampling or a final classification. An asterisk indicates a class that is not defined.

Table 10: OLS Regression Models for State-level Per Capita Incidents

		Line Length		Voting		Absentee	
Model	Covariate	coef.	s.e.	coef.	s.e.	coef.	s.e.
(a)	(Intercept)	.248	.011	.241	.0067	.272	.0056
	Voter ID	.091	.014	.074	.0093	.016	.0076
	Early Voting	.297	.014	.190	.0087	.117	.0070
	ID × Early Voting	−.456	.019	−.331	.0119	−.257	.0094
(b)	(Intercept)	.150	.014	.148	.0086	.184	.0078
	Voter ID	.197	.019	.184	.0122	.131	.0105
	EV plus In-person Abs.	.366	.016	.273	.0099	.201	.0086
	ID × EV+In-person Abs.	−.496	.022	−.406	.0139	−.347	.0116
(c)	(Intercept)	.192	.012	.184	.0075	.209	.0063
	Voter ID	.038	.014	.034	.0090	−.025	.0074
	No Excuse Absentee	.346	.015	.250	.0091	.192	.0075
	ID × No Excuse Absentee	−.300	.020	−.228	.0122	−.165	.0096
(d)	(Intercept)	−6.8	.06	−5.7	.04	−5.3	.03
	White	12.3	.12	10.3	.09	9.5	.07
	Bachelor's plus	33.9	.25	28.1	.19	26.0	.14
	White × Bachelor's plus	−59.0	.55	−48.5	.41	−44.0	.31

Note: ordinary least squares regression coefficients and standard errors estimated using classified incident types among the 176,468 incident Tweet texts for which state information could be extracted. "Line Length" models: n = 13602. "Voting" models: n = 25776. "Absentee" models: n = 41737.

Table 11: Types of Incidents Regressed on User Description Cosine Similarities

	Long Line	No or Short Line	Problem Voting	Neutral or Successful Voting	Problem Registering	Neutral or Successful Registration
(Intercept)	−0.81	−1.69	−4.55	−1.29	−0.63	−1.86
	(2.19)	(207)	(0.80)	(0.84)	(1.85)	(1.33)
republican	−0.85	−0.45	0.42	−0.70	0.44	−0.13
	(0.06)	(0.15)	(0.08)	(0.03)	(0.07)	(0.05)
democrat	0.36	0.13	−0.76	0.63	−0.80	0.79
	(0.06)	(0.16)	(0.09)	(0.03)	(0.07)	(0.05)
echo parentheses	0.29	0.23	0.20	0.02	0.07	−0.24
	(0.08)	(0.23)	(0.12)	(0.05)	(0.11)	(0.08)
looks-like-a-bot	−0.10	−0.69	−0.22	−0.23	−0.48	0.61
	(0.09)	(0.32)	(0.12)	(0.05)	(0.11)	(0.05)
irrelevant	−1.83	−2.64	−1.94	−0.14	−1.98	−1.43
	(0.25)	(1.00)	(0.31)	(0.05)	(0.22)	(0.13)
(Intercept)	−0.79	−16.8	−4.54	−1.30	−0.49	−2.02
	(2.20)	(207)	(0.80)	(0.84)	(1.86)	(1.33)
trump	−0.25	−0.11	−0.28	0.06	−0.07	−0.91
	(0.05)	(0.13)	(0.07)	(0.03)	(0.06)	(0.04)
clinton	−0.62	−0.45	0.06	−0.39	−0.07	1.26
	(0.05)	(0.13)	(0.07)	(0.03)	(0.06)	(0.04)
echo parentheses	0.29	0.23	0.18	0.03	0.06	−0.26
	(0.08)	(0.23)	(0.12)	(0.05)	(0.11)	(0.08)
looks-like-a-bot	−0.09	−0.68	−0.23	−0.22	−0.49	0.61
	(0.09)	(0.32)	(0.12)	(0.05)	(0.11)	(0.05)
irrelevant	−1.83	−2.63	−1.91	−0.15	−1.95	−1.51
	(0.25)	(1.00)	(0.31)	(0.05)	(0.22)	(0.13)
(Intercept)	−0.77	−1.69	−4.54	−1.32	−0.50	−1.95
	(2.13)	(0.02)	(0.80)	(0.84)	(1.85)	(1.33)
donald	−1.26	−0.57	−0.26	−0.36	−0.08	0.16
	(0.06)	(0.16)	(0.09)	(0.03)	(0.07)	(0.05)
hillary	0.44	−0.00	0.08	0.03	−0.03	0.24
	(0.06)	(0.17)	(0.09)	(0.04)	(0.07)	(0.05)
echo parentheses	0.25	0.21	0.18	0.01	0.06	−0.23
	(0.08)	(0.23)	(0.12)	(0.05)	(0.11)	(0.08)
looks-like-a-bot	−0.10	−0.69	−0.23	−0.22	−0.49	0.63
	(0.09)	(0.32)	(0.12)	(0.05)	(0.11)	(0.05)
irrelevant	−1.83	−2.63	−1.90	−0.16	−1.94	−1.47
	(0.25)	(1.00)	(0.31)	(0.05)	(0.22)	(0.13)

Note: logistic regressions of type of Tweet on cosine similarities based on users' `description` strings. $n = 315,180$. Table reports coefficient estimates (standard errors in parentheses). Coefficients for state dummy variables and for friend, follower and triad polynomials are not shown.

Table 12: Types of Incidents Regressed on User Description Cosine Similarities

	Long Line	No or Short Line	Problem Voting	Neutral or Successful Voting	Problem Registering	Neutral or Successful Registration
(Intercept)	−0.80	−1.69	−4.55	−1.29	−0.52	−2.02
	(2.16)	(207)	(0.80)	(0.84)	(1.85)	(1.33)
realdonaldtrump	−0.93	−0.63	−0.01	−0.38	0.19	−0.30
	(0.04)	(0.10)	(0.06)	(0.02)	(0.04)	(0.03)
hillaryclinton	0.09	0.17	−0.30	0.06	−0.47	1.09
	(0.04)	(0.10)	(0.06)	(0.02)	(0.05)	(0.03)
echo parentheses	0.27	0.22	0.19	0.02	0.08	−0.26
	(0.08)	(0.23)	(0.12)	(0.05)	(0.11)	(0.08)
looks-like-a-bot	−0.08	−0.69	−0.22	−0.22	−0.47	0.59
	(0.09)	(0.32)	(0.12)	(0.05)	(0.11)	(0.05)
irrelevant	−1.80	−2.61	−1.90	−0.15	−1.95	−1.45
	(0.25)	(1.00)	(0.31)	(0.05)	(0.22)	(0.13)
(Intercept)	−0.78	−1.69	−4.54	−1.31	−0.52	−1.94
	(2.19)	(207)	(0.80)	(0.84)	(1.85)	(1.33)
maga	−0.73	−0.38	−0.25	−0.21	−0.05	−0.48
	(0.04)	(0.10)	(0.06)	(0.02)	(0.05)	(0.03)
strongertogether	−0.13	−0.26	0.14	−0.19	0.00	0.91
	(0.04)	(0.12)	(0.06)	(0.02)	(0.05)	(0.04)
echo parentheses	0.28	0.23	0.18	0.03	0.06	−0.28
	(0.08)	(0.23)	(0.12)	(0.05)	(0.11)	(0.08)
looks-like-a-bot	−0.09	−0.68	−0.23	−0.22	−0.49	0.62
	(0.09)	(0.32)	(0.12)	(0.05)	(0.11)	(0.05)
irrelevant	−1.81	−2.62	−1.90	−0.15	−1.95	−1.46
	(0.25)	(1.00)	(0.31)	(0.05)	(0.22)	(0.13)

Note: logistic regressions of type of Tweet on cosine similarities based on users' `description` strings. $n = 315,180$. Table reports coefficient estimates (standard errors in parentheses). Coefficients for state dummy variables and for friend, follower and triad polynomials are not shown.

Figure 1: General Election Flowchart for Making Hits Decisions in American Twitter Election Comments, Part 1

An **observation** is a statement that refers to a (probably) real situation with which the Tweeter (probably) had personal familiarity: either the person witnessed the situation or personally knew the person who did; in cases where there is ambiguity about the directness of the personal involvement, the observation report will be treated as if it were personal. So descriptions that are entirely about news reports are generally excluded, but if it's not clear that the item comes from a news report we'll include it.

