Online Appendix For:

Trumping Hate on Twitter?

Online Hate Speech in the 2016 US Election Campaign and its Aftermath

Alexandra A. Siegel[*], Evgenii Nikitin, Pablo Barberá, Joanna Sterling, Bethany Pullen, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker[†]

Email: alexandra.siegel@colorado.edu or joshua.tucker@nyu.edu

This PDF file includes:

- Supplementary Text
- Figures A1 to A43
- Tables A1 to A30

List of Figures

[*] Corresponding Author

[†] Corresponding Author

List of Tables

A.1 Supplementary Text

A.1.1 Human Coding of Tweets

Human coders (both undergraduate student volunteers and paid coders on Crowdflower) were asked to code tweets by answering the following multiple choice questions when coding a smaller random sample of 5,400 tweets containing hate speech and white nationalist language dictionary terms. The remaining 20,000 tweets were coded answering only questions 1a and 2a in the interest of creating shorter, less expensive tasks:

1a) *Does this tweet contain hate speech, defined as "Bias-motivated, hostile and malicious language targeted at a person/group because of their actual or perceived innate characteristics, especially when the group is unnecessarily labeled (e.g., "jew bankers", "n– hooligans")." The tweet contains hate speech if slurs or derogatory language are used toward any person or group regardless of whether or not the target is a member of the slur-target group. (e.g. a tweet calling Trump a "f**" or "n*****" still contains hate speech).*

- Yes
- No
- Unclear

1b) If yes:
Is the hate speech directed at...

- Donald Trump
- Hillary Clinton
- Barack Obama
- One or more of Trump's Family Members
- Another Democratic Politician
- Another Republican Politician
- Liberals or Democrats in general
- Republicans or Conservatives or Trump Supporters in general
- None of the above

1c) If no:
Is the tweet denouncing or condemning the use of hate speech?

- Yes

- No
- Unclear

2a) *Does this tweet contain white nationalist rhetoric? This includes any rhetoric or content that praises known white-nationalist groups, shows excessive pride in the white race, puts forth white supremacist or white separatist ideologies, or focuses on the alleged inferiority of nonwhites (eg. #whitegenocide, #whitefamily, #whitevalues)?*

- Yes
- No
- Unclear

2b) If no:
Is the tweet denouncing or condemning the use of white nationalist rhetoric?

- Yes
- No
- Unclear

Our random sample of human coded tweets included up to 3,000 tweets from each of our nine categories, for a total of 25,000 tweets. After initially using trained undergraduate coders to ensure that tweets could be accurately classified, crowd-sourced coding was done using Figure8 (formerly Crowdflower), a data enrichment platform that allows a researcher to launch microtasks to a "crowd" of over five million contributors. For a recent overview of how to use Figure8 in political science research, see (**?**). Tweets were each labeled by three coders. Test questions for quality control ensured that the contributors coding tweets were responding to tasks truthfully and conscientiously. If a contributor answered test questions incorrectly, he or she was removed from the job and his or her data was erased.

A.1.2 Classifying Tweets (Dictionary-Based Method)

Before training Naive Bayes classifiers, we pre-processed the data by stemming and lowercasing words, and removing punctuation except for Twitter relevant symbols (@ and #). Using 70% of the data as a training dataset and 30% as a test dataset from our random sample of about 25,000 human coded tweets containing hate speech and white nationalist dictonary terms, our hate speech classifier performed with 94% accuracy (true positives + true negatives/ total cases); 95% precision (true positives/true positives + false positives) and 90% recall (true positives/true positives + false negatives). Given that our dictionary method brought in as many hate speech tweets as non-hate speech tweets, this classification represents a marked improvement over a pure dictionary method. All categories of hate speech were trained using the same classifier because so many tweets contain hate speech directed at a variety of targets.

Because white nationalist rhetoric was much less common in our sample of 25,000 human coded tweets, the precision and recall of our classifier were a bit lower, though overall accuracy was quite high. Our white nationalist classifier performed with 97% accuracy (true positives + true negatives/ total cases); 71% precision (true positives/true positives + false positives) and 88% recall (true positives/true positives + false negatives). This still represents an improvement over the pure dictionary based method, though it did not perform quite as well as our hate speech classifier.

A.2 Additional Results

A.2.1 Top Hate Speech Dates

While our initial goal was to examine whether or not we observed a more sustained Trump effect either following his election or over the course of the campaign, it is possible that particular campaign events or shifts in Trump's rhetoric might be producing shorter term Trump effects throughout the period under study that may have contributed to the perceived increase. Tables A4, A5, and A6 in the Appendix display the top 10 dates with the highest proportion of hate speech and white nationalist rhetoric in each dataset, capturing the largest spikes in the data.

In the Clinton Dataset, the largest spike in hate speech occurred on April 14, 2016 during the last debate of the Democratic primary between Bernie Sanders and Hillary Clinton. Looking at spikes in the data by hate speech type, this spike is primarily driven by an uptick in misogynistic hate speech during the debate. The next three largest spikes in the Clinton dataset occurred on December 16, 2016 when Clinton first publicly addressed the impact of Russian hacking on the 2016 election, on December 24, 2016 when Trump praised Putin for attacking Clinton, and on January 3, 2017 when the Clintons announced they would attend Trump's inauguration. Again these spikes are primarily driven by spikes in misogynistic hate speech, which is the most prevalent type of hate speech in our political datasets. The next largest spikes include the November 22, 2015 rally of Democratic primary candidates, the second debate between Trump and Clinton on October 9, 2016, the release of the Comey Letter about Clinton's emails on October 28, 2016, and September 25, 2016 during the first Clinton-Trump debate. This brief examination of the largest spikes in the Clinton data suggest that contentious political events in general—rather than a Trump effect—are driving spikes in hateful speech in the Clinton data, much of which may have been directed at Clinton herself.

In the Trump Dataset, the three largest spikes in hate speech occurred on January 31, 30, and February 1, 2017, after Donald Trump fired acting US attorney general Sally Yates when she told justice department lawyers not to defend his travel ban executive order. This period was also the first time Clinton spoke out after the election, criticizing the travel ban. These daily spikes are captured by the large monthly increase in hate speech that we display in Figure 3 in late January/early February 2017. As we highlight in the manuscript, this was largely driven by an increase in misogynistic language at least in part directed at Yates and Clinton. The next largest spikes occurred about a week later on February 8 and 9 2017 when Trump's travel ban was overruled. The next three largest spikes occurred on January 29, 30, and 31, 2016 after Trump held his own rally on January 28, 2016 instead of participating in the GOP primary debate. Another spike occurred on July 13, 2015 after Trump made inflammatory comments about Mexican immigrants. While this spike is perhaps indicative of a brief Trump effect, the other major spikes in the Trump data appear to correspond to moments of political contention rather than a response to particular rhetoric or actions by Trump on the campaign trail.

In the random sample data, the largest spike in the data occurred on February 9, 2017, the day Trump's travel ban was overruled in the San Francisco court of appeals. There are also a series of spikes in June 2017 largely driven by anti-Muslim tweets following the June 2017 London terror attacks and the March against Sharia Law protests in mid June 2017. We also see a spike in late May 2017 following the Portland train attack when a white nationalist stabbed three people shouting anti-Muslim slurs. Finally we see a spike in misogynistic language in October 2016 following a Trump-Clinton debate.

Together, the majority of these spikes in hateful language do not appear to be explicitly connected to Trump, but rather to times of political contention when incivility may be more likely online as well as in the aftermath of exogenous events like terror attacks that tend to be accompanied by a rise in hate speech. While certainly consequential, these spikes do not appear to correspond to a pattern of sustained popularity or a larger number of spikes over time. Instead, the daily proportion of hate speech seems to shift in response to particular political and non-political events and then to re-equilibrate quickly. These general patterns persist whether we are examining the daily volume or proportion of tweets, as well as number or proportion of unique users. Again, here we have chosen to include retweets to better capture the popularity of tweets. When we exclude retweets most of the spikes that we observe disappear. This is evident in Figures A7, A8, and A9 in the supplementary materials.

A.2.2 Reddit-based (Non-Dictionary) Robustness Analysis

One of the potential pitfalls of using dictionary-based methods for identifying hate speech—no matter how sophisticated the application of these approaches—is that they force the analyst to rely on a corpus of words used in the"past" to code speech in the present. In most contexts, this is unlikely to be problematic, due to the long life span of slurs and derogatory language. However, given our surprising finding that hate speech and white nationalist rhetoric did not increase persistently either over the course of the 2016 election campaign or in its aftermath, we must seriously consider the fact that we have somehow failed to identify a significant subset of hate speech on Twitter.

There were many alt-right code words with alternative meanings used in the period under study. For example, "'googles' signified the n-word; 'skypes' referred to Jews (k****); and 'yahoos' was used in place of 's***"'(**?**). A dictionary-based method that did not contain these code words would therefore be missing the occurrence of hate speech. Perhaps even more problematically, if an event both led to an increase in hate speech *and* to the use of new code-words for hate speech, we would completely miss the impact of this event. As an aside, there is an existential question here as to whether hate speech that uses sanitized words ought to be considered hate speech at all. Addressing this issue is beyond the scope of the current paper, although it seems that if everyone knows what is being referred to by"kill all the skypes" then it is difficult to see why this would not be considered hate speech. For now, though, we simply note that the alternative approach described in the remainder of this section is flexible enough to pick up any instances of hate speech that might be missed by dictionary methods, regardless of whether they rely on hateful words that are simply not in the dictionaries we are using or if they represent some sort of new vocabulary—coded or not—for depicting hateful speech.

With this concern in mind, we repeat our analyses using a non-dictionary-based classification method to measure the prevalence of hateful language. The main idea underlying this alternative method is to find an example of "hate speech in the wild," or a large collection of labeled text that contains the types of hostile rhetoric people actually use online. For this task, we rely on publicly available comments posted on Reddit.com for our reference corpus. Reddit.com is a popular news aggregation and discussion web-site. Because Reddit entries are organized into forums with specific topics of interest—"subreddits"—they are already classified for us. Some of these subreddits are infamous for their explicitly racist, hateful and extreme alt-right content. Examples of these subreddits include /r/Coontown, /r/WhiteRights, /rAntiPOZi, and /r/european. Many of them were eventually banned or quarantined by the Reddit administration, but the comments that were posted in these subreddits are still available for downloading.

To the extent that these subreddits contain "real world" hate speech, we can measure how much of this rhetoric is present on Twitter by examining the probability that the language in our tweet collections might be found in a particular subreddit or group of subreddits. Now there will of course be differences in how speech is used across different platforms, so any absolute measures of these probabilities are difficult to interpret. However, *relative* measures—such as whether tweets after Trump's election were more similar to alt-right subreddit content than tweets at the beginning of the campaign—help us assess whether or not hate speech increased on Twitter over time. Indeed, we can again leverage daily Twitter data, just as we did with our machine-learning-augmented dictionary methods. But this time, instead of comparing the proportion of classified tweets containing slurs, we assess the likelihood of text from our tweet collections on a given day appearing on white nationalist subreddits.

Reddit's structure has two important properties that were particularly useful for developing and testing the proposed method:

1. Users subscribe to subreddits that they find interesting and only see content posted in these subreddits on their main page. It leads to emergence of the echo-chambers, where users mostly communicate with people who share their views. Even though redditors (Reddit users) who do not support alt-right views can still post comments in the alt-right subreddits, the number of such comments is relatively small.

2. We can further reduce noisiness of the data by filtering out comments with negative

user rating. Redditors upvote and downvote comments depending on how they feel about ideas expressed in these comments. For example, members of the alt-right communities tend to downvote comments that go against their ideology. Figure A38 shows two comments posted in one of the threads in */r/DebateAltRight* subreddit with negative and positive ratings respectively.

A.2.3 Reddit Model

The intuition behind the method proposed here lies in replacing manual labeling of individual documents by using large naturally annotated corpora of text, such as Reddit comments. For example, it does not take much effort and time to determine that */r/soccer* is mainly dedicated to discussing soccer, even though each individual comment is not necessarily directly related to soccer. On average, soccer is going to be a much more popular topic of discussion in */r/soccer* than in, say, */r/movies*. The same logic applies to political subreddits. Donald Trump's political views are much more popular in */r/The_Donald* than in */r/hillaryclinton* or */r/EnoughTrumpSpam*. Any kind of naturally annotated text corpus that we have some background knowledge about can potentially be used—newspaper articles, party manifestos, politicians' tweets, etc. The choice of training data depends on the nature of the task at hand.

Now imagine that we want to analyze how the popularity of soccer among Twitter users changed over time compared to other topics of interest. In order to do that, we can train a classifier that outputs the probability that a particular document (tweet) belongs to a particular group (subreddits about soccer). This probability measures how semantically similar this document is to this subreddit (or group of subreddits) as compared to other subreddits. To put it differently, we check whether words and phrases used in the tweet are similar to words and phrases that are popular in the different subreddits. We do not need to explicitly provide a dictionary of soccer-related terms and phrases to the model as it can automatically learn them from the corpus of comments. This method seems like a particularly good fit for measuring popularity of the alt-right movement, considering that there are plenty of communities on Reddit and other online platforms that openly declare their alt-right views.

Generally speaking, any kind of supervised classifier can be used to apply this method. In this paper, we used a supervised version of fastText model (**?**), which is conceptually similar to the skip-gram version of word2vec. Unlike word2vec, instead of learning word representations in an unsupervised fashion, fastText updates these embeddings to optimize the particular text classification task. Here, we trained the model to predict which subreddit (or a group of subreddits) each of the comments in the corpus was posted in. The model learned semantic commonalities of comments in each subreddit. An architecture of the model is shown in Figure A39.

A single document (e.g., "Make America Great Again"), which is split into tokens, constitutes an input layer of the model. Each word in the document is initialized with a random vector representation (word embedding). These representations are stored in an embedding matrix of size $K \times n$, where K is the number of unique words in vocabulary and n is the dimensionality of the model (say, 100). A vector representation of the document is then calculated as a component-wise mean of the token vectors (”make” + ”america” + ”great” + ”again”). This sentence vector constitutes a hidden layer d of length n. Vector d can be interpreted as a vector representation of the current document. This vector is then multiplied by model weights matrix U (also initialized randomly) in order to obtain an output activation for each class (e.g., */r/The_Donald, /r/hillaryclinton, /r/StarWars*). Matrix U has dimensionality $m \times n$, where m is the number of possible classes (3 in the current example). A softmax activation function is then applied to the output activations to make sure that final class activations sum up to 1. These activations can be interpreted as predicted probabilities for each class.

Parameters of this model that are learned during training phase are (a) embedding matrix and (b) weights matrix U. The embedding matrix stores word representations that were optimized for a classification task. These representations can later be used for computing vector representations for new documents. For this project, we are also interested in matrix U. The rows of this matrix can be thought of as embeddings of the classes (**?**). The classes that are semantically more similar to each other should have vectors that are closer to each other in the vector space.

After training the model to predict which subreddit (or group of subreddits) comments were posted in, it can be used to calculate class probabilities for each tweet in our collection. Finally, average daily probabilities can be calculated for each class, and changes in these probabilities will show us the dynamics of the relative popularity of a particular topic on Twitter.

A.2.4 Reddit Data

The main source of data for this analysis is Reddit.com. Reddit comments were obtained via Google BigQuery platform[1]. The analysis presented in this paper uses Reddit comments posted from November 2016 to July 2017.

Since the model has to be able to distinguish between a variety of different topics and sentiments, a diverse dataset should be used for training. Here we selected 65 subreddits that covered all relevant political ideologies and movements (liberal, conservative, alt-right, pro-Trump, anti-Trump, feminist, etc.). We also included a variety of non-political subreddits in the dataset. The full list of subreddits is displayed in Figure A40.