- Personal involvement does not mean the observation refers to a personal experience: statements about collective situations such as the electoral system, voter registration procedures and electoral administration are also to be included.
- The observation may be embedded in an opinion, comment or advocacy statement, but advocacy statements per se are to be excluded. The observation may be adjacent to a news report that is relinked but news reports per se are to be excluded.
- Notice that an observation may take the form of an image, so remember to consider what any URLs contain. Keep that point in mind as you move down the flow chart.
- If the Tweet contains editorializing comments, be sure to identify the specific observation about the electoral process that the comments may be making.

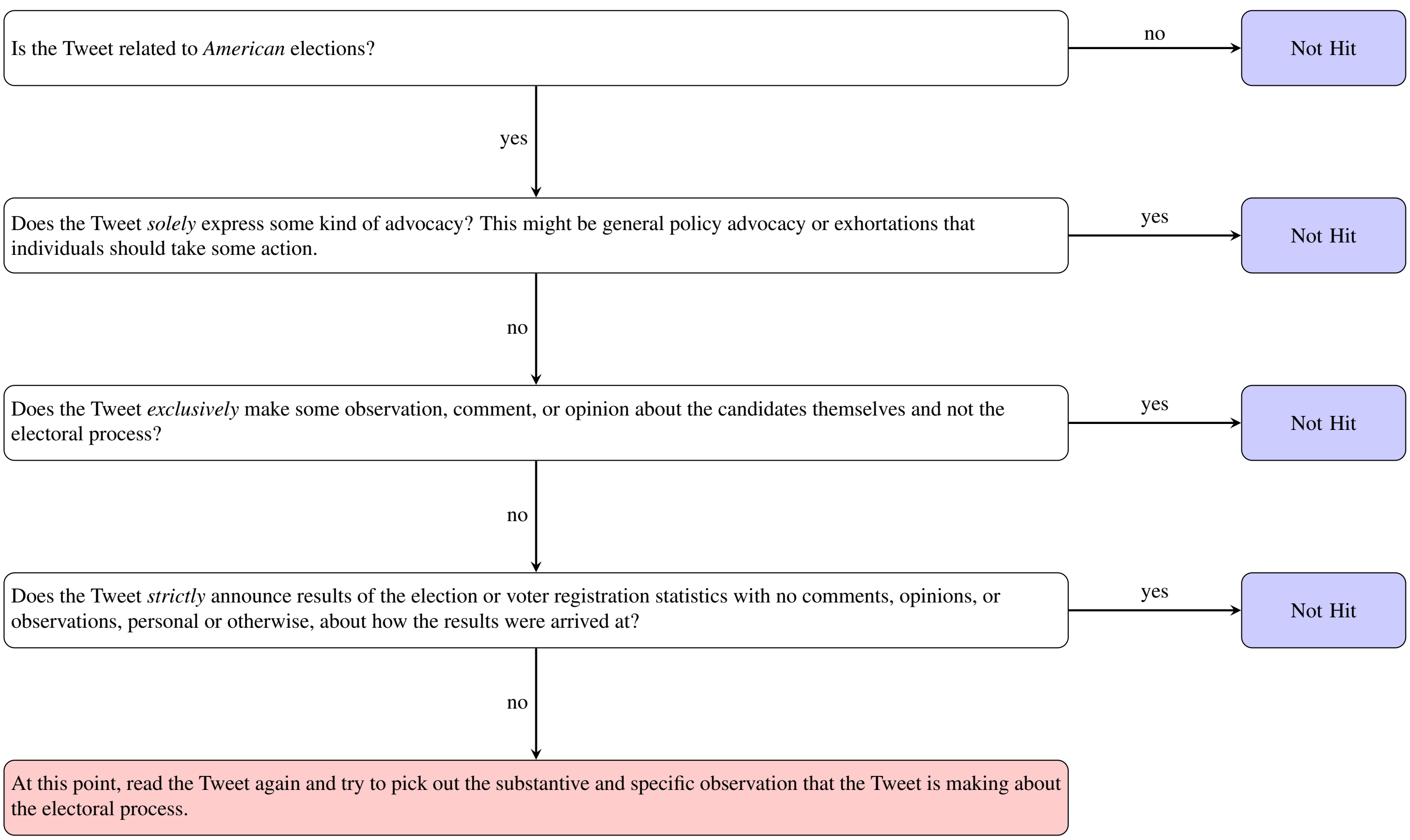


Figure 2: General Election Flowchart for Making Hits Decisions in American Twitter Election Comments, Part 2

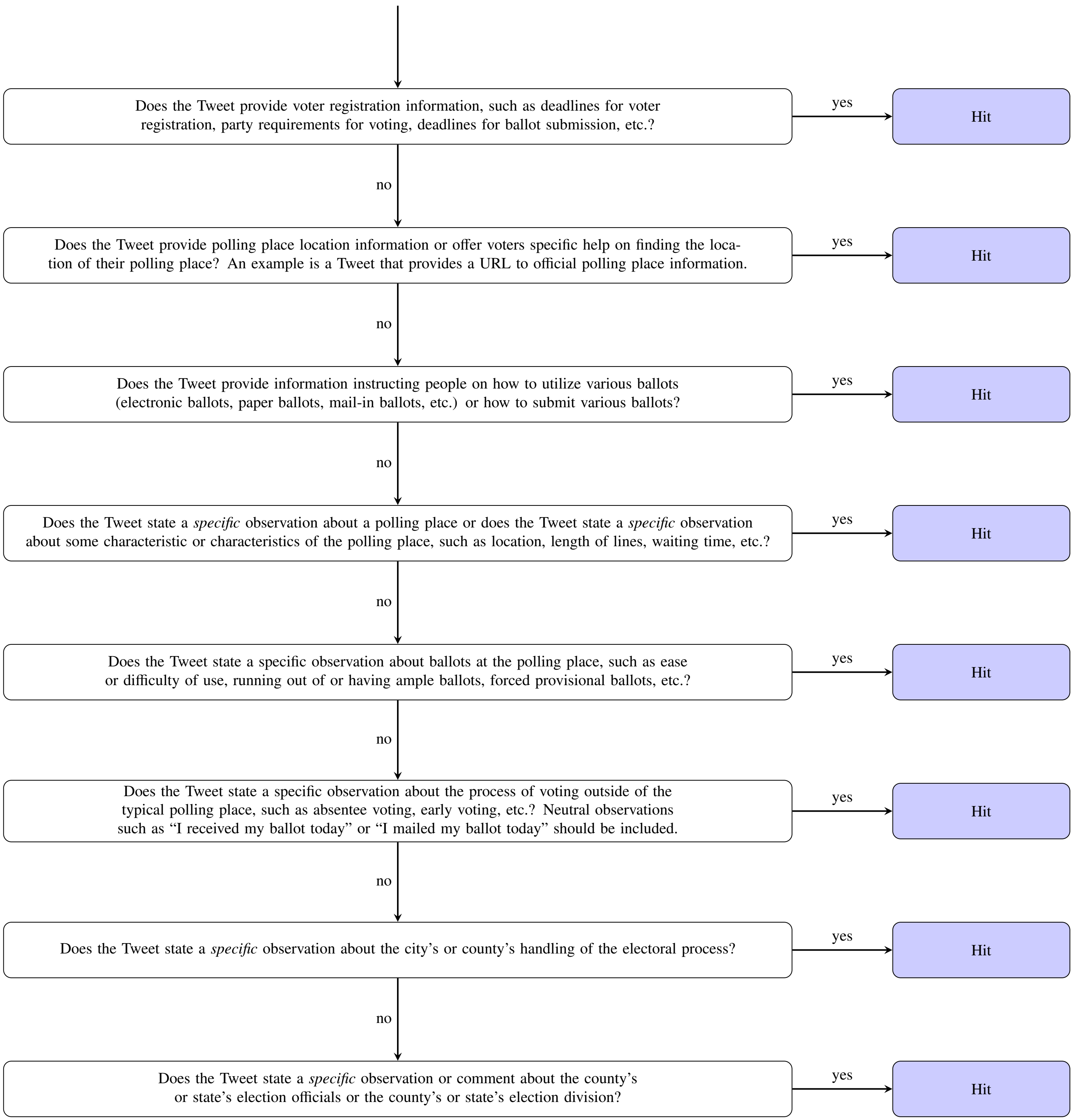


Figure 3: General Election Flowchart for Making Hits Decisions in American Twitter Election Comments, Part 3

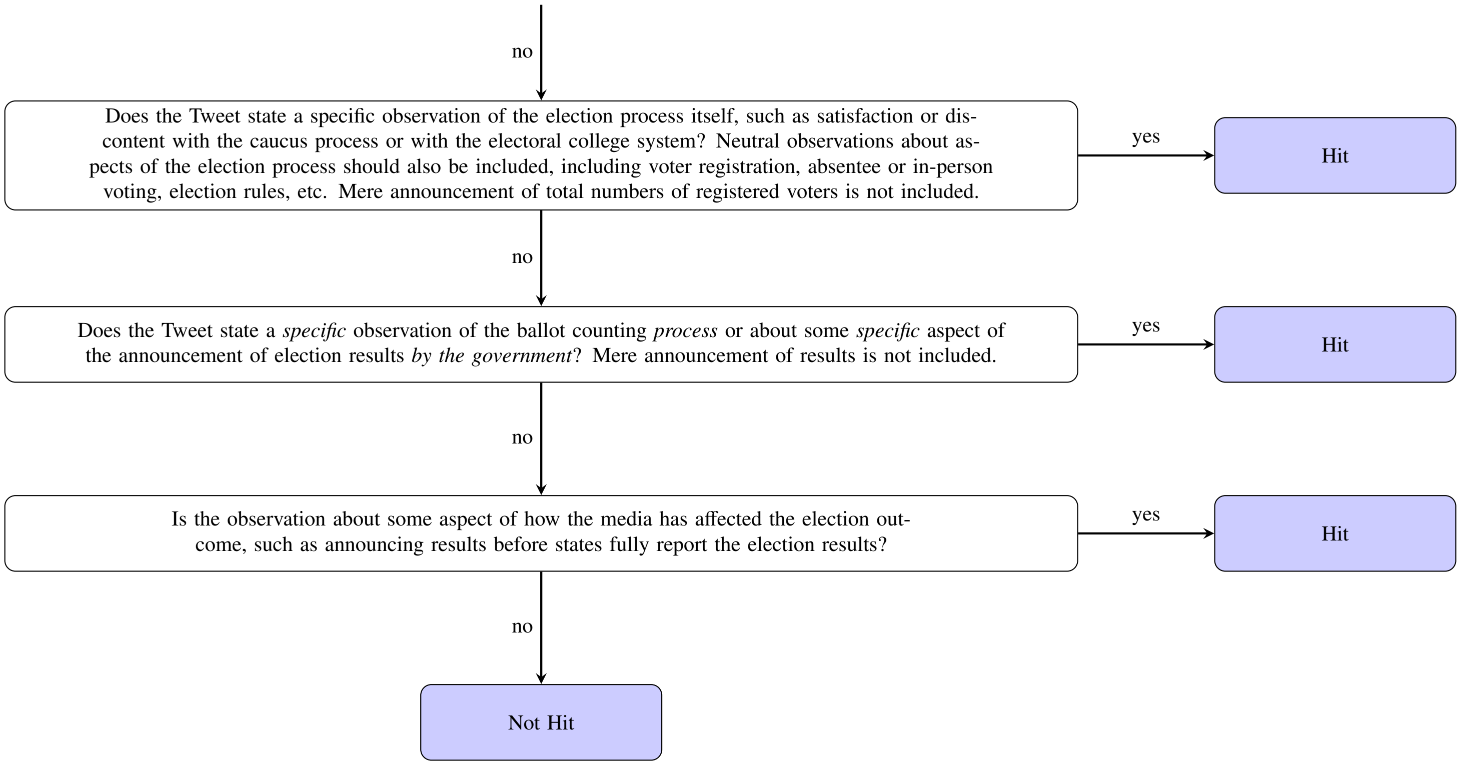
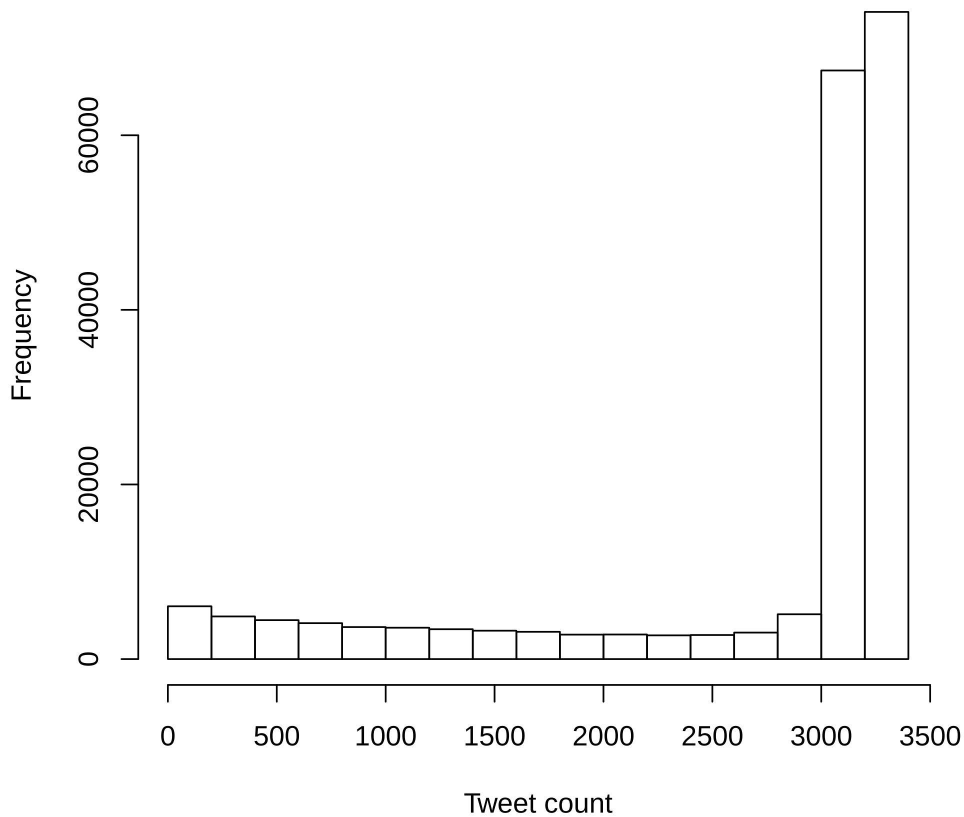


Figure 4: General Election Timelines

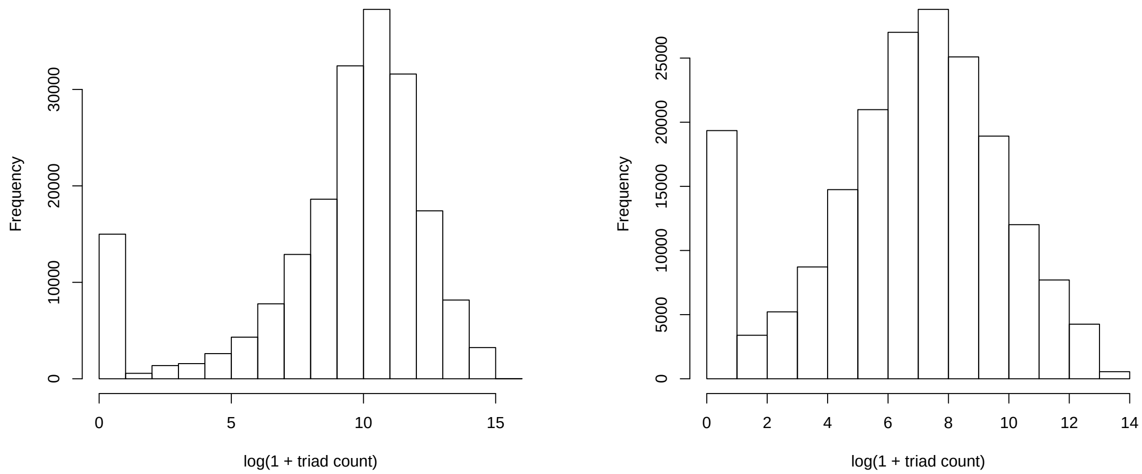
(a) Tweets in each timeline



Note: counts of Tweets in timelines for users who sent a Tweet classified as referring to an incident (total $n = 527,961,969$ Tweets).