All comments posted in selected subreddits from November 2016 to August 2017 were initially included in the dataset.[2] We chose this time frame because many relevant subreddits were created after or just before the election day. We then applied the following procedure to the dataset:

- Comments with negative user scores (more downvotes than upvotes) were removed from the dataset.

[1] https://bigquery.cloud.google.com/dataset/fh-bigquery:reddit_comments

[2] Anti-black subreddit */r/CoonTown* was banned in 2015, but it was included in the dataset as a separate category

- Popular subreddits (more than 100,000 comments) were downsampled to tje top 100,000 comments with the highest ratio $\frac{score}{number\ of\ comments\ in\ the\ thread}$ (most upvoted comments are more likely to accurately represent content from each subreddit).

- All comments posted by Reddit bots and comments and tweets which contained less than two relevant tokens were removed. Additionally, all names of subreddits (and their common abbreviations), usernames, and URLs were removed. Numbers, dates, and times were replaced by special tokens "NUMBER", "DATE", and "TIME" respectively. All quotes were removed to prevent "cross-contamination" (e.g., comments and posts from other subreddits are often quoted in order to make fun of them). Stemming, stop-words removing, punctuation removing, and lowercasing preprocessing steps were tuned as hyperparameters of the model. For all tweets used for validation and analysis, a hashtag splitting procedure was performed as a lot of information is often contained in the hashtags (*#makeamericagreatagain* becomes *make america great again*).

- Non-English comments and tweets were removed from all datasets by applying trigram-based language detection procedure (**?**).

Hyperparameter and preprocessing pipeline tuning was performed by a grid search procedure. Cross-validation log-loss score (5 folds) was computed for each unique combination of parameters, and the model with the highest score was chosen as the final one.

A.2.4.1 Reddit Validation Data

Two additional Twitter datasets were used to demonstrate how well the model was able to distinguish between different topics and sentiments. First, we should expect tweets labeled as hate speech using our dictionary-based method to be more similar to the alt-right group of subreddits. Hereafter, this dataset is called **hate speech collection**. Second we used the Twitter API on November 11, 2017 to collect last 3,200 tweets from the Twitter accounts of politicians, journalists, civil rights activists, and prominent alt-right figures (**Popular collection**). The full list of the accounts with their descriptions can be found in Table A25.

1. AntiSJW – anti-"social justice warriors" subreddits that make fun of popular liberal ideas such as social justice and new-wave feminism (*/r/CringeAnarchy, /r/SocialJusticeInAction, /r/sjwhate, /r/KotakuInAction, /r/4chan, /r/whiteknighting*).

2. Anti-alt-right – subreddits that oppose ideas popular in the alt-right movement (white nationalism, racial realism, etc.) (*/r/altright, /r/DebateAltRight, /r/AntiPOZi, /r/WhiteRights*)

3. */r/CoonTown* – infamous anti-black community that was banned in 2015

4. Alt-right – subreddits that openly declare their alt-right and white nationalist views

5. */r/uncensorednews* – subreddit with strong anti-immigration views

6. Conspiracy – subreddits where users discuss conspiracy theories, such as PizzaGate or 9/11 conspiracies

7. */r/The_Donald* – one of the largest pro-Trump online communities (over 500,000 subscribers); infamous for its radical anti-immigration and anti-progressive views

8. Liberal – liberal and Democrat subreddits (*/r/SandersForPresident, /r/hillaryclinton, /r/Liberal, /r/democrats*)

9. Anti-Trump – subreddits where large proportion of submissions is dedicated to opposing Donald Trump and his policies (*/r/esist, /r/politics, /r/EnoughTrumpSpam*)

10. Conservative – conservative and Republican subreddits (*/r/Conservative, /r/Republican, /r/AskTrumpSupporters, /r/AskThe_Donald*)

11. Misc – non-political subreddits (*/r/AskReddit, /r/funny, /r/food, /r/AnimalsBeingJerks, /r/gaming, /r/EarthPorn, /r/Fitness, /r/science, /r/books, /r/movies, /r/Music, /r/anime, /r/news, /r/technology*)

12. Sport – subreddits about sport (*/r/sports, /r/soccer, /r/baseball*)

13. Black – subreddits popular among African Americans (*/r/BlackPeopleTwitter, /r/Blackfellas, /r/blackladies*)

14. Anti-feminist – anti-feminist subreddits (*/r/TheRedPill, /r/MGTOW*)

15. LGBT – pro-LGBT subreddits (*/r/lgbt, /r/ainbow*)

16. Feminist – feminist subreddits (*/r/AskFeminists, /r/Feminism, /r/socialjustice101, /r/GenderCritical, /r/TwoXChromosomes*)

17. Religion – subreddits dedicated to discussions of topics about religion (*/r/islam, /r/Judaism, /r/Christianity*)

A.2.5 Reddit Model Validation

A.2.5.1 Subreddit Representations

Subreddit embeddings which are contained in matrix U can be used to facilitate the process of grouping subreddits into categories. This grouping is important for two main reasons. First, some subreddits, especially in important categories such as alt-right and anti-Trump subreddits, contain relatively few comments (less than 30,000). This could potentially affect predictive performance of the model. Second, it is undesirable for the model to learn semantic peculiarities of specific communities (e.g., slang and abbreviations used by members of particular subreddit). Instead we want the model to learn semantic constructions that are common among all people with similar ideologies and interests.

Subreddits that are semantically similar to each other should have vector representations that are closer to each other in the vector space of the model. In order to visualize the final vectors, we applied a hierarchical clustering procedure. All subreddits start off in the separate clusters in the beginning of the procedure. At each step, the two most similar clusters are merged together. Between-cluster distance can be calculated in many different ways. Here, we measured it as an average cosine distance between subreddit vectors. Figure A40 plots results of the clustering procedure as a dendrogram. The x axis represents cosine distance, so points where clusters of subreddits merge together show average cosine distance between these two clusters.

As A40 suggests, the model was able to learn some meaningful semantic differences between subreddits. For example, all explicitly alt-right subreddits (*/r/altright, /r/DebateAltRight, /r/WhiteRights, /r/AntiPOZi,* etc.) ended up having vectors that are quite close to each other in the vector space of the model.

In addition, we applied a t-SNE dimensionality reduction technique **?** on subreddit vectors to visualize semantic relationships between subreddits. Figure A41 shows that clustering of subreddits mainly matches our expectations.

Based on this clustering, background knowledge, and reviewing the sidebars, rules, FAQs, submissions, and comments, we grouped 65 subreddits into the 17 categories displayed in Figure A42.

A.2.5.2 Popular Collection Validation

How can we demonstrate that learned semantic differences between groups of subreddits are not Reddit-specific and can be successfully applied to measure trends on Twitter? One way to do so without coding individual tweets is to compile a list of popular Twitter accounts with easily identifiable ideological leanings. For example, the trained model should on average classify Richard Spencer's tweets as alt-right since he is one of the leaders of the alt-right movement, and the vast majority of his tweets are devoted to popular issues within the alt-right community.

To conduct this analyiss, we used the Twitter API in November 2017 to obtain the most recent 3,200 tweets from 48 popular Twitter accounts, displayed in Table A27. We then calculated predicted probabilities of belonging to one of the groups of subreddits for each tweet. We then averaged these probabilities over Twitter accounts. Each Twitter account was assigned a label corresponding to the highest average predicted probability. Table A27 displays these classification results.

Overall, classification results are in line with what we would expect. Most importantly, all explicitly alt-right accounts get classified as belonging to the "Alt-right" group. There are a few accounts that probably do not get classified correctly (e.g., *@SenJohnMcCain* classified as */r/The_Donald*), but their number is very small.

A.2.5.3 Hate Speech Collection Validation

We can also use our dictionary-based data to validate the quality of the model. It is safe to assume that alt-right tweets on average contain more racial slurs and hate speech towards minorities than non-alt-right tweets. Therefore, tweets that were coded as hateful in hate speech collection should on average receive higher probability of belonging to the Alt-right group than non-hateful tweets. Figure A43 shows results of the validation.

Figure A43a plots average predicted probabilities of belonging to each of the classes for tweets that were coded as hateful by the trained coders. Figure A43b shows the difference in average predicted probabilities between tweets that were coded as hateful and tweets that were coded as non-hateful. Alt-right, alt-lite, and anti-immigration groups of subreddits are printed in red. These figures suggest that hateful tweets are semantically much more similar to these groups of subreddits than to the other groups.

A.2.6 Additional Analysis of Trump, Clinton, and Random Sample Collections

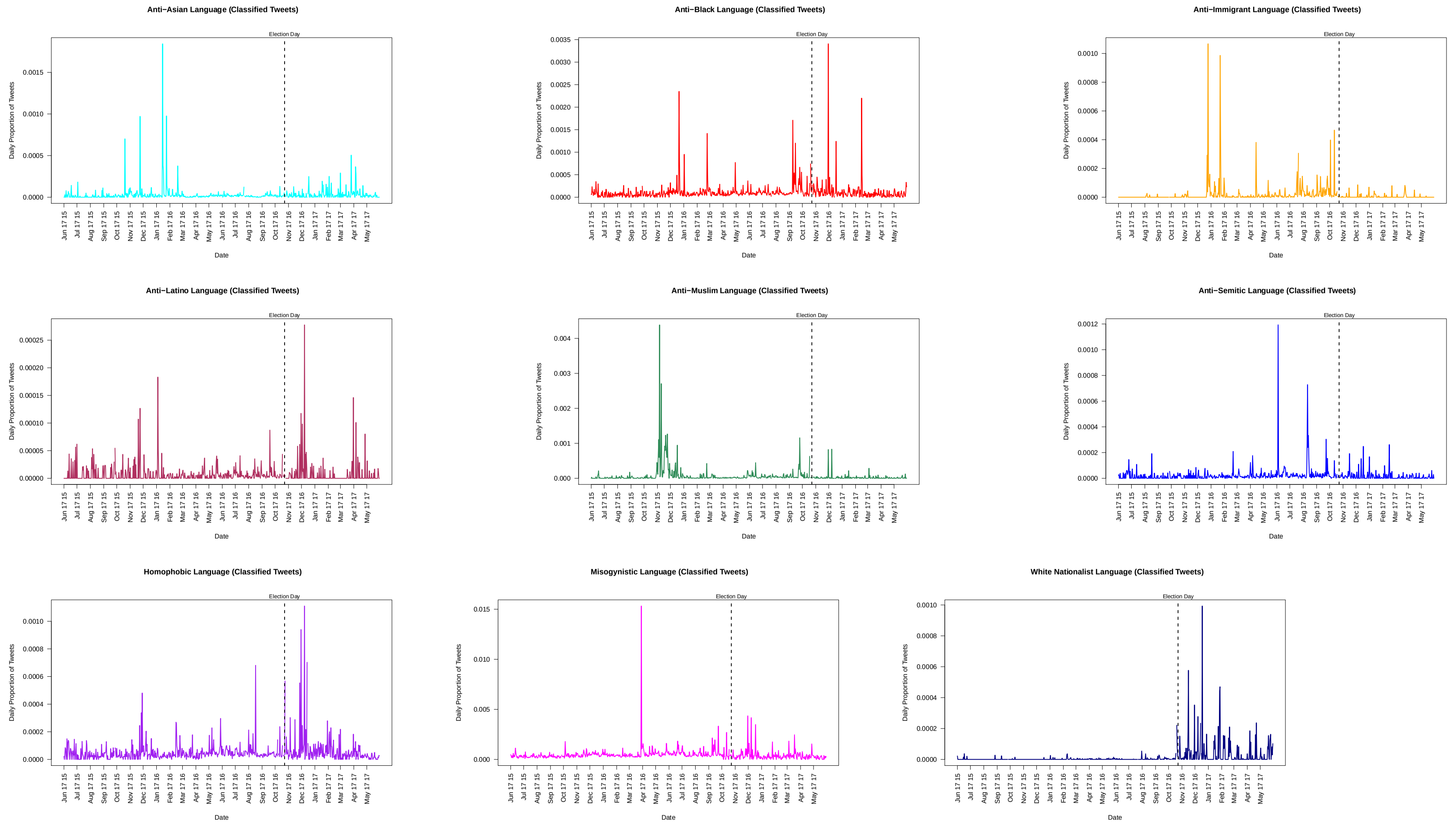
In order to provide a non-dictionary based test the degree to which hate speech or white nationalist increased over the course of the 2016 election campaign or following Trump's election, we examine the average daily probabilities that tweets in our Trump, Clinton, and random sample Twitter collections might appear in a cluster of alt-right subreddits. The alt-right subreddits that form this group are as follows:

- **/r/altright**: This is the first and the largest alt-right subreddit, created in 2010, and dedicated to the discussion of the alt-right political ideology. The gorup was banned in February 2017 for doxing.
- **/r/DebateAltRight**: This is another active alt-right subreddit. People from "the outside" are allowed to ask questions and to debate various aspects of the alt-right movement, but "anti-white" comments get downvoted pretty quickly. The sidebar of this subreddit contains the following quote, "Tenets of the Alt-Right: We believe in the Self-Advocacy, Self-Determination, and Self-Preservation of European peoples and nations."
- **/r/WhiteRights**: This is a self-proclaimed community for "discussing white interests, white identity and the culture of the West."

By measuring the daily likelihood that tweets in the datasets might be found in this collection of alt-right subreddits, we can provide a useful robustness check on our dictionary-based analysis. Consistent with our dictionary-based analysis, we do not observe an increase in white nationalist rhetoric in any of the datasets over the course of the campaign. Trump's election also has no effect on the use of alt-right rhetoric in the Trump, Clinton or random sample datasets, though there are some spikes in the Clinton and random sample datasets that may correspond to the small increases we observed using our dictionary-based methods. These results are provided in Tables A28-A30.