Figure 5: General Election Triad Counts

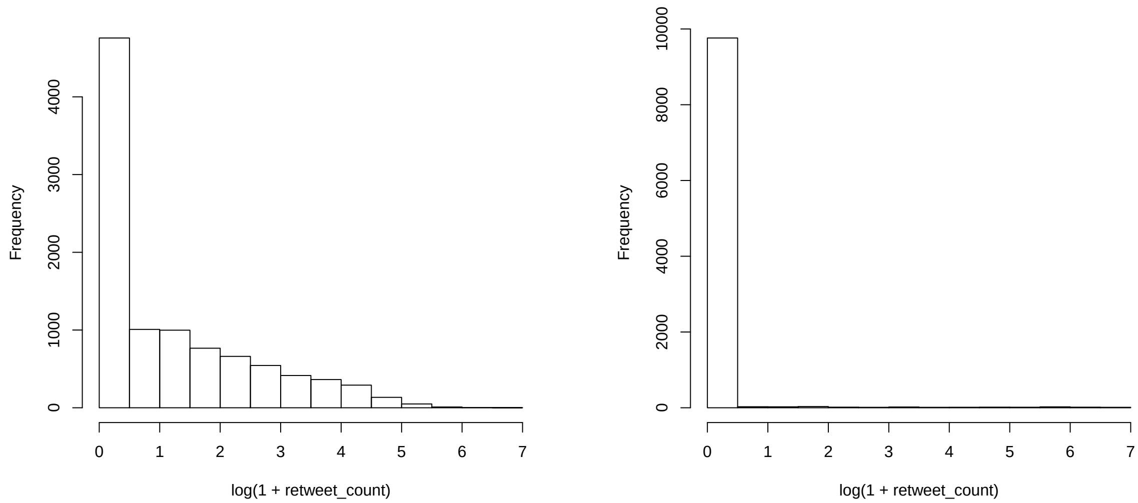
(a) user-friend-friend triads (b) user-follower-follower triads



Note: triads include only users, friends and followers who sent a Tweet classified as referring to an incident, and all three vertex IDs in a triad are unique (cyclical triads are not included). (a) natural log of counts of user-friend-friend triads per user (total $n = 21,243,891,283$ triads). (b) natural log of counts of user-follower-follower triads per user (total $n = 3,057,780,205$ triads).

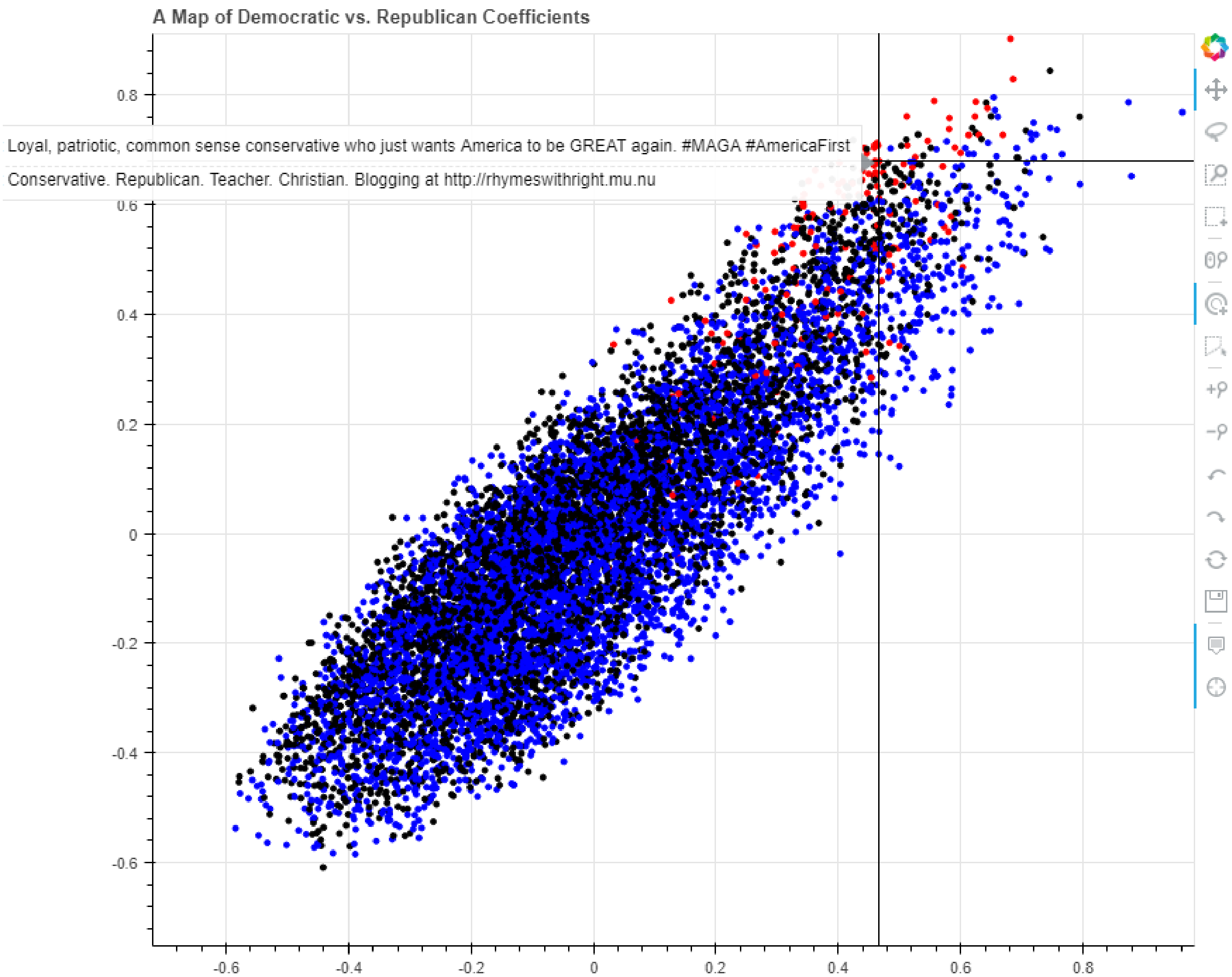
Figure 6: Counts of “Ideological” Retweets in Timelines

(a) liberal retweets (b) conservative retweets



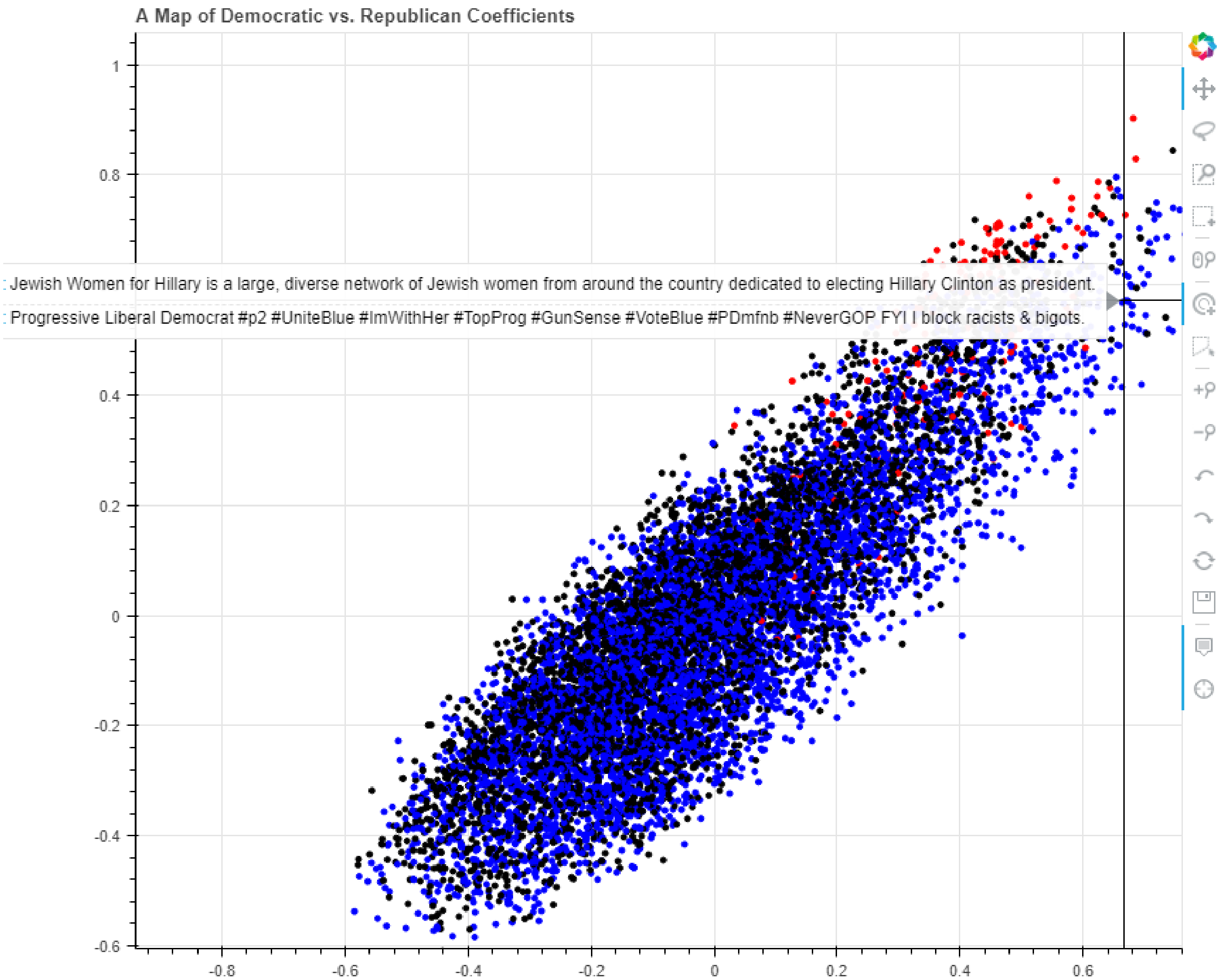
Note: counts per user of retweets of liberal or conservative accounts in timelines for users who sent a Tweet classified as referring to an incident (total $n = 10,000$ users), transformed using $\log(1 + \text{count})$.

Figure 7: Cosine Similarities: “Democrat” by “Republican”; Republican Illustrative Descriptions



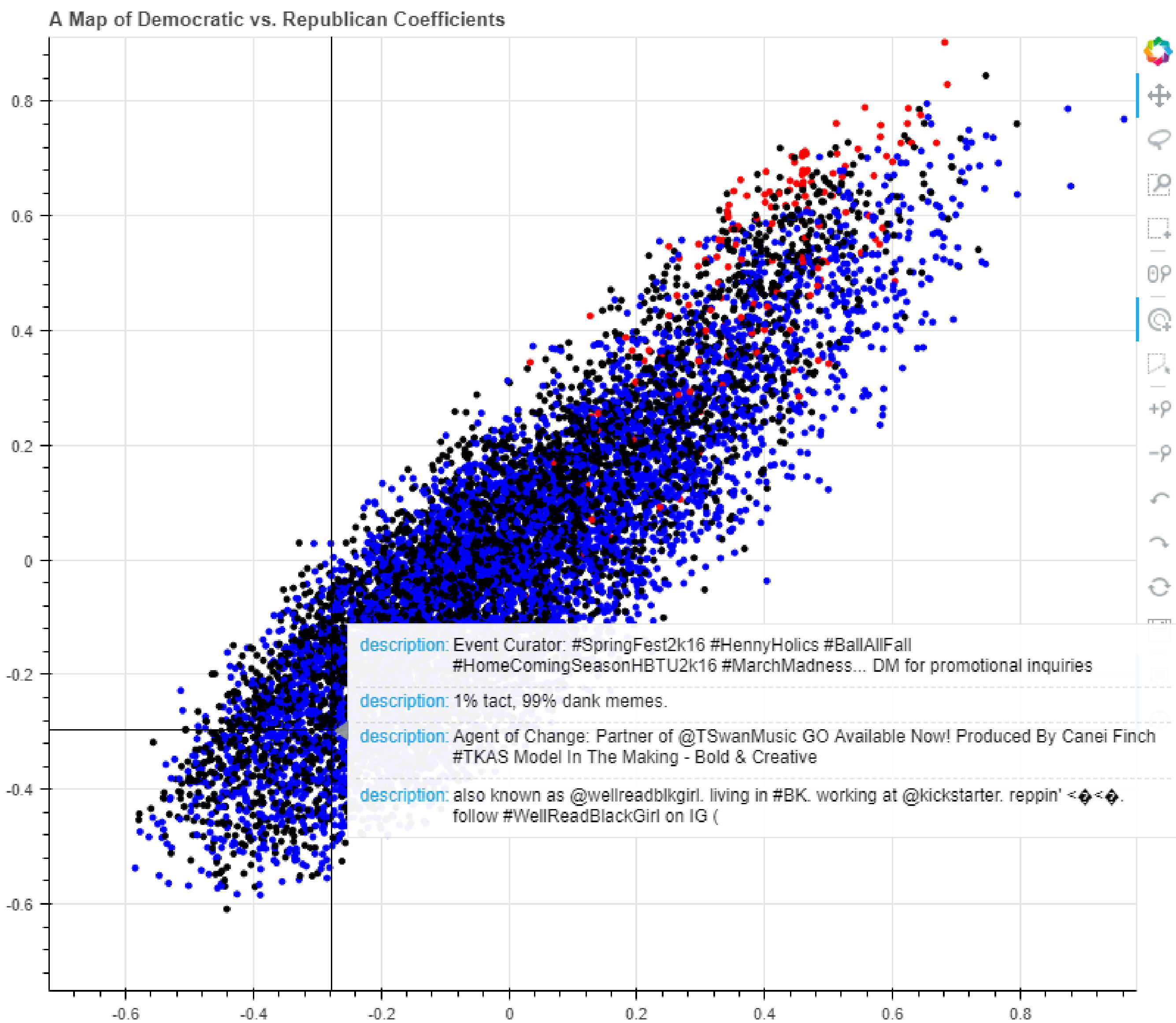
Note: scatterplot of “republican” and “democrat” cosine similarity coefficients for a sample of $n = 10000$ incident-Tweeting users. Colors indicate user timeline retweeting behavior: red, more conservative retweets; blue, more liberal retweets; black, no conservative or liberal retweets. Illustrative user descriptions come from a region with high cosine similarity to “republican.”

Figure 8: Cosine Similarities: “Democrat” by “Republican”; Democratic Illustrative Descriptions



Note: scatterplot of “republican” and “democrat” cosine similarity coefficients for a sample of n = 10000 incident-Tweeting users. Colors indicate user timeline retweeting behavior: red, more conservative retweets; blue, more liberal retweets; black, no conservative or liberal retweets. Illustrative user descriptions come from a region with high cosine similarity to “democrat.”

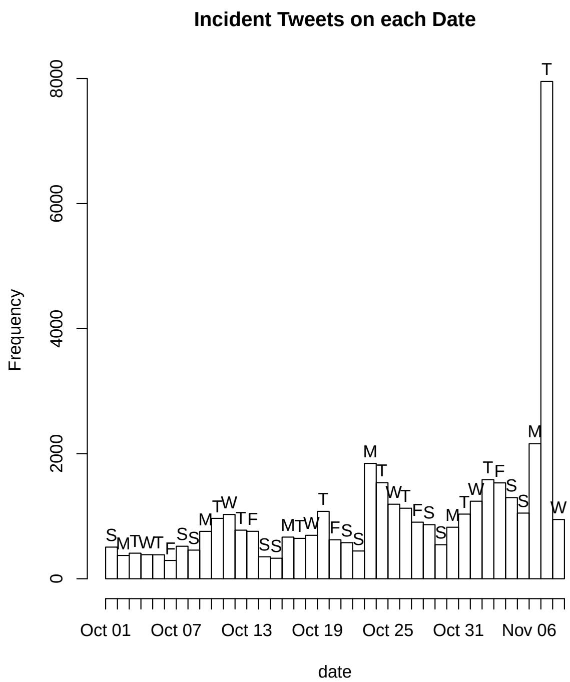
Figure 9: Cosine Similarities: “Democrat” by “Republican”; Nonpartisan Illustrative Descriptions



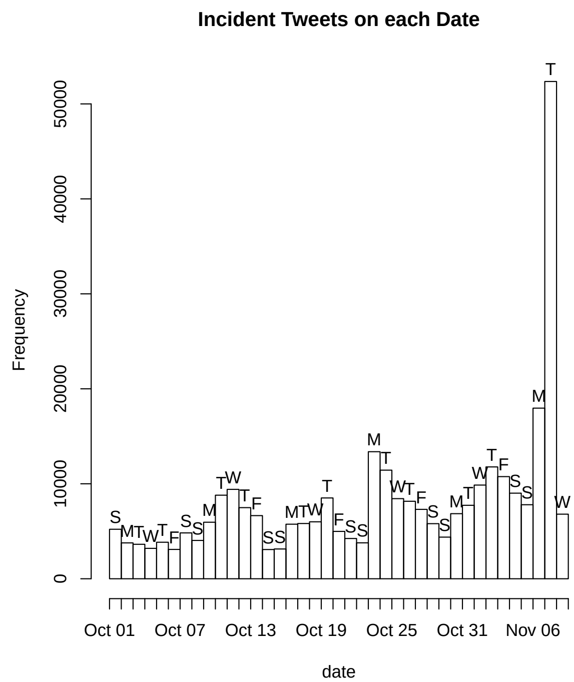
Note: scatterplot of “republican” and “democrat” cosine similarity coefficients for a sample of n = 10000 incident-Tweeting users. Colors indicate user timeline retweeting behavior: red, more conservative retweets; blue, more liberal retweets; black, no conservative or liberal retweets. Illustrative user descriptions come from a region that has about -0.3 cosine similarity with both “republican” and “democrat.”