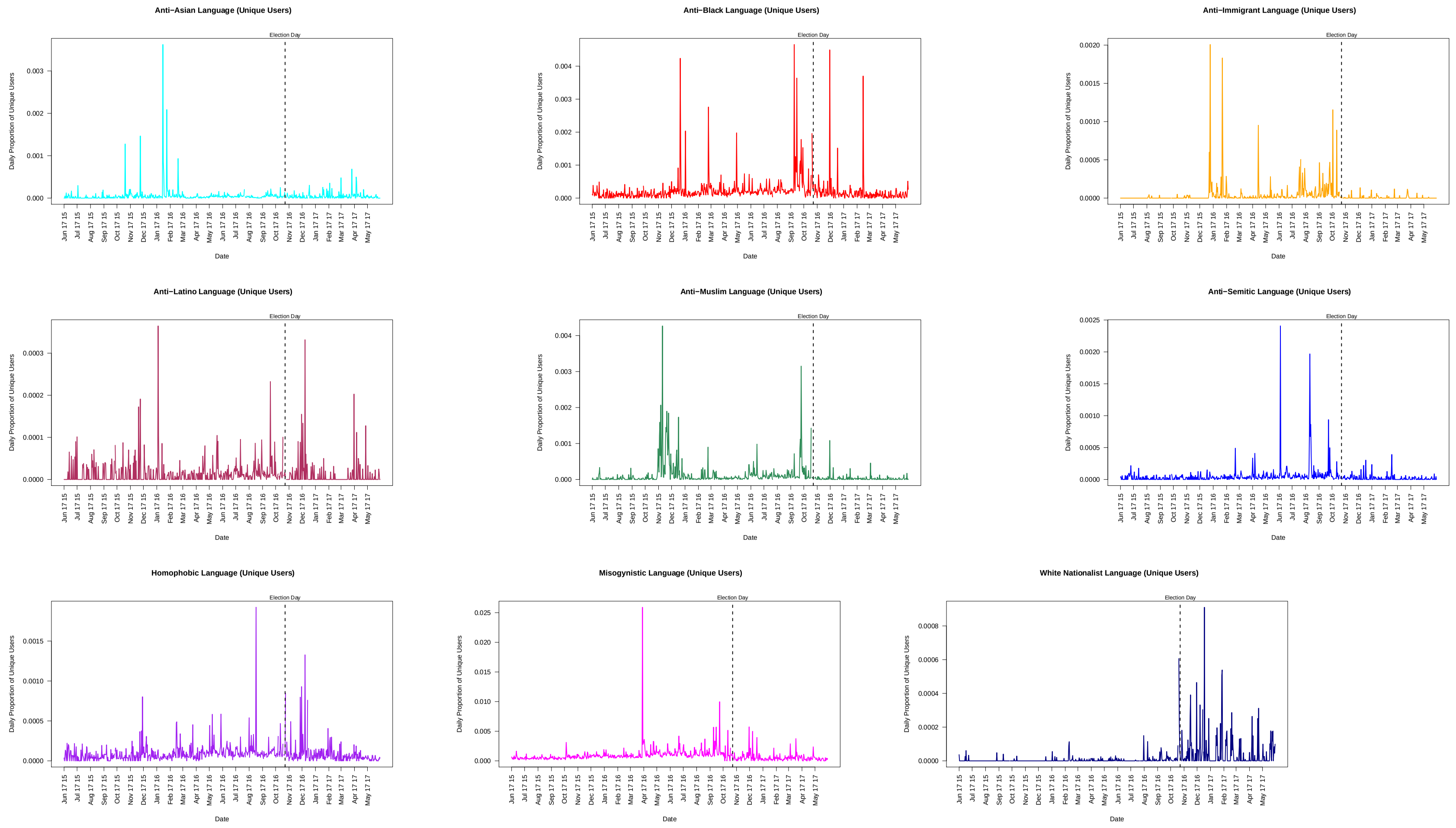
A.3 Figures and Tables

Fig. A1: Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Clinton Dataset)



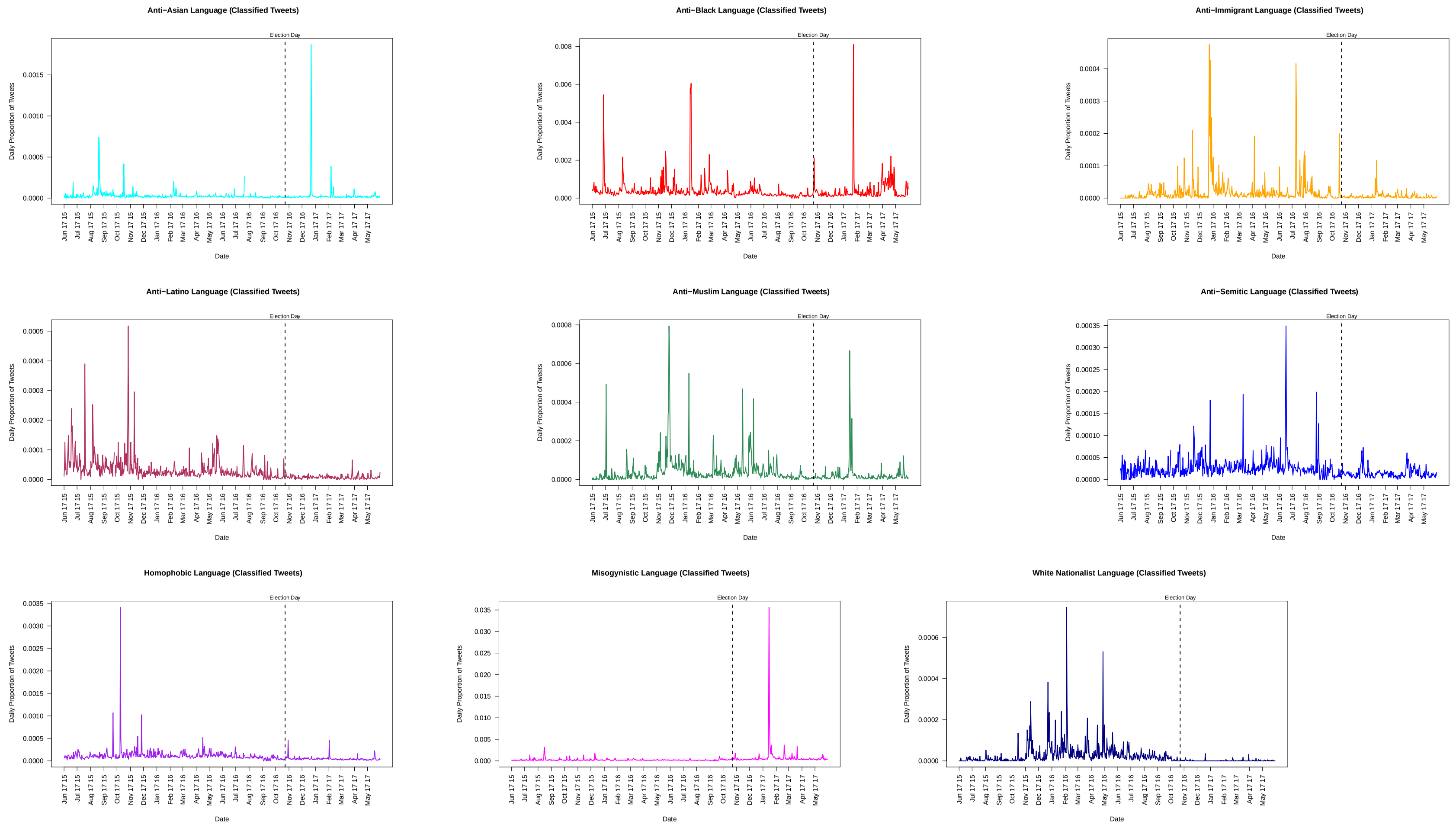
This plot shows the daily proportion of tweets containing hate speech and white nationalist rhetoric in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A2: Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Clinton Dataset)



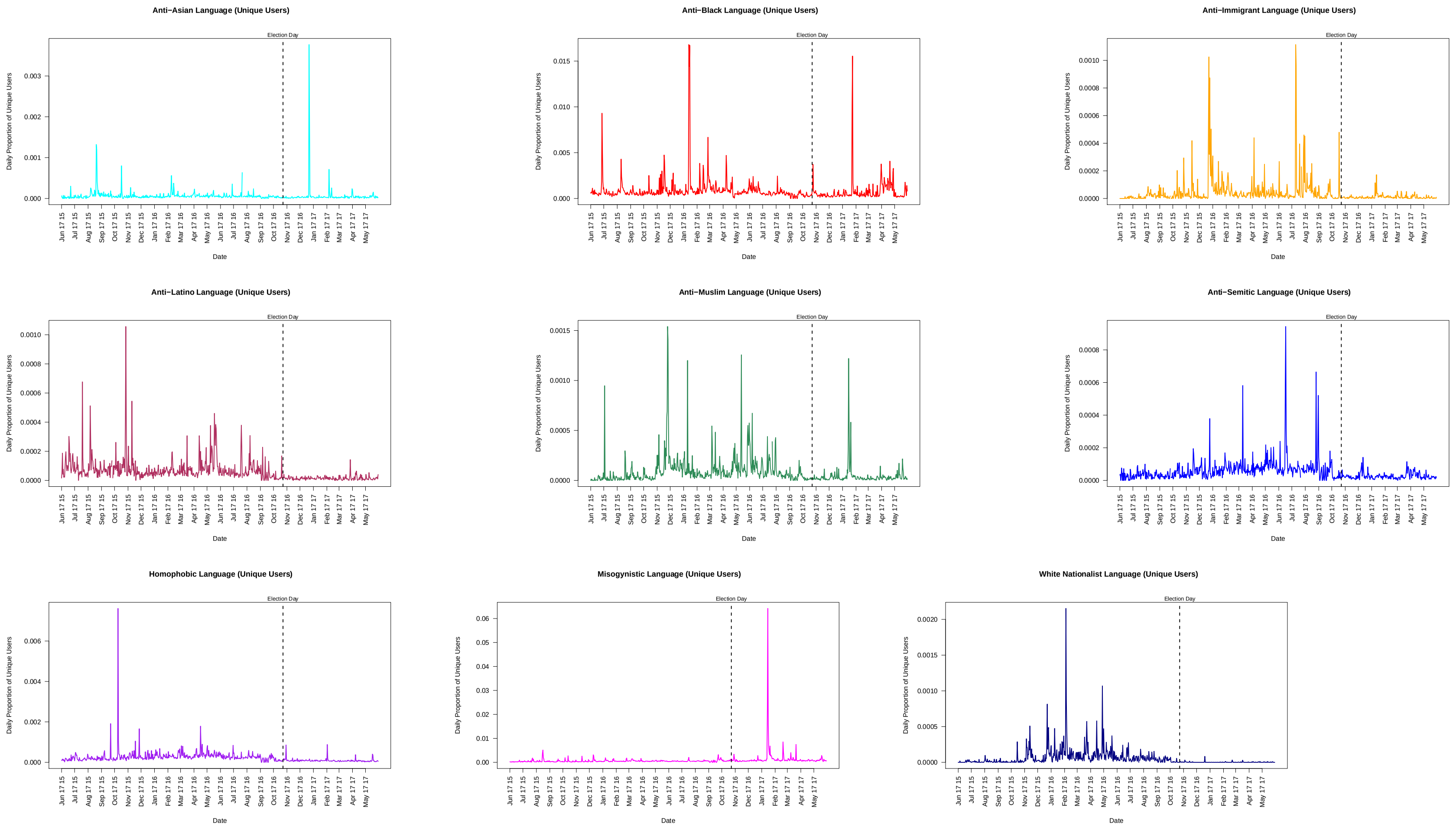
This plot shows the daily proportion of unique users tweeting hate speech and white nationalist rhetoric in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's Streaming API between June 15, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A3: Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Trump Dataset)



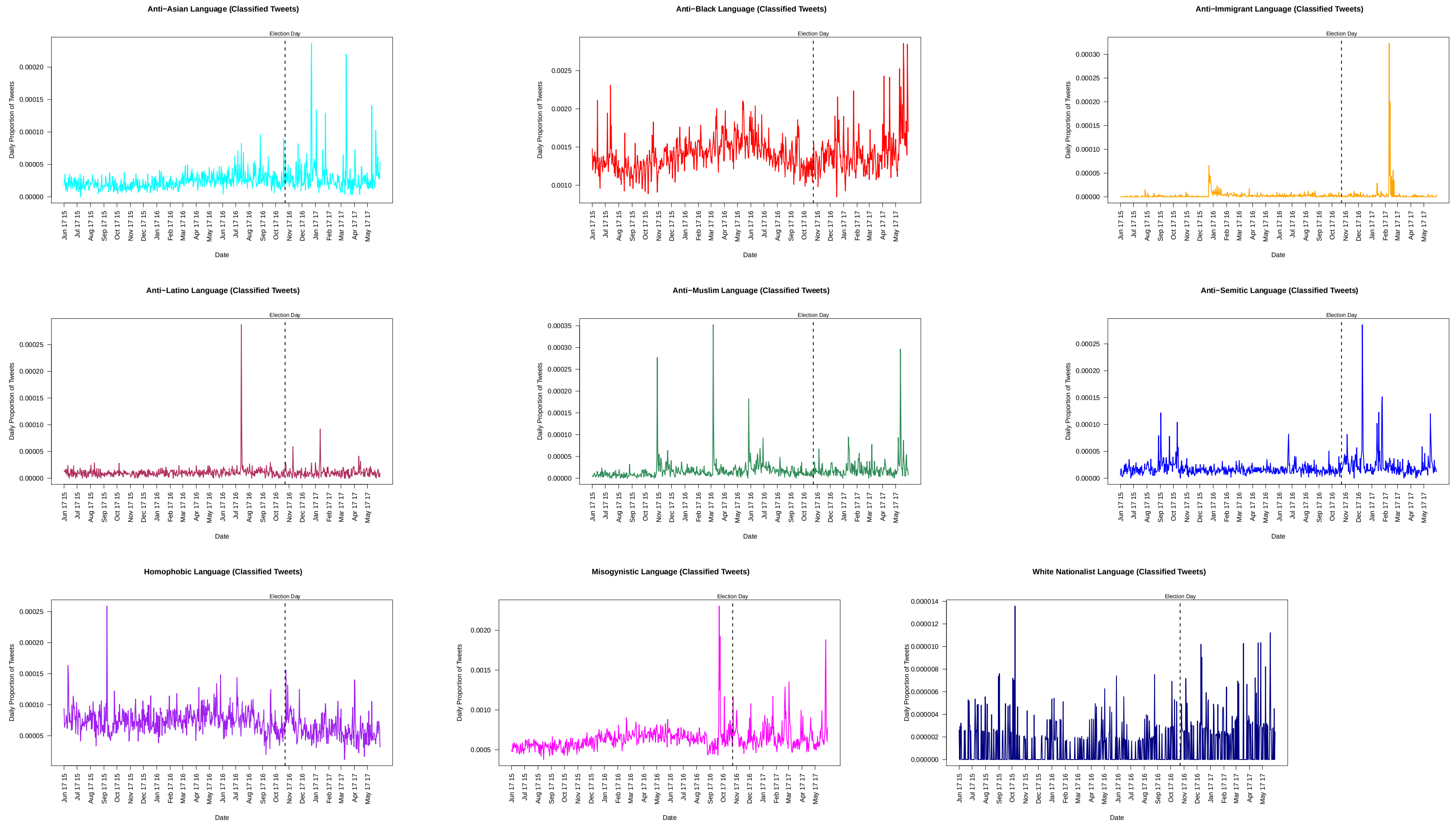
This plot shows the daily proportion of tweets containing hate speech and white nationalist rhetoric in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A4: Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Trump Dataset)



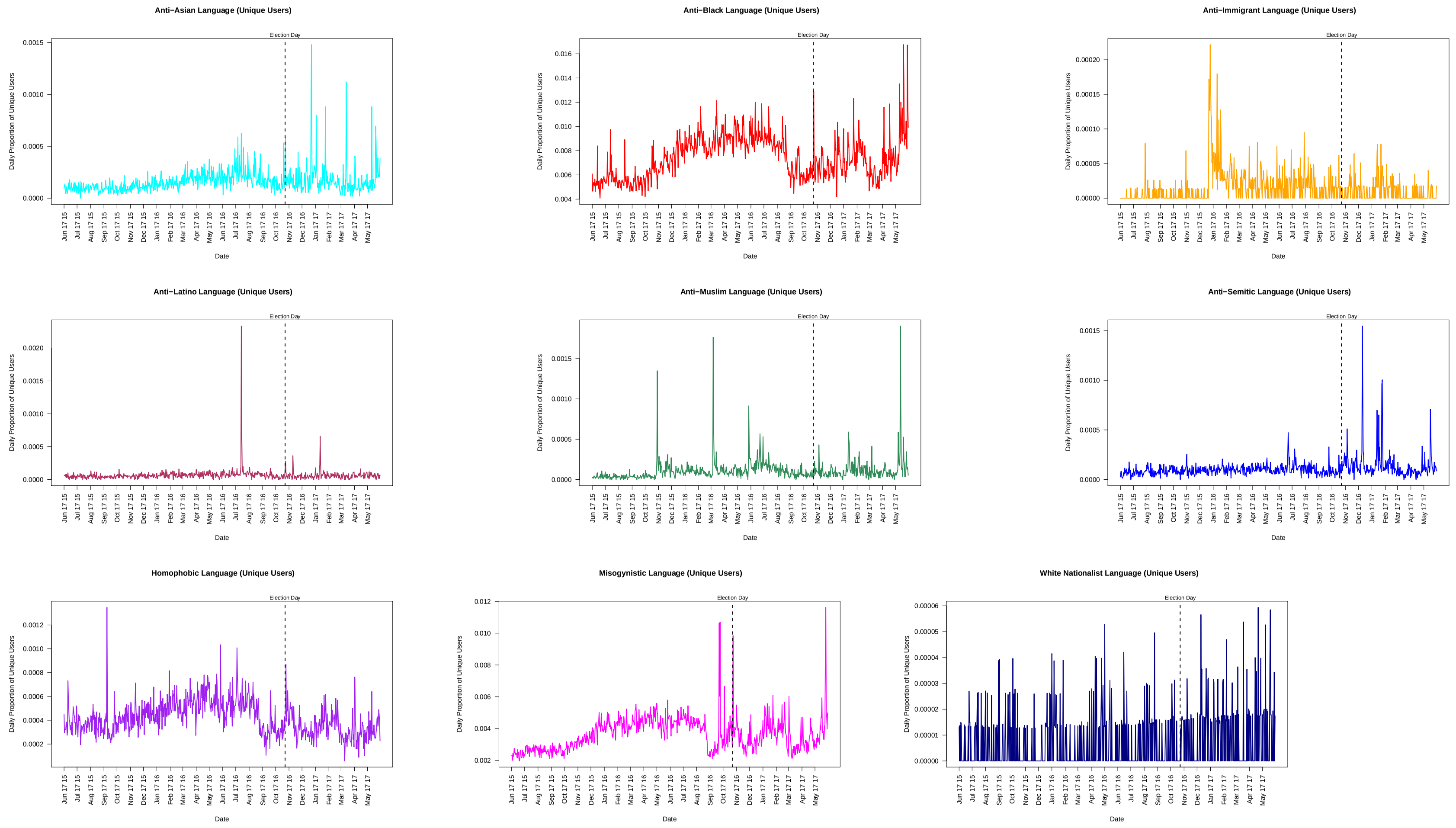
This plot shows the daily proportion unique users tweeting hate speech and white nationalist rhetoric in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A5: Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Random Sample Dataset)



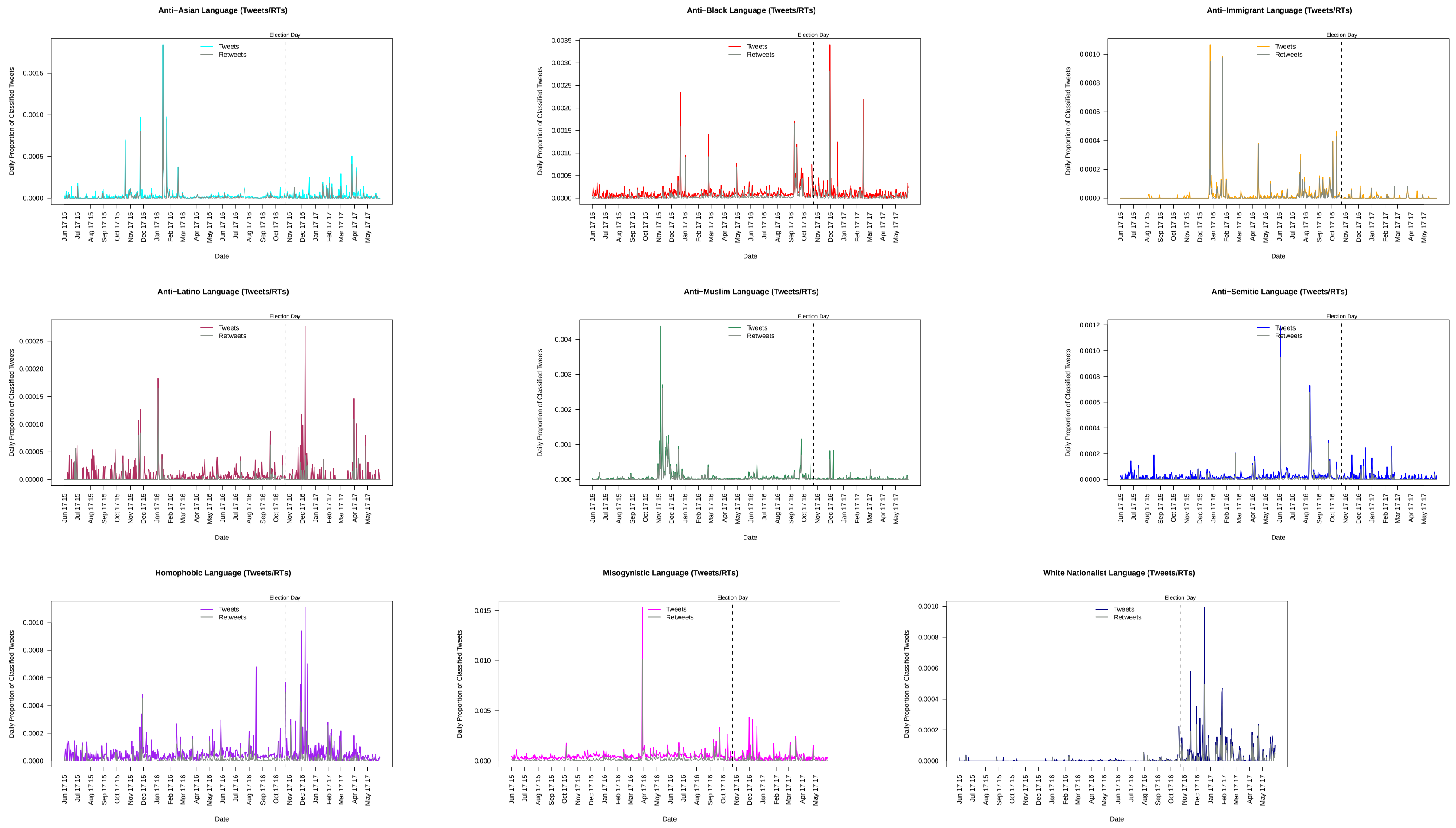
This plot shows the daily proportion of tweets containing hate speech and white nationalist rhetoric in a dataset of almost 400 million tweets sent by a random sample of 500,000 American Twitter users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A6: Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Random Sample Dataset)



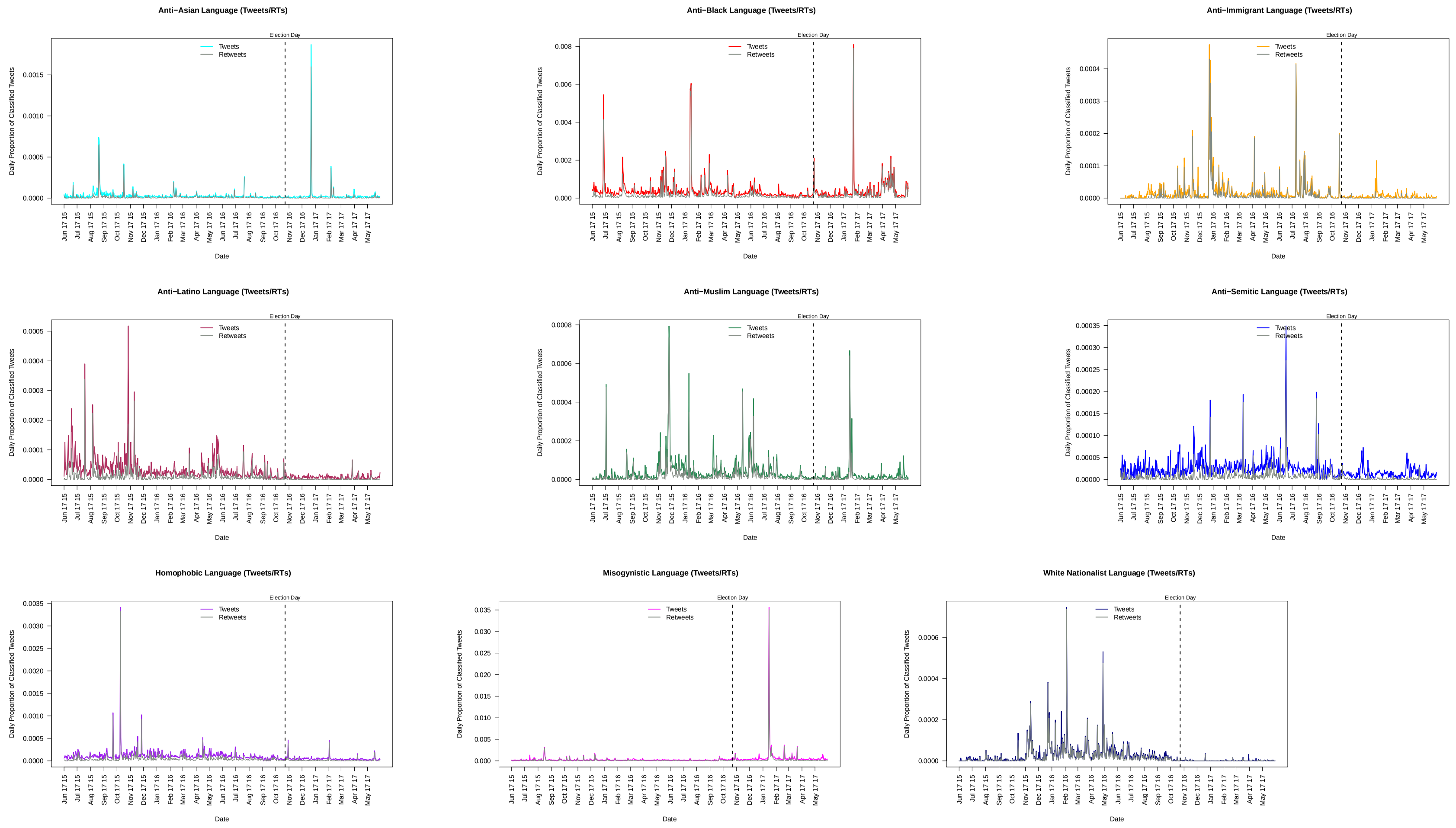
This plot shows the daily proportion of unique users tweeting hate speech and white nationalist rhetoric in a dataset of almost 400 million tweets sent by a random sample of 500,000 American Twitter users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A7: Daily Proportion of Tweets vs. Retweets Containing Hate Speech or White Nationalist Language (Clinton Dataset)



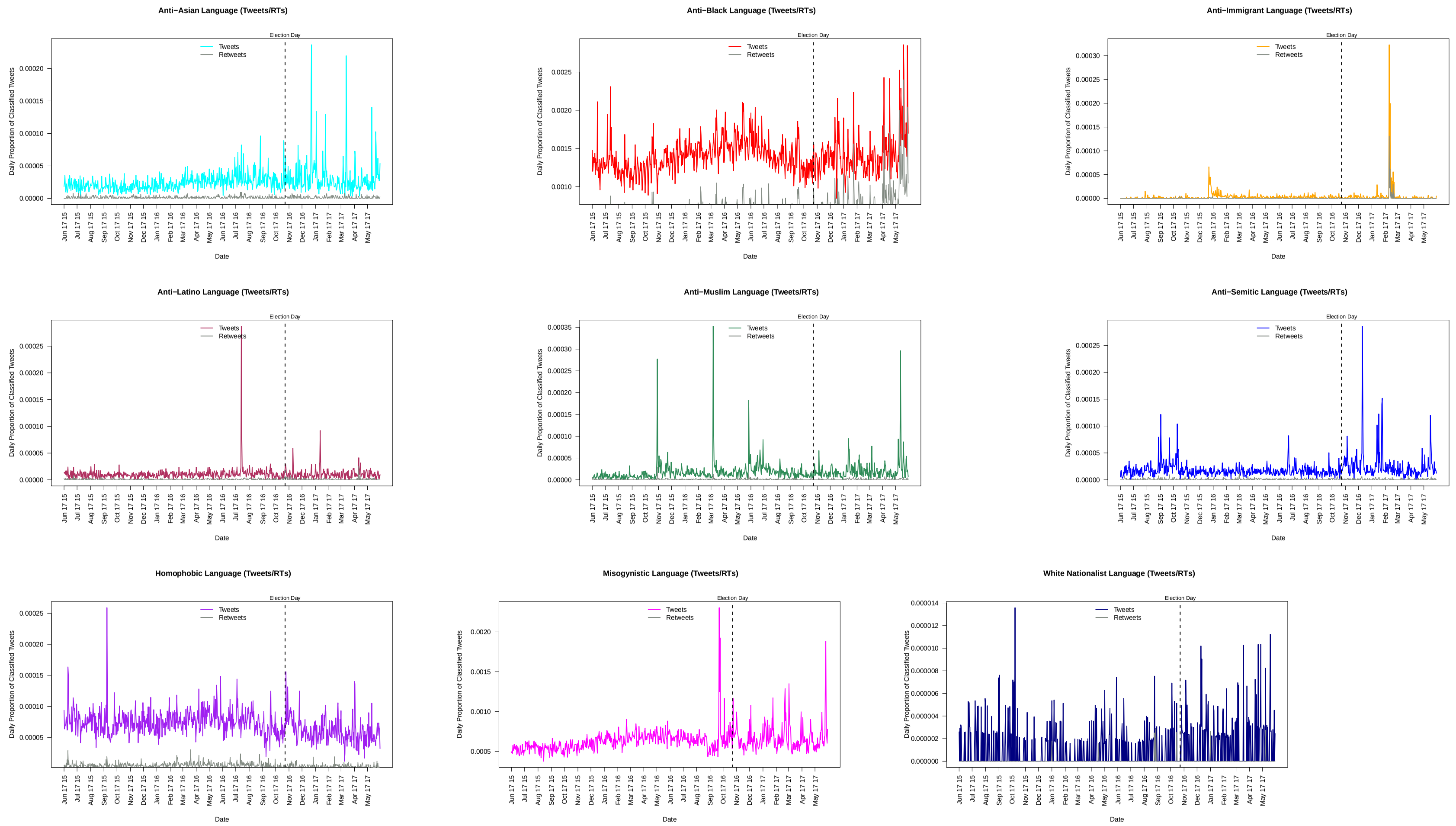
This plot shows the daily proportion of tweets vs. retweets containing hate speech and white nationalist rhetoric in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A8: Daily Proportion of Tweets vs. Retweets Containing Hate Speech or White Nationalist Language (Trump Dataset)