Figure 10: General Election Incident Observations by Date

(a) Tweets with `place` information

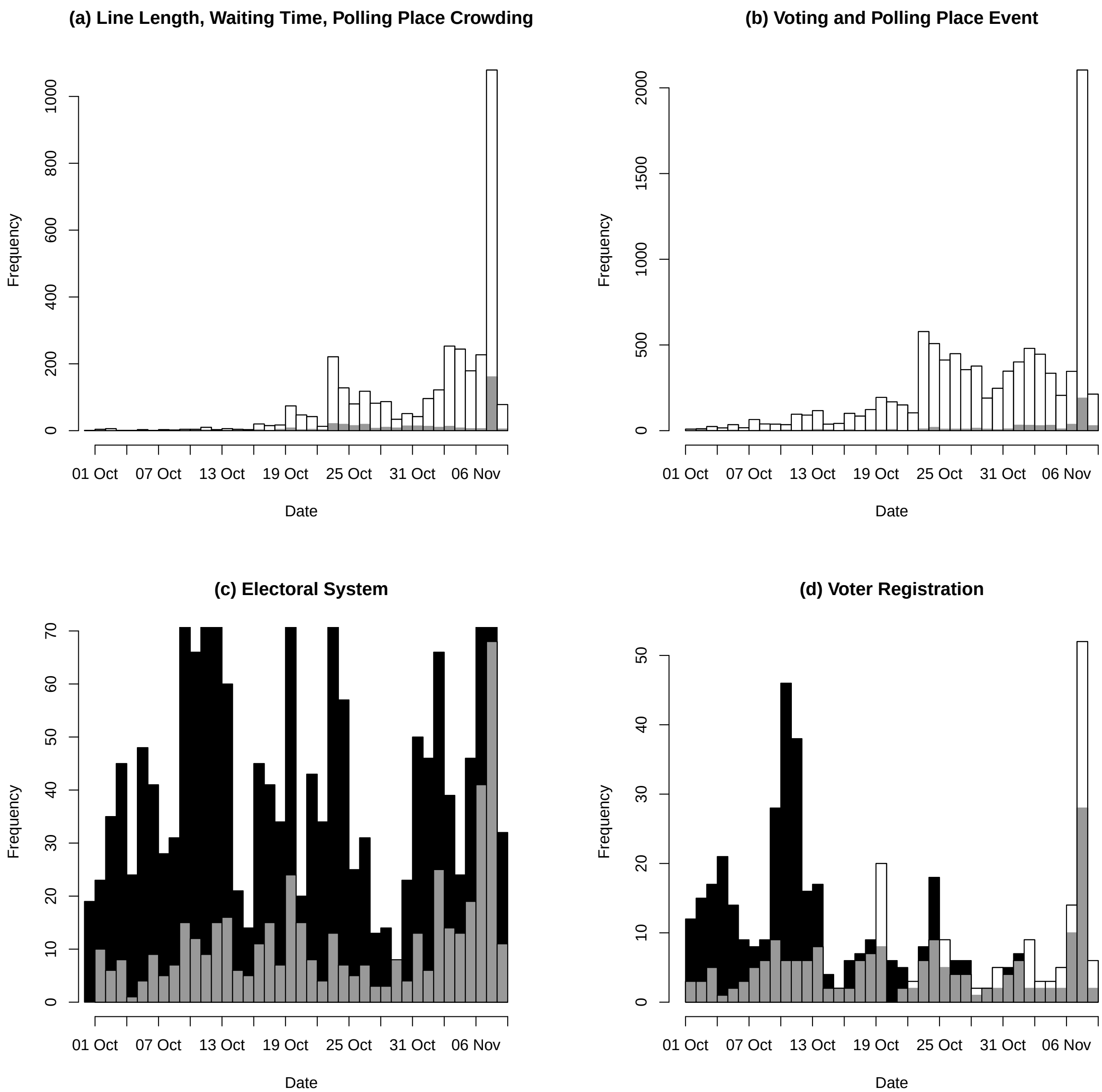


(b) all Tweets



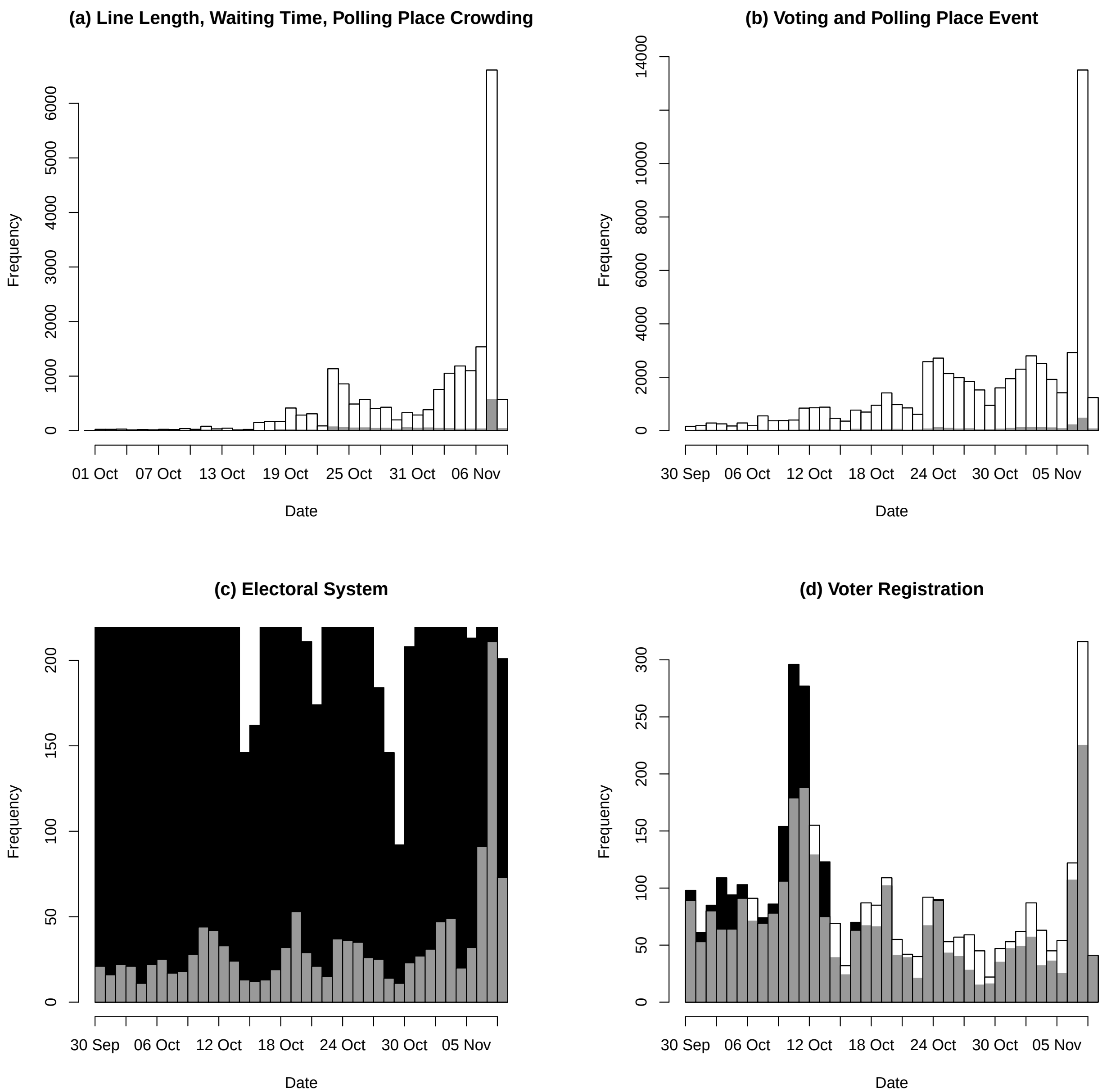
Note: (a) incidents by day for 40,678 Tweets with `place` information. (b) incidents by day for 315,180 Tweets with or without `place` information.

Figure 11: General Election Incident Observations by Type and Date



Note: using classifications for all Tweets that have `place` information. (a) white, long line or wait; black, no wait/line or short wait/line. (b) white, in-person, early or absentee voting success; black, voting problem. (c) white, election system problem; black, election system success. (d) white, voter registration problem; black, registration success.

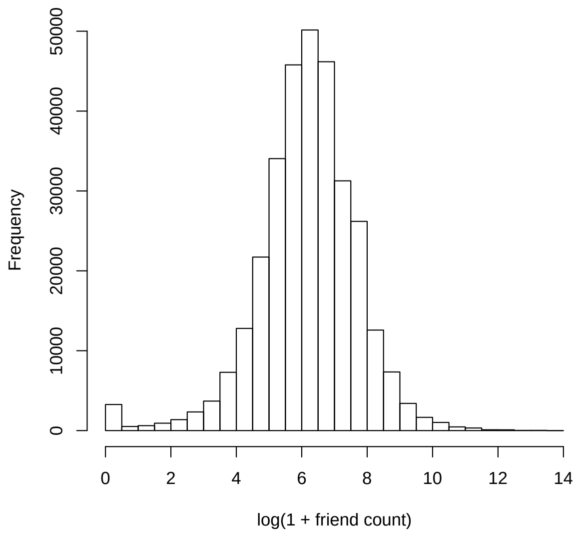
Figure 12: General Election Incident Observations by Type and Date



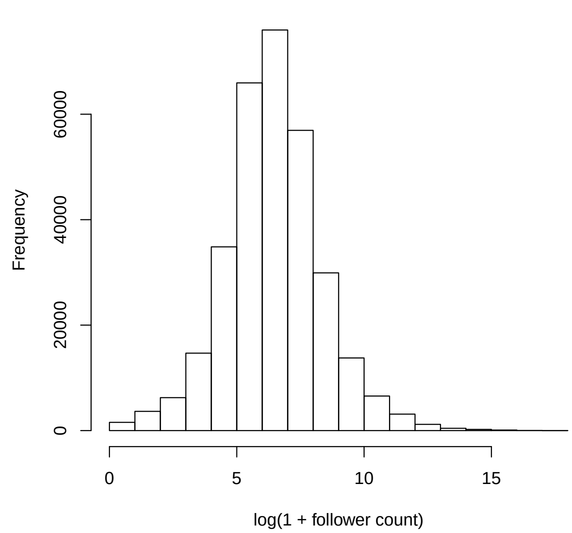
Note: using classifications for all Tweets with or without `place` information. (a) white, long line or wait; black, no wait/line or short wait/line. (b) white, in-person, early or absentee voting success; black, voting problem. (c) white, election system problem; black, election system success. (d) white, voter registration problem; black, registration success.

Figure 13: General Election Friends, Followers and Triads by Tweet

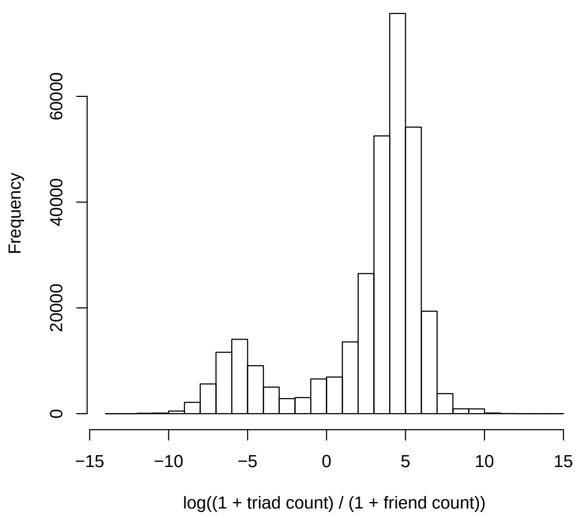
(a) friends



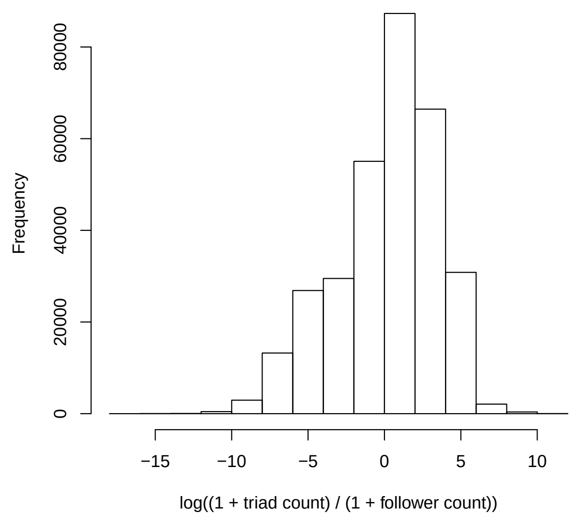
(b) followers



(a) user-friend-friend triads per friend



(b) user-follower-follower triads per follower



Note: counts are taken from each of the 315,180 Tweets that are classified as referring to incidents. Triads include only users, friends and followers who sent a Tweet classified as referring to an incident, and all three vertex IDs in a triad are unique (cyclical triads are not included). (a) natural log of `friends_count` values by Tweet. (b) natural log of `followers_count` values by Tweet. (c) natural log of count of user-friend-friend triads divided by `friends_count` value by Tweet. (d) natural log of count of user-follower-follower triads divided by `followers_count` value by Tweet.

Figure 14: Incident Types Nonparametrically Regressed on by-Tweet Friends and Triads

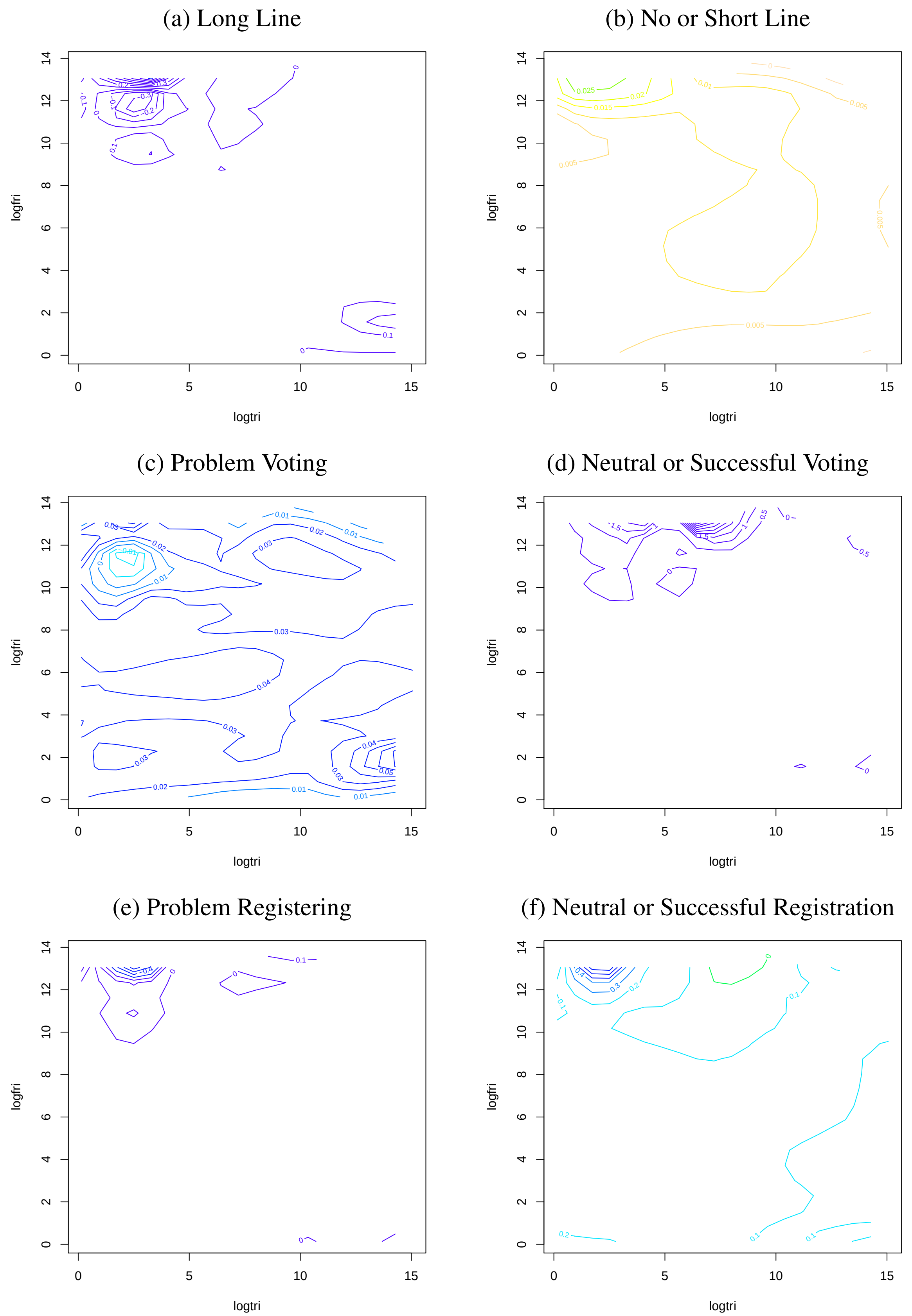


Figure 15: Incident Types Nonparametrically Regressed on by-Tweet Followers and Triads