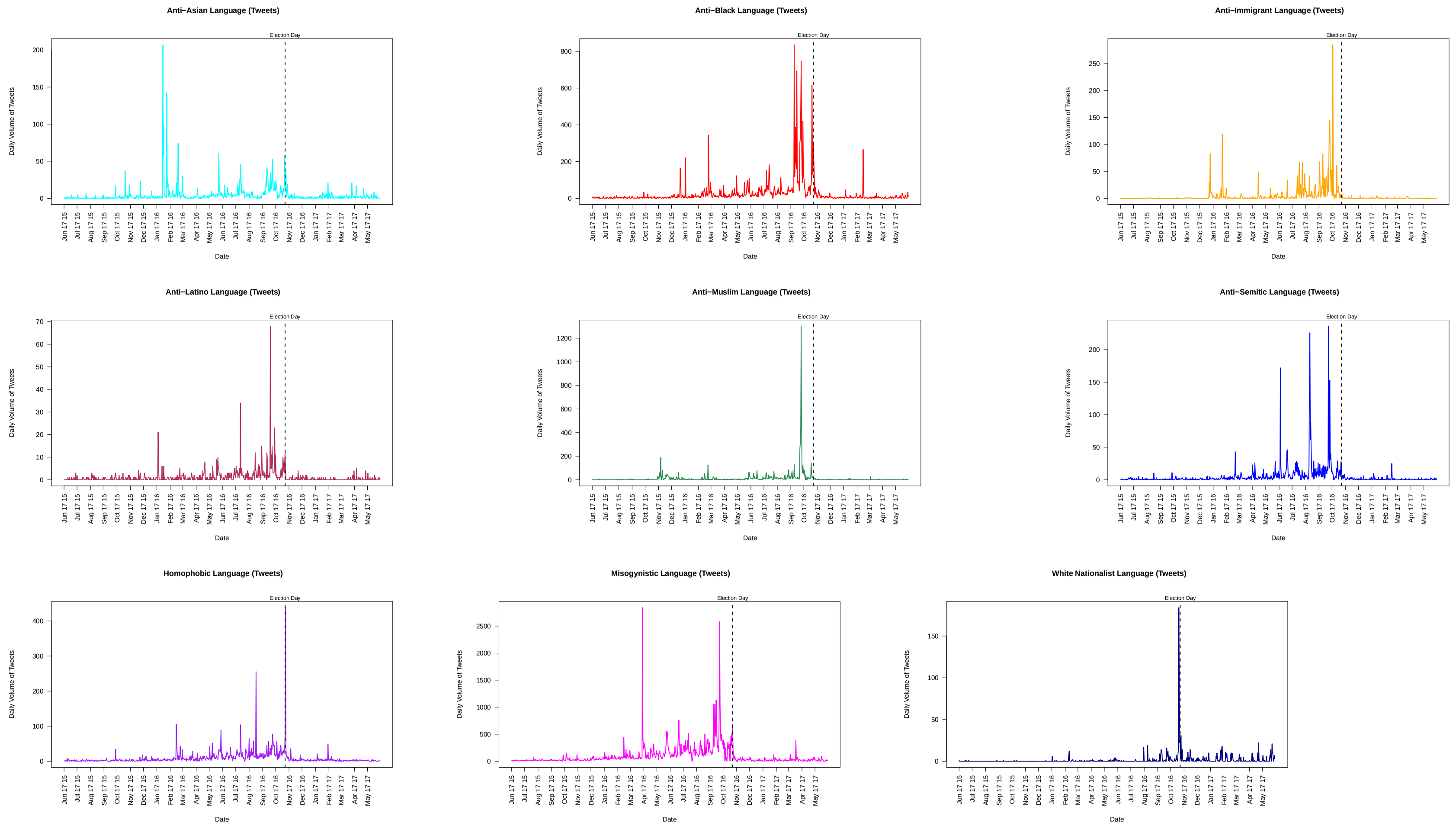
This plot shows the daily proportion of tweets vs. retweets containing hate speech and white nationalist rhetoric in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A9: Daily Proportion of Tweets vs. Retweets Containing Hate Speech or White Nationalist Language (Random Sample Dataset)



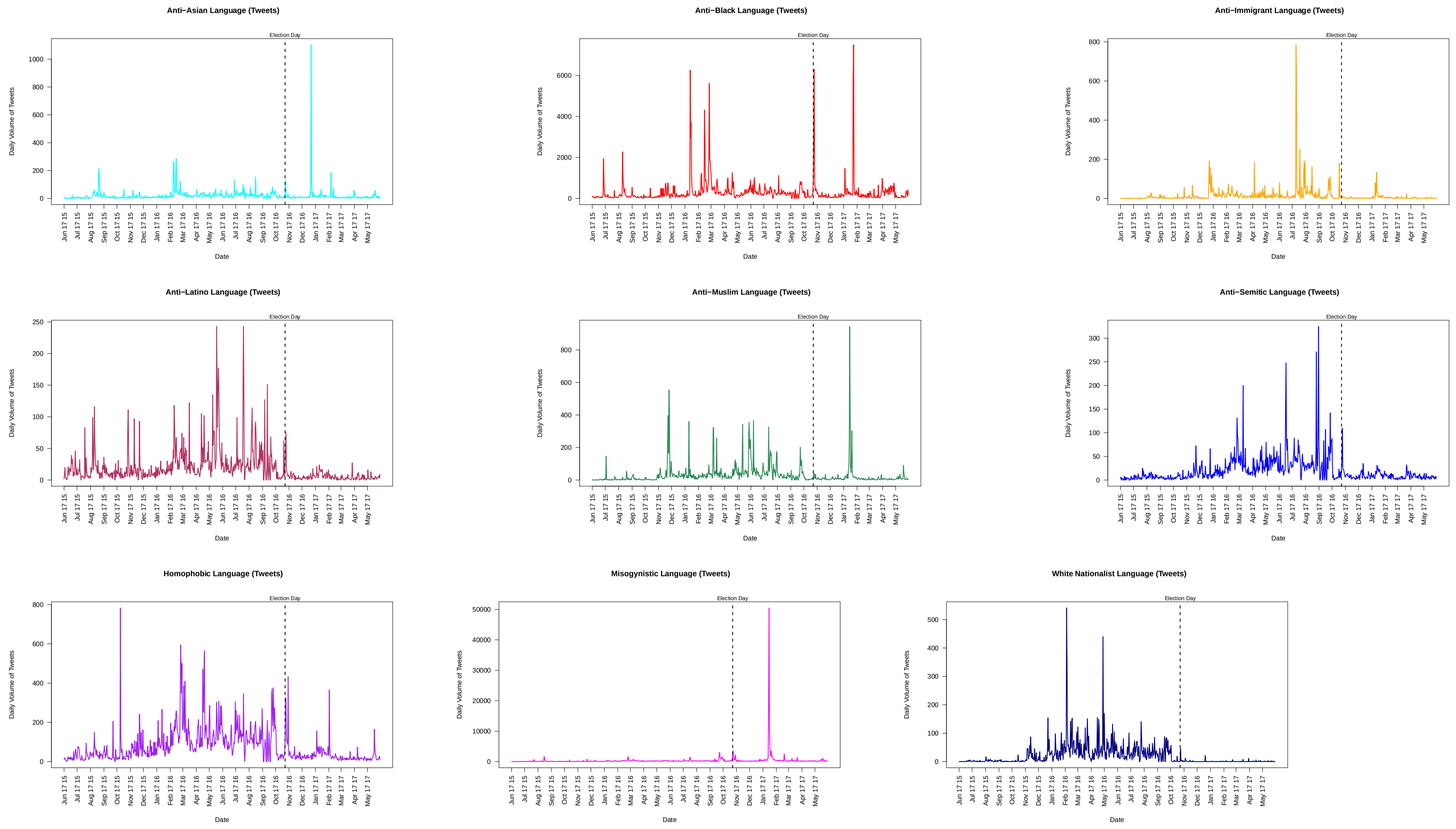
This plot shows the daily proportion of tweets vs. retweets containing hate speech and white nationalist rhetoric in a dataset of almost 400 million tweets sent by a random sample of 500,000 American Twitter users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A10: Daily Volume of Tweets Containing Hate Speech or White Nationalist Language (Clinton Dataset)



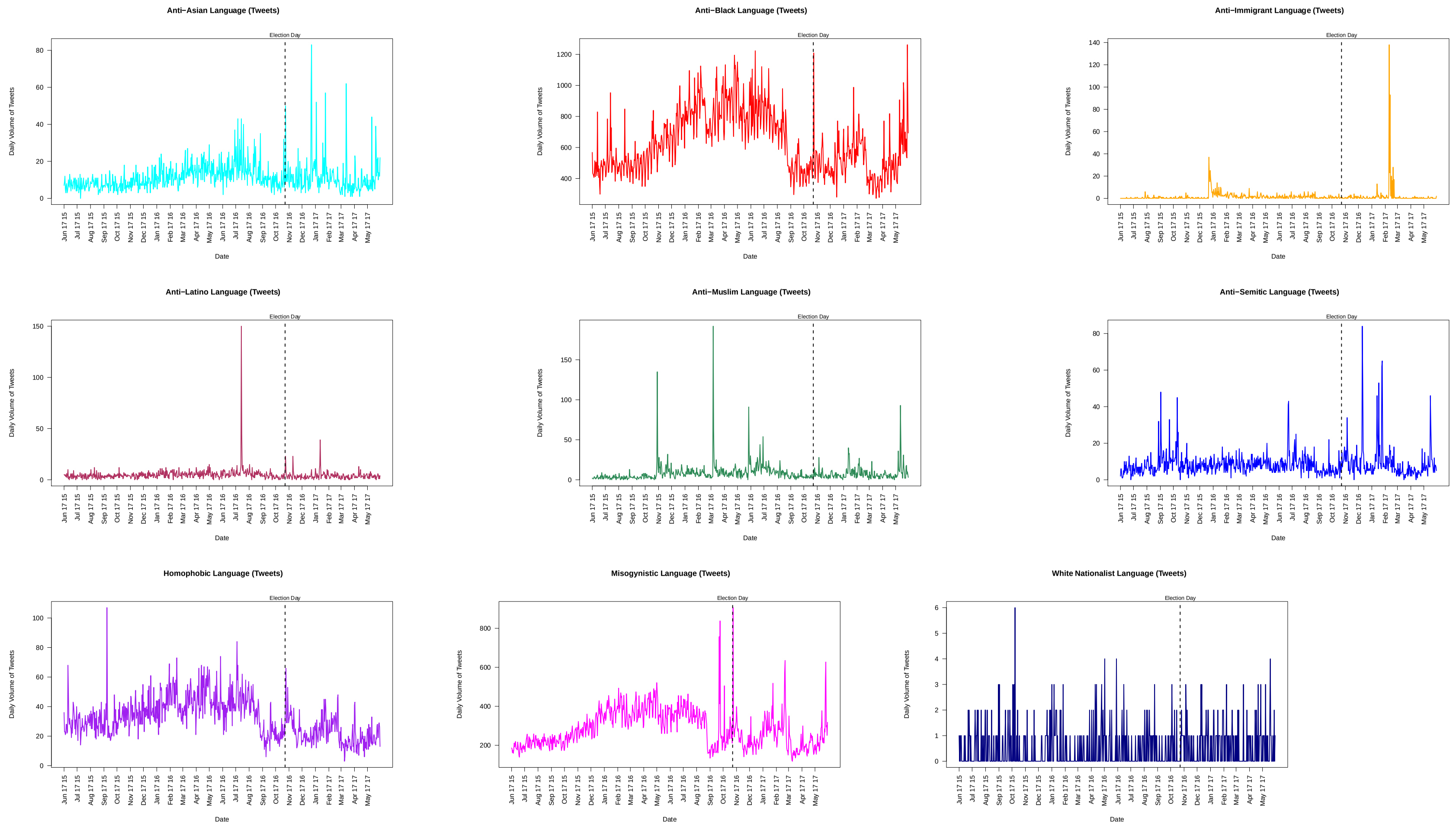
This figure shows the daily volume of classified hate-speech tweets in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A11: Daily Volume of Tweets Containing Hate Speech or White Nationalist Language (Trump Dataset)



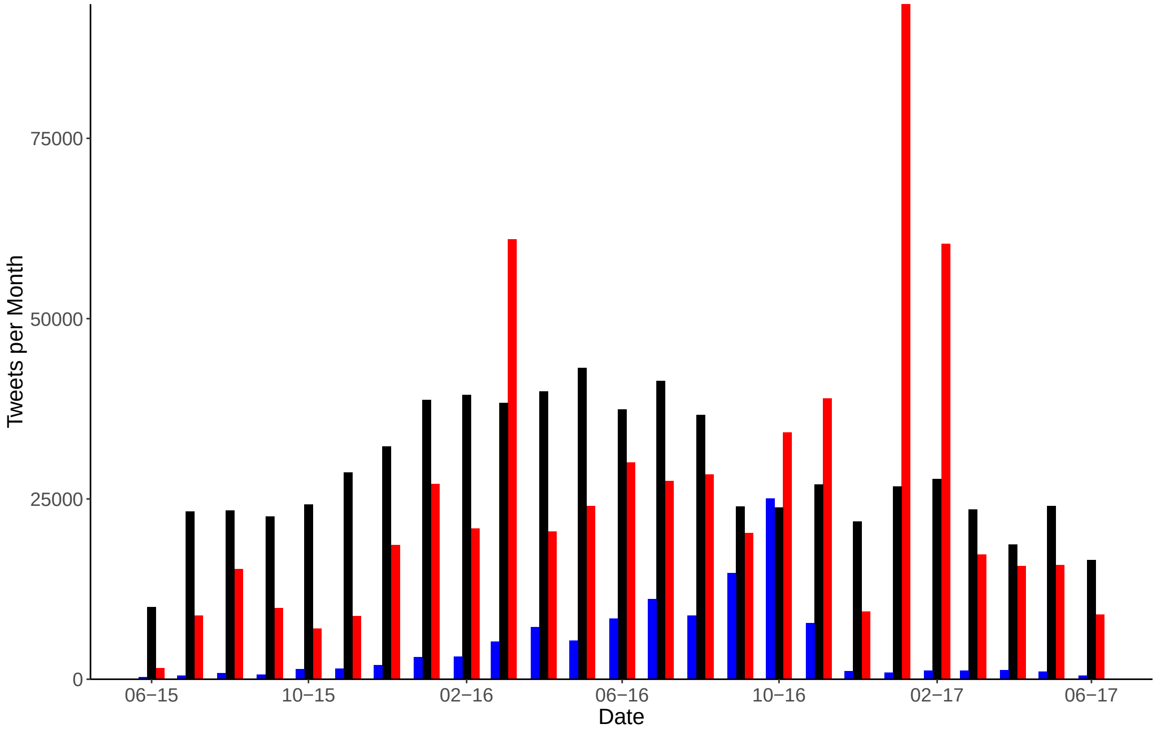
This figure shows the daily volume of classified hate-speech tweets in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A12: Daily Volume of Tweets Containing Hate Speech or White Nationalist Language (Random Sample Dataset)



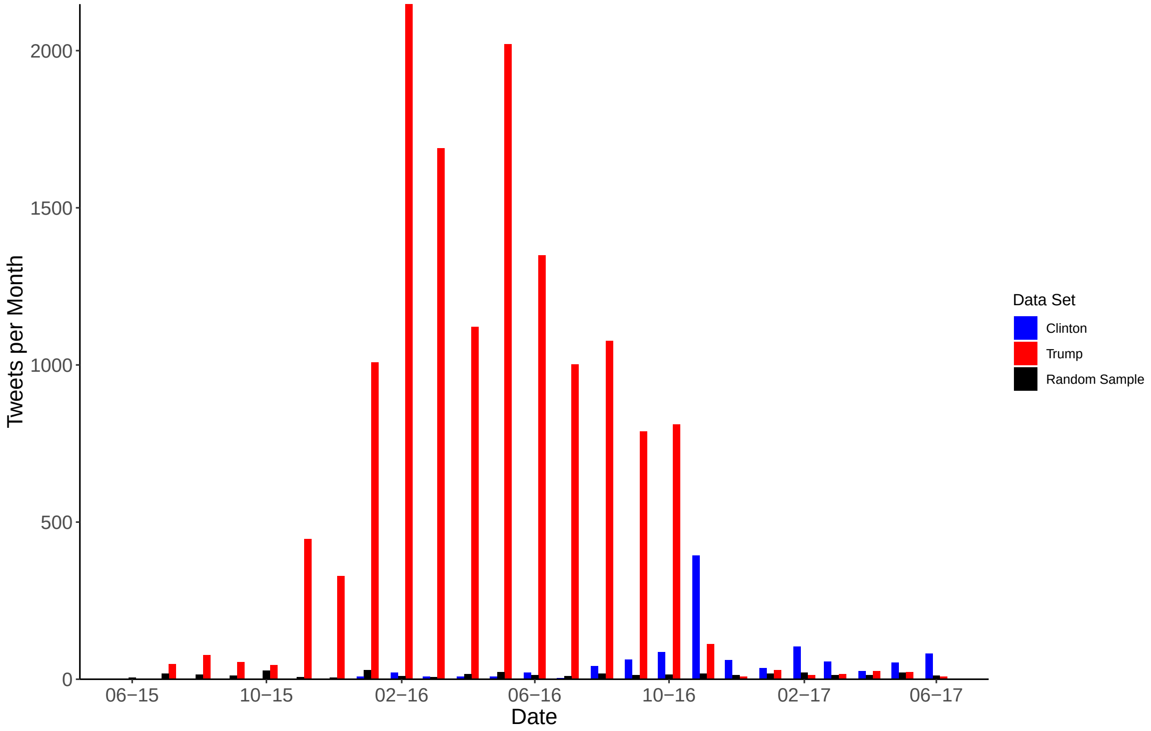
This figure shows the daily volume of classified hate-speech tweets in a dataset of almost 400 million tweets sent by a random sample of 500,000 American Twitter users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A13: Monthly Volume of Hate Speech Tweets
(Clinton, Trump, and Random Sample Datasets)



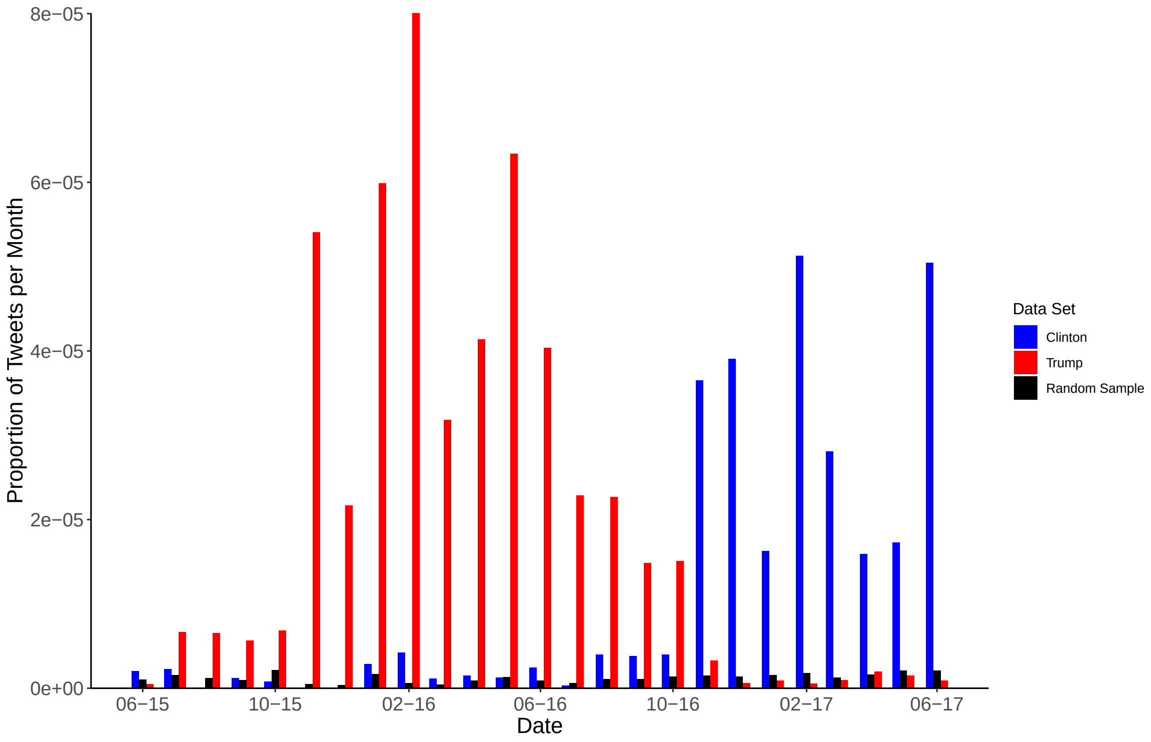
This figure shows the monthly volume of classified hate-speech tweets in the Clinton, Trump, and random sample datasets.

Fig. A14: Monthly Volume of White Nationalist Language Tweets
(Clinton, Trump, and Random Sample Datasets)



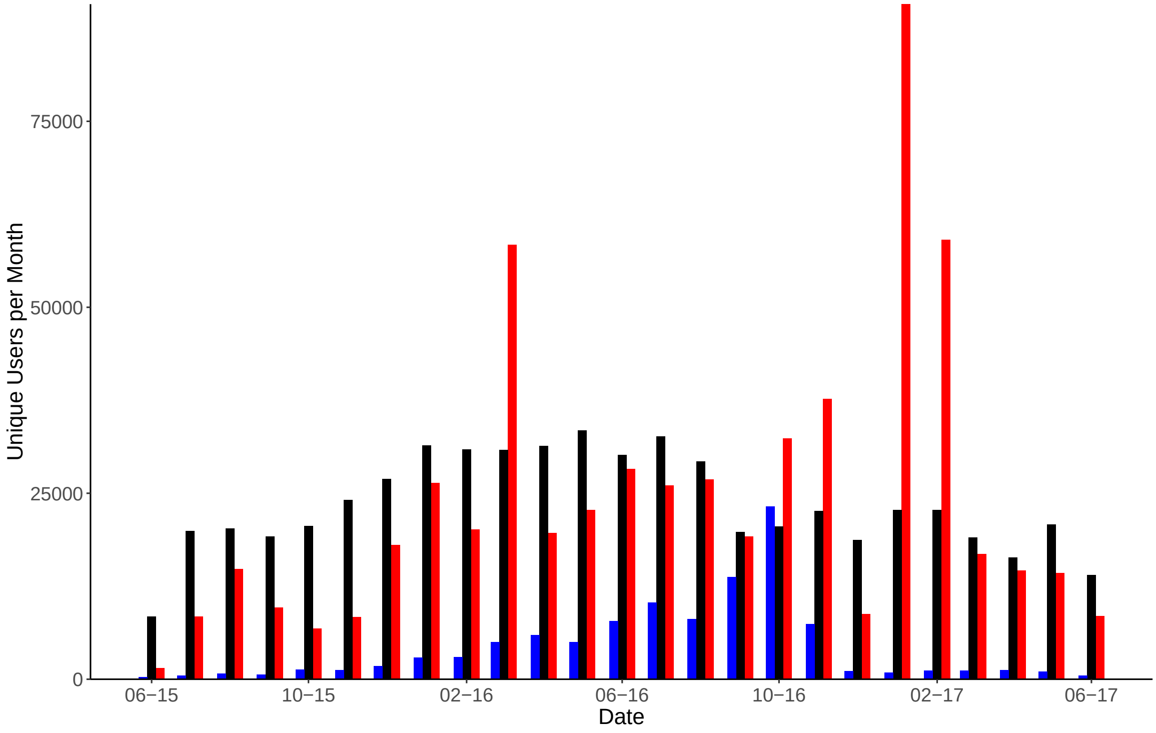
This figure shows the monthly volume of classified white nationalist tweets in the Clinton, Trump, and random sample datasets.

Fig. A15: Monthly Proportion of White Nationalist Language Tweets (Clinton, Trump, and Random Sample Datasets)



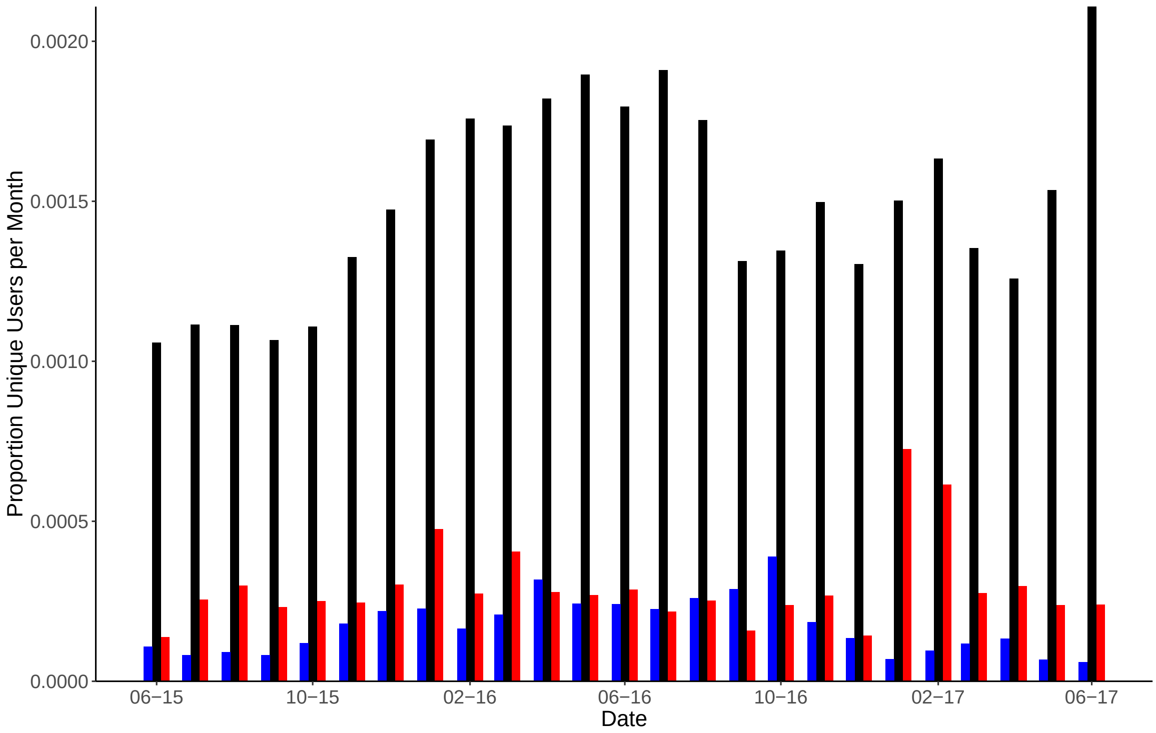
This figure shows the monthly proportion of classified white nationalist tweets in the Clinton, Trump, and random sample datasets.

Fig. A16: Monthly Unique Users Tweeting Hate Speech
(Clinton, Trump, and Random Sample Datasets)



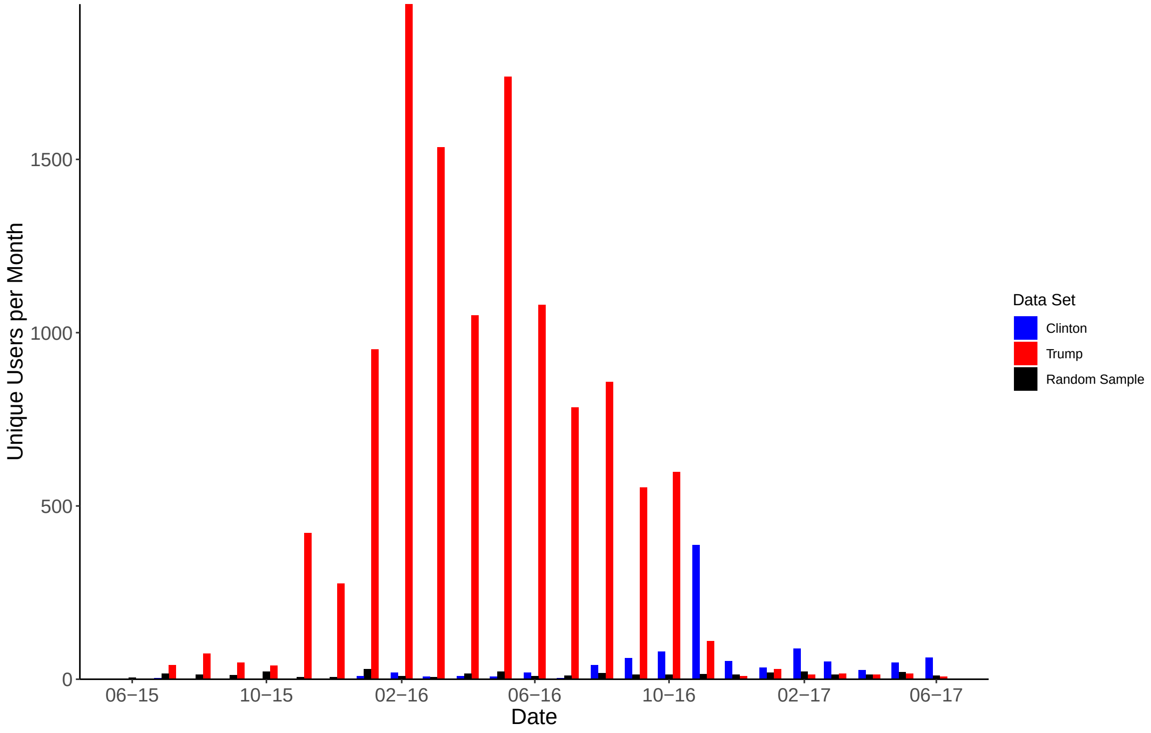
This figure shows the monthly number of unique users tweeting classified hate-speech tweets in the Clinton, Trump, and random sample datasets.

Fig. A17: Monthly Proportion of Unique Users Tweeting Hate Speech
(Clinton, Trump, and Random Sample Datasets)



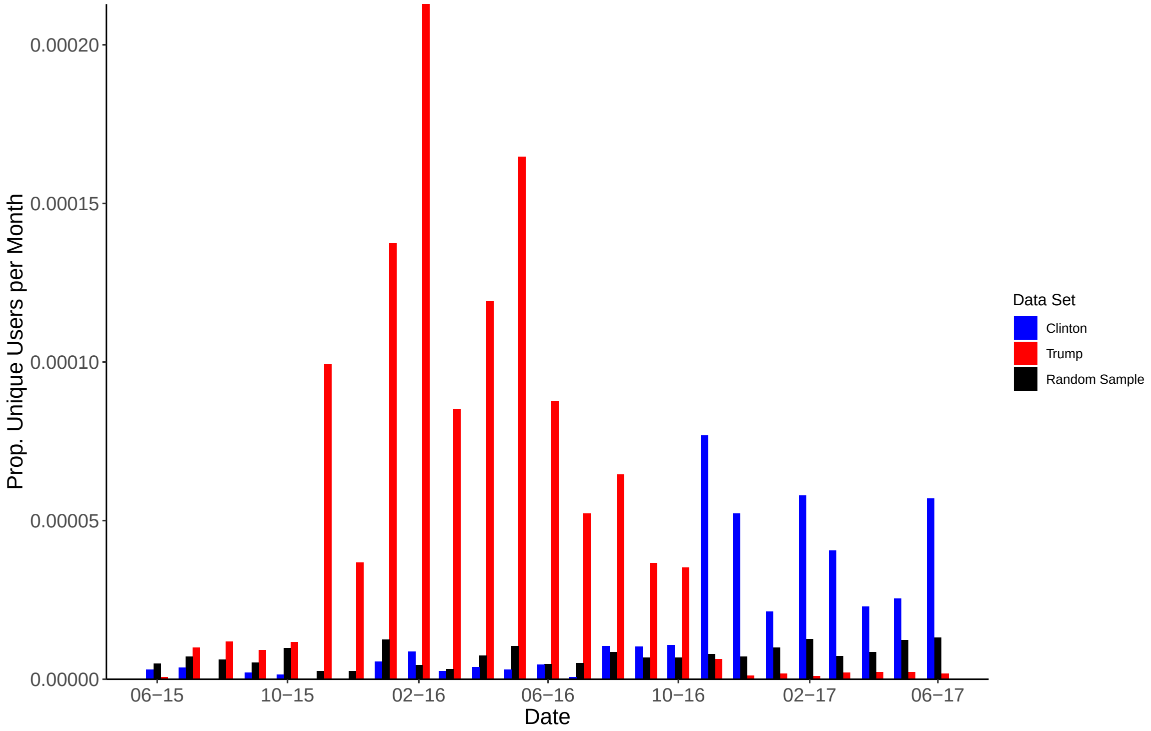
This figure shows the monthly proportion of unique users tweeting classified hate-speech tweets in the Clinton, Trump, and random sample datasets.