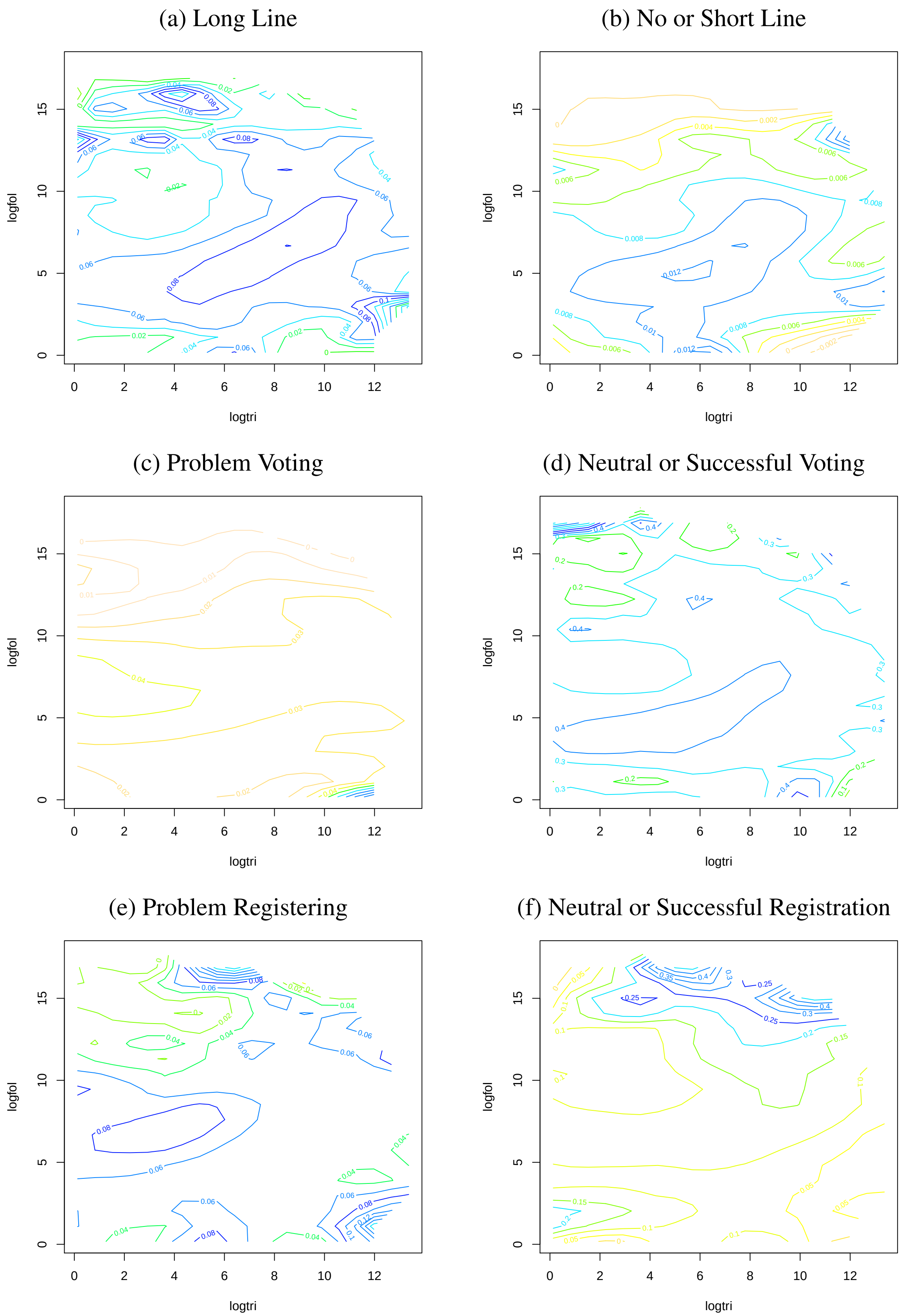
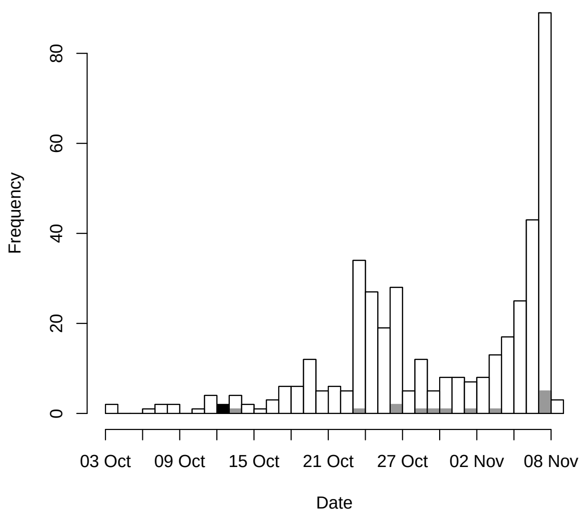
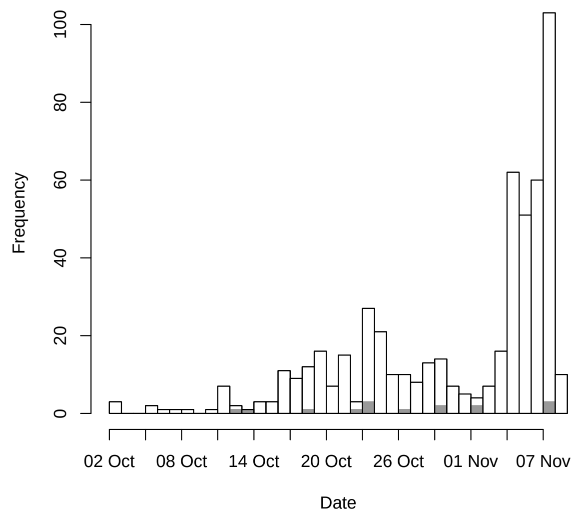


Figure 16: General Election Incident Observations by Type, Date and User Attributes

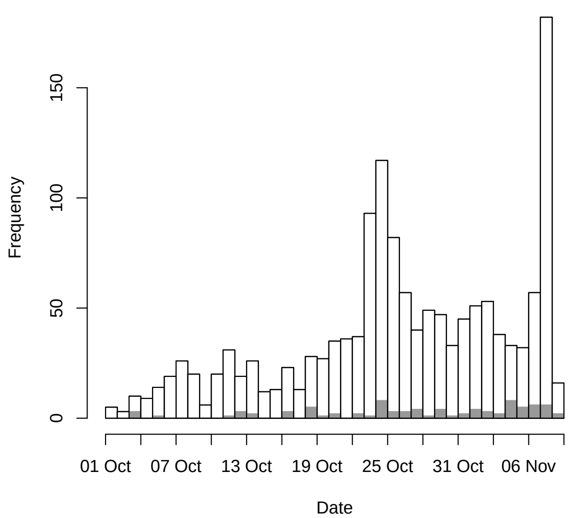
(a) Line Length, Polling Place Crowding: Trump



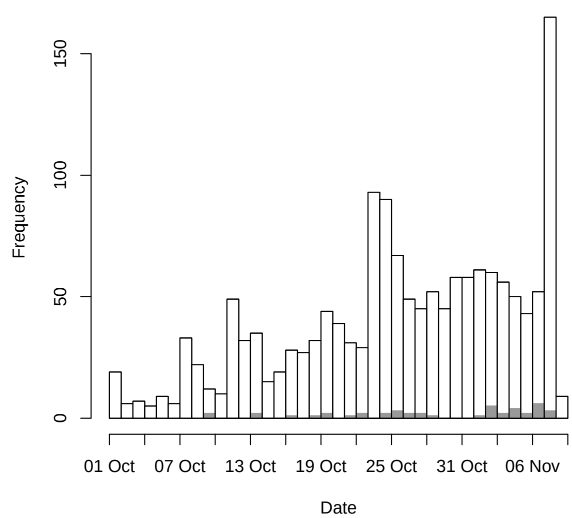
(b) Line Length, Polling Place Crowding: Clinton



(c) Voting and Polling Place Event: Trump



(d) Voting and Polling Place Event: Clinton



Note: using classifications for all Tweets. (a,b) white, long line or wait; black, no wait/line or short wait/line. (c,d) white, in-person, early or absentee voting success; black, voting problem.