Fig. A18: Monthly Unique Users Tweeting White Nationalist Language (Clinton, Trump, and Random Sample Datasets)



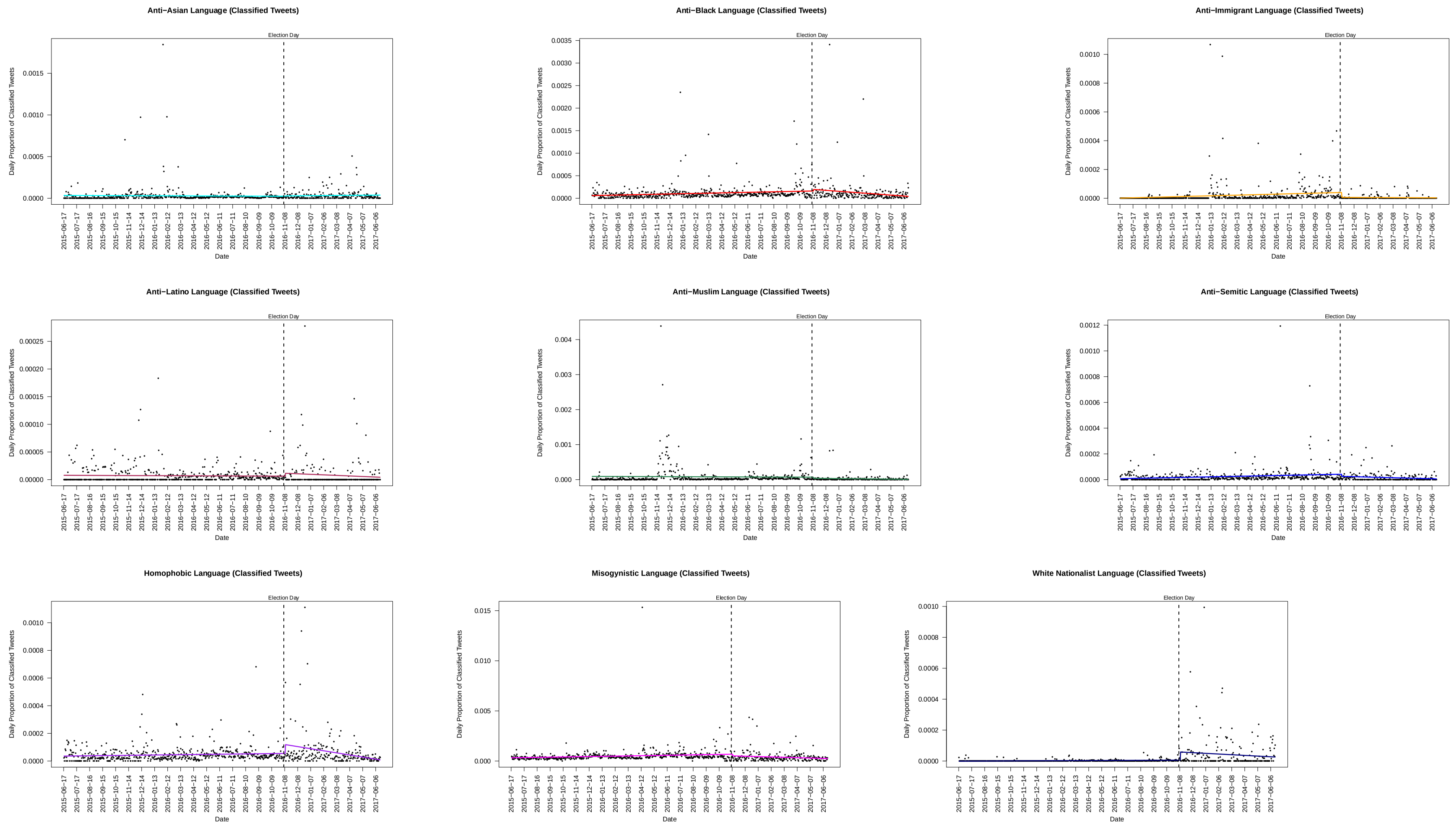
This figure shows the monthly number of unique users tweeting classified white nationalist tweets in the Clinton, Trump, and random sample datasets.

Fig. A19: Monthly Proportion of Unique Users Tweeting White Nationalist Language (Clinton, Trump, and Random Sample Datasets)



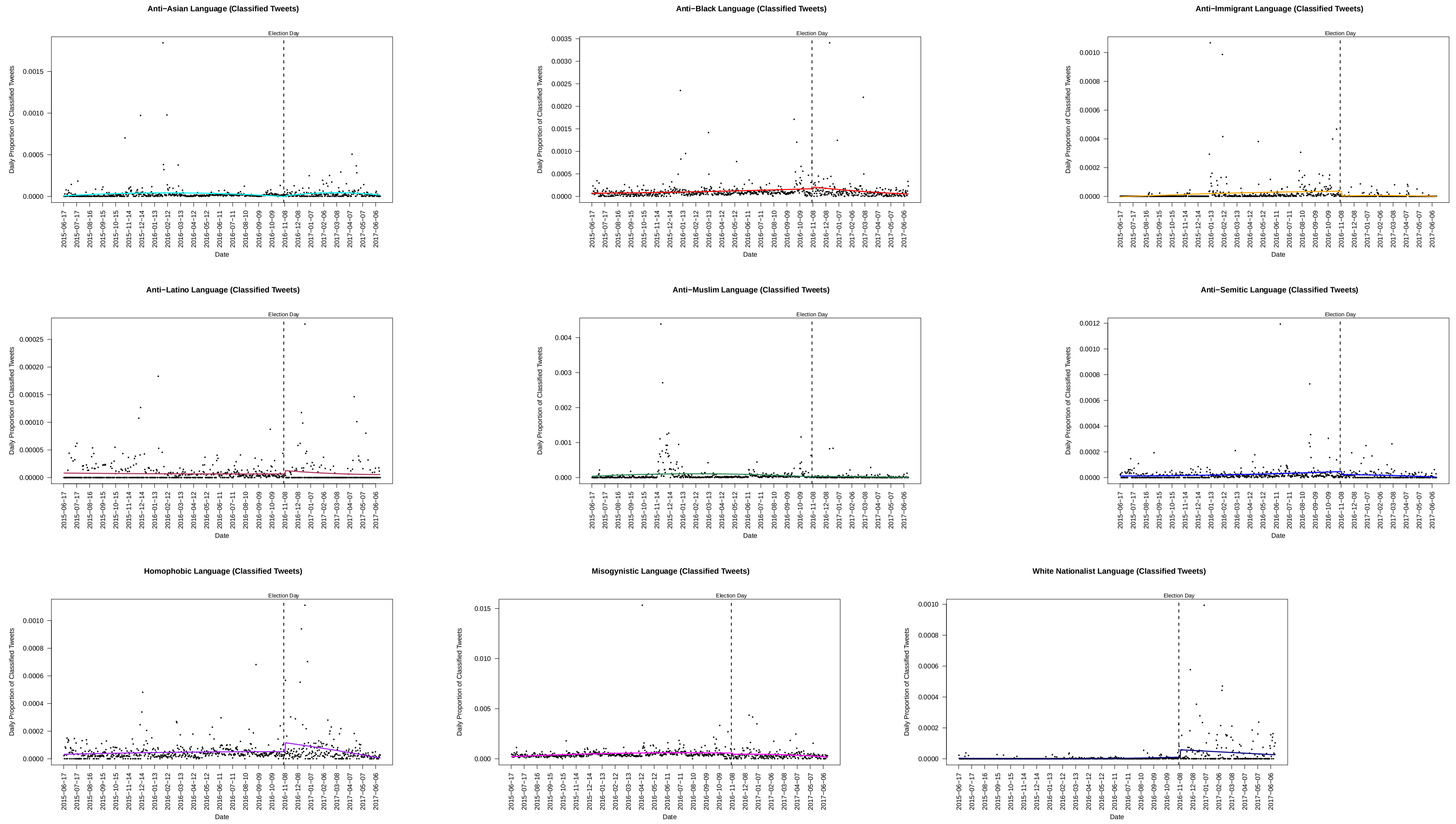
This figure shows the monthly proportion of unique users tweeting classified white nationalist tweets in the Clinton, Trump, and random sample datasets.

Fig. A20: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Clinton Dataset)



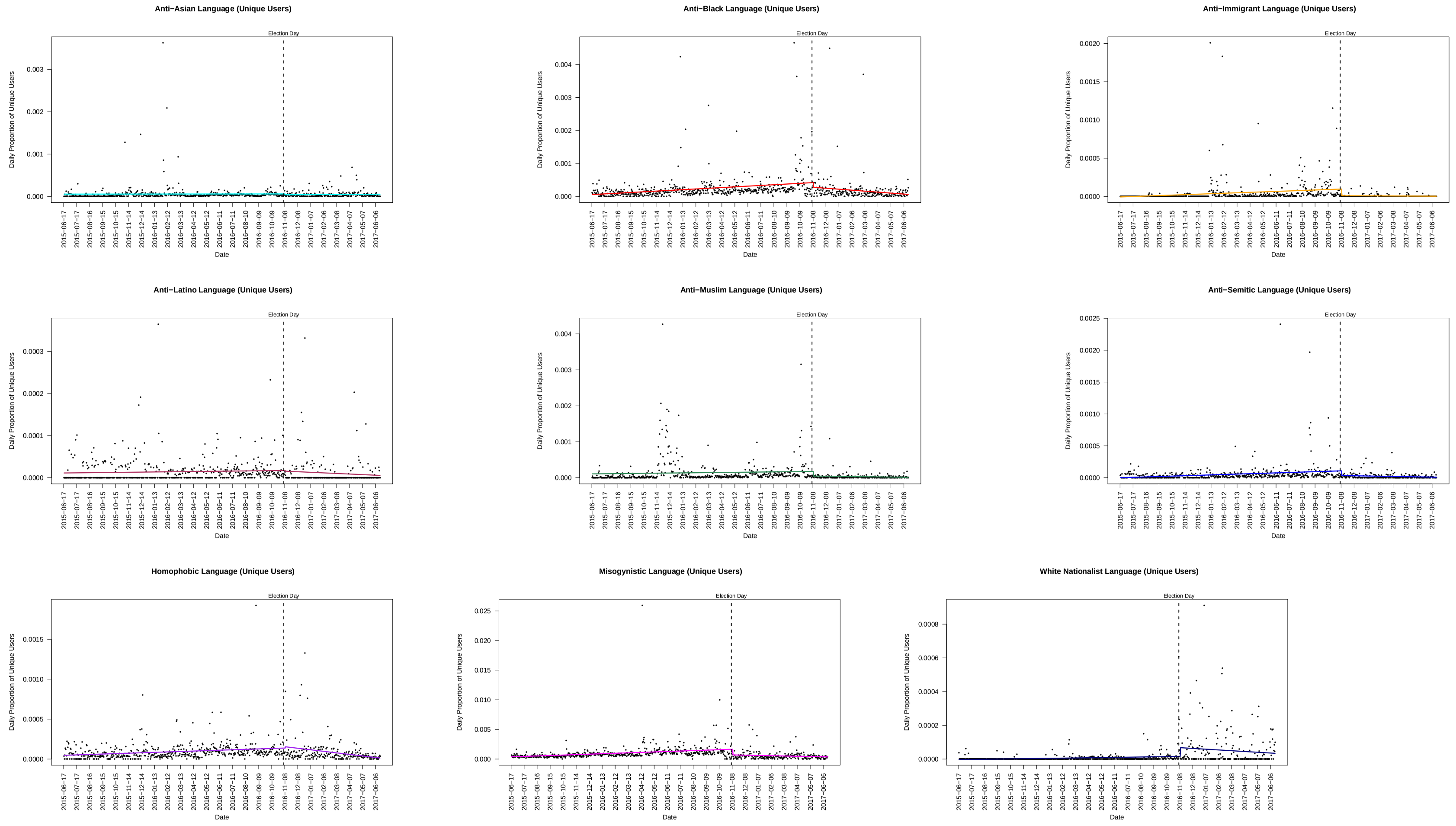
This plot shows the predicted trend lines based on our AR1 linear interrupted time series regression model plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in the dataset of over 150 million tweets referencing Hillary Clinton collected between June 17, 2015, and June 15, 2017.

Fig. A21: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language Including Quadratic Terms (Clinton Dataset)



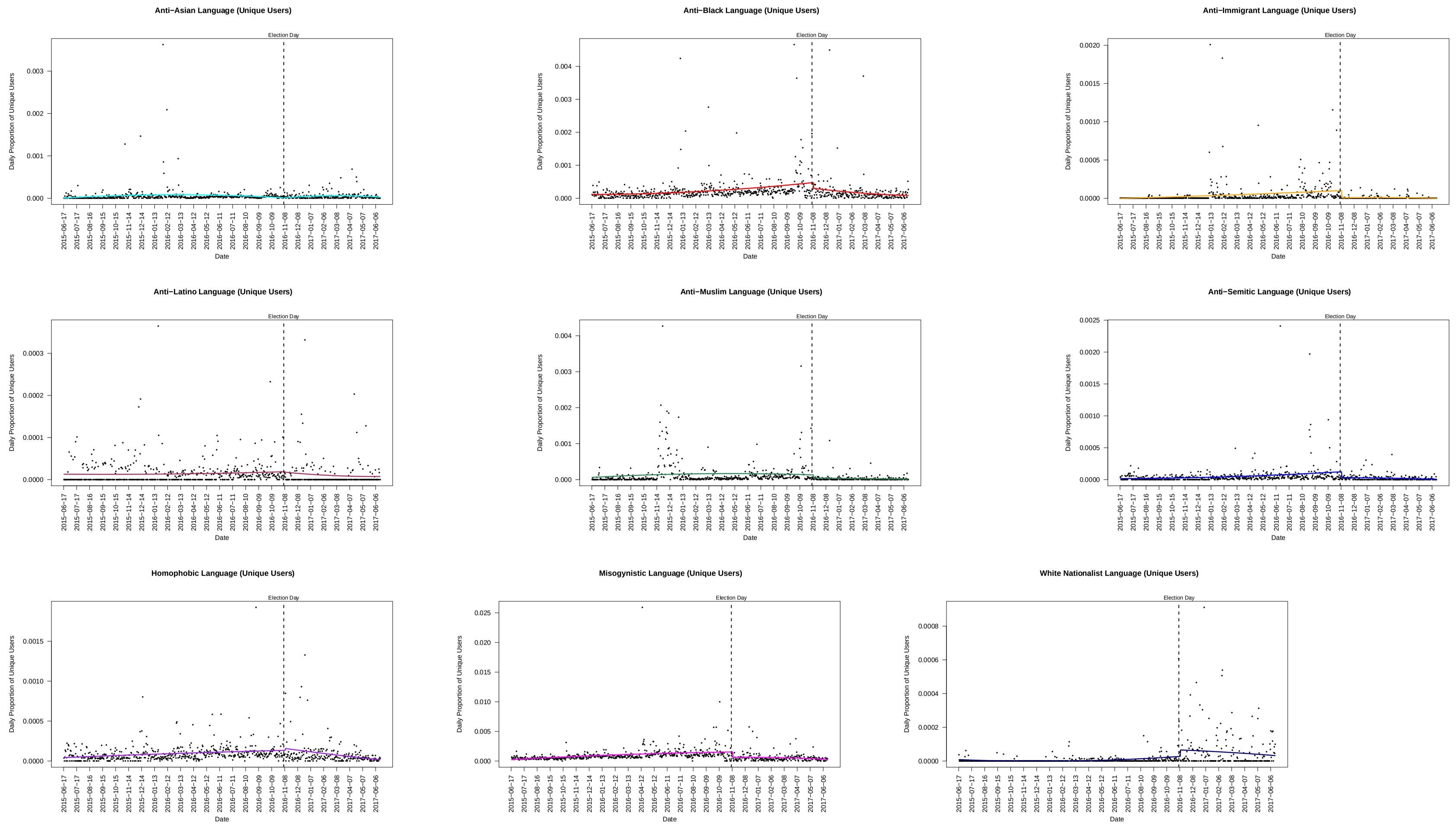
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model including quadratic terms, plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in the dataset of over 150 million tweets referencing Hillary Clinton collected between June 17, 2015, and June 15, 2017.

Fig. A22: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Clinton Dataset)



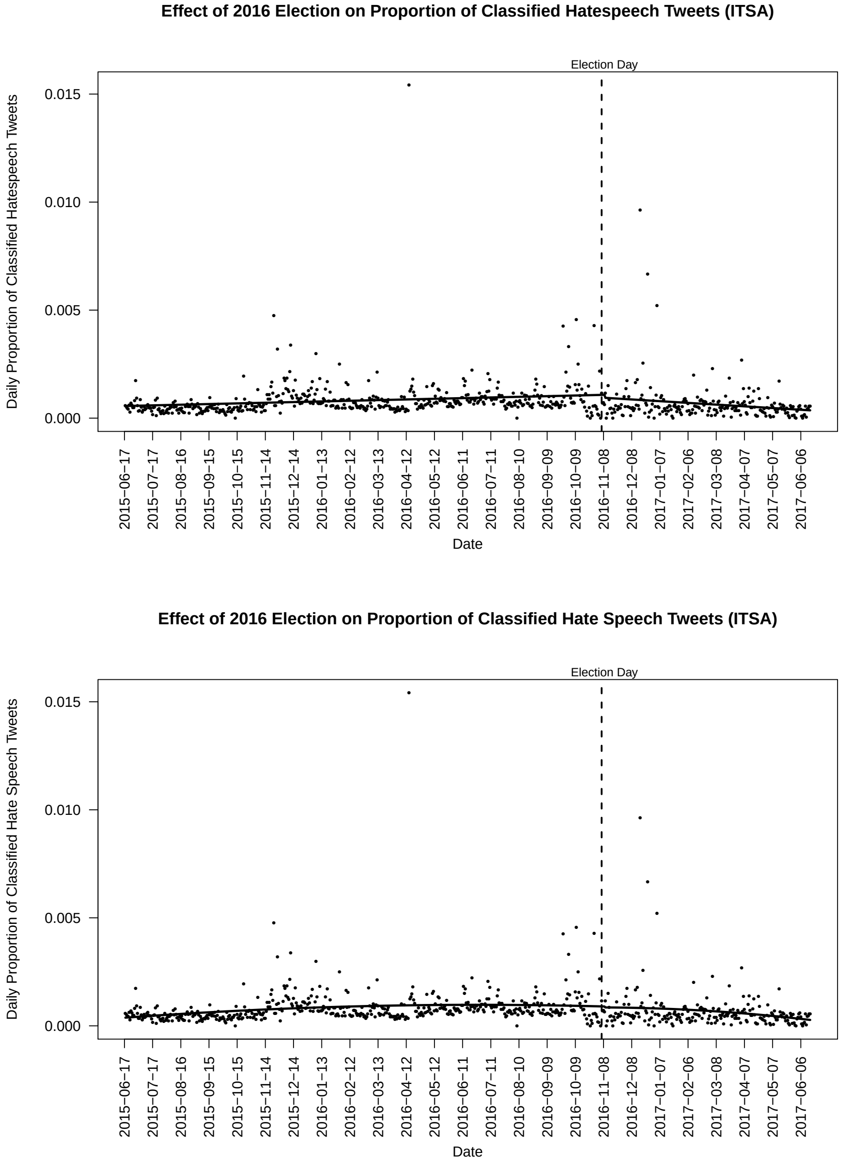
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model plotted against the observed daily proportion number of unique users producing hate speech and white nationalist rhetoric in a dataset of over 150 million tweets mentioning Hillary Clinton collected between June 17, 2015, and June 15, 2017.

Fig. A23: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language Including Quadratic Terms (Clinton Dataset)



This plot shows the predicted trend lines based on our AR1 interrupted time series regression model with quadratic terms plotted against the observed daily proportion number of unique users producing hate speech and white nationalist rhetoric in a dataset of over 150 million tweets mentioning Hillary Clinton collected between June 17, 2015, and June 15, 2017.

Fig. A24: Aggregate Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language
(Clinton Dataset)

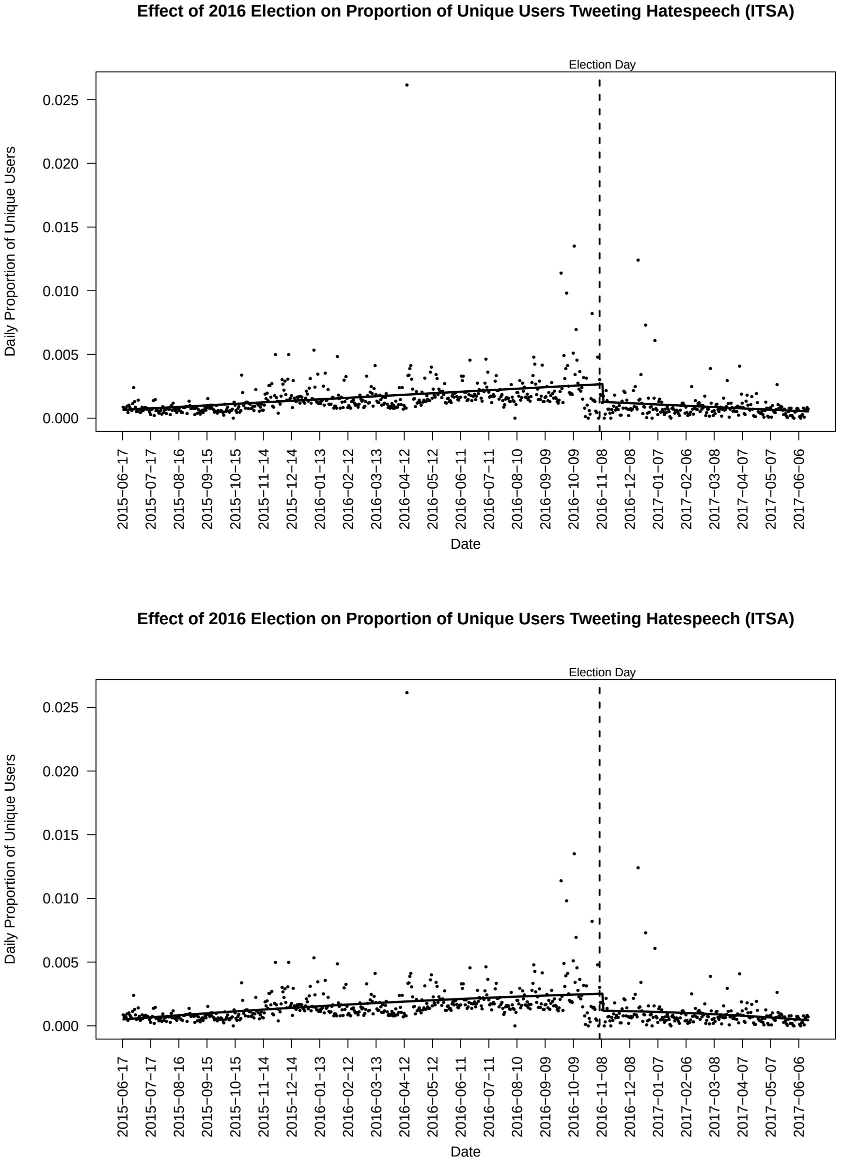


Effect of 2016 Election on Proportion of Classified Hatespeech Tweets (ITSA)

Effect of 2016 Election on Proportion of Classified Hate Speech Tweets (ITSA)

These plots show the predicted trend line of the proportion of hate speech tweets based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of hate speech tweets (black dots) in the dataset of over 125 million tweets referencing Hillary Clinton collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A25: Aggregate Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech
(Clinton Dataset)

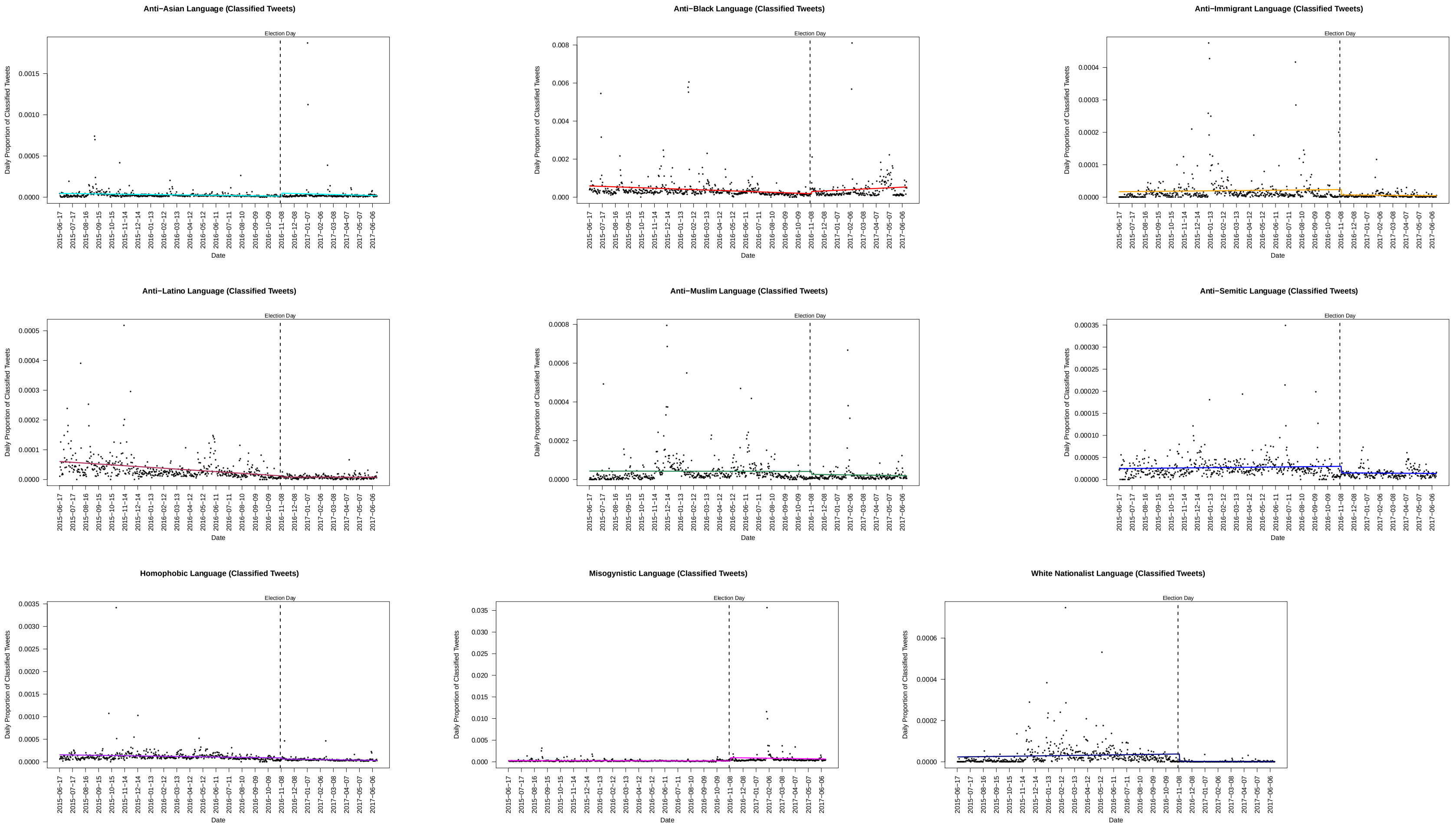


Effect of 2016 Election on Proportion of Unique Users Tweeting Hatespeech (ITSA)

Effect of 2016 Election on Proportion of Unique Users Tweeting Hatespeech (ITSA)

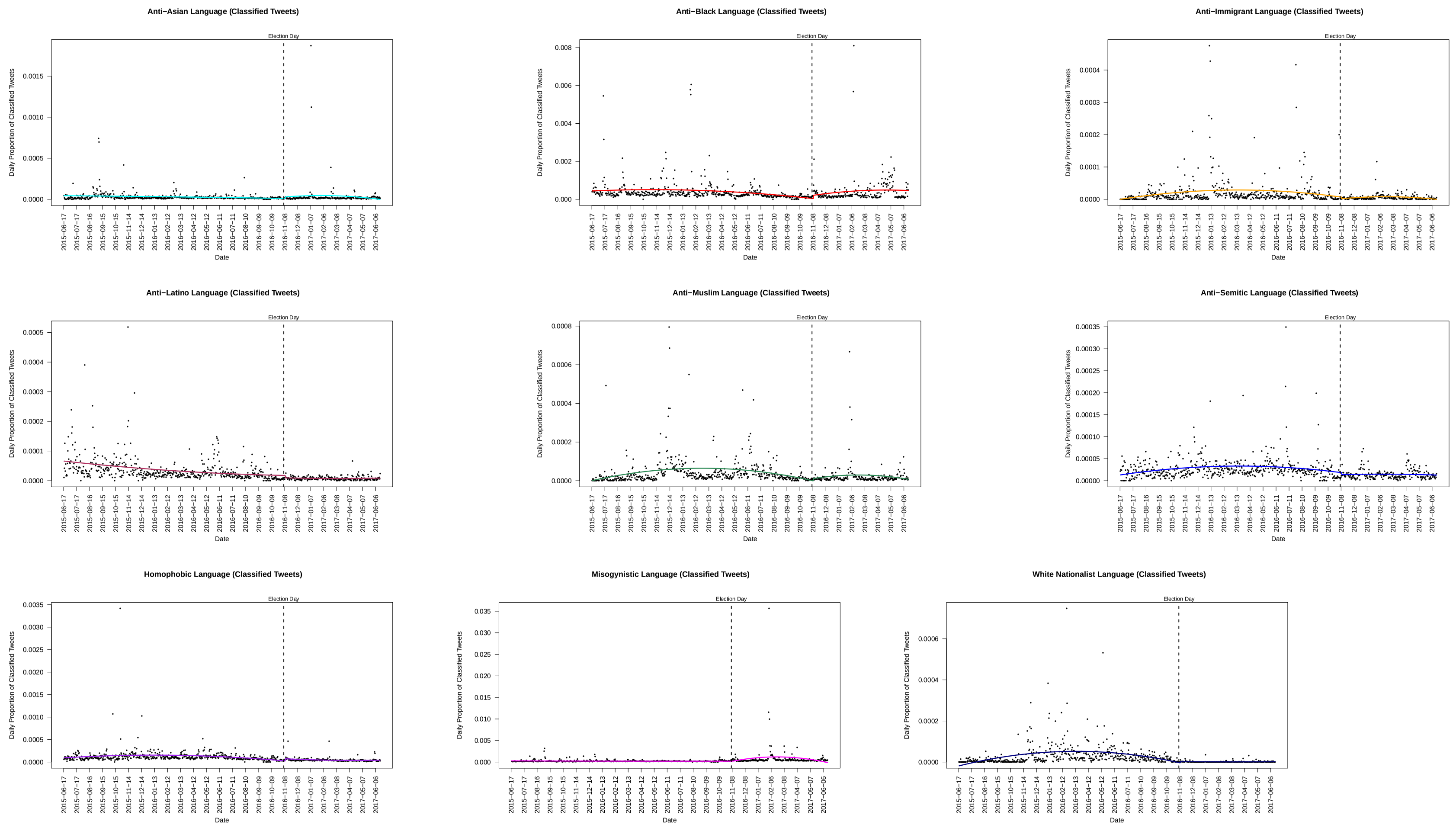
These plots show the predicted trend line of the daily proportion of unique users tweeting hate speech based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of unique users producing hate speech tweets (black dots) in the dataset of over 150 million tweets referencing Hillary Clinton collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A26: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Trump Dataset)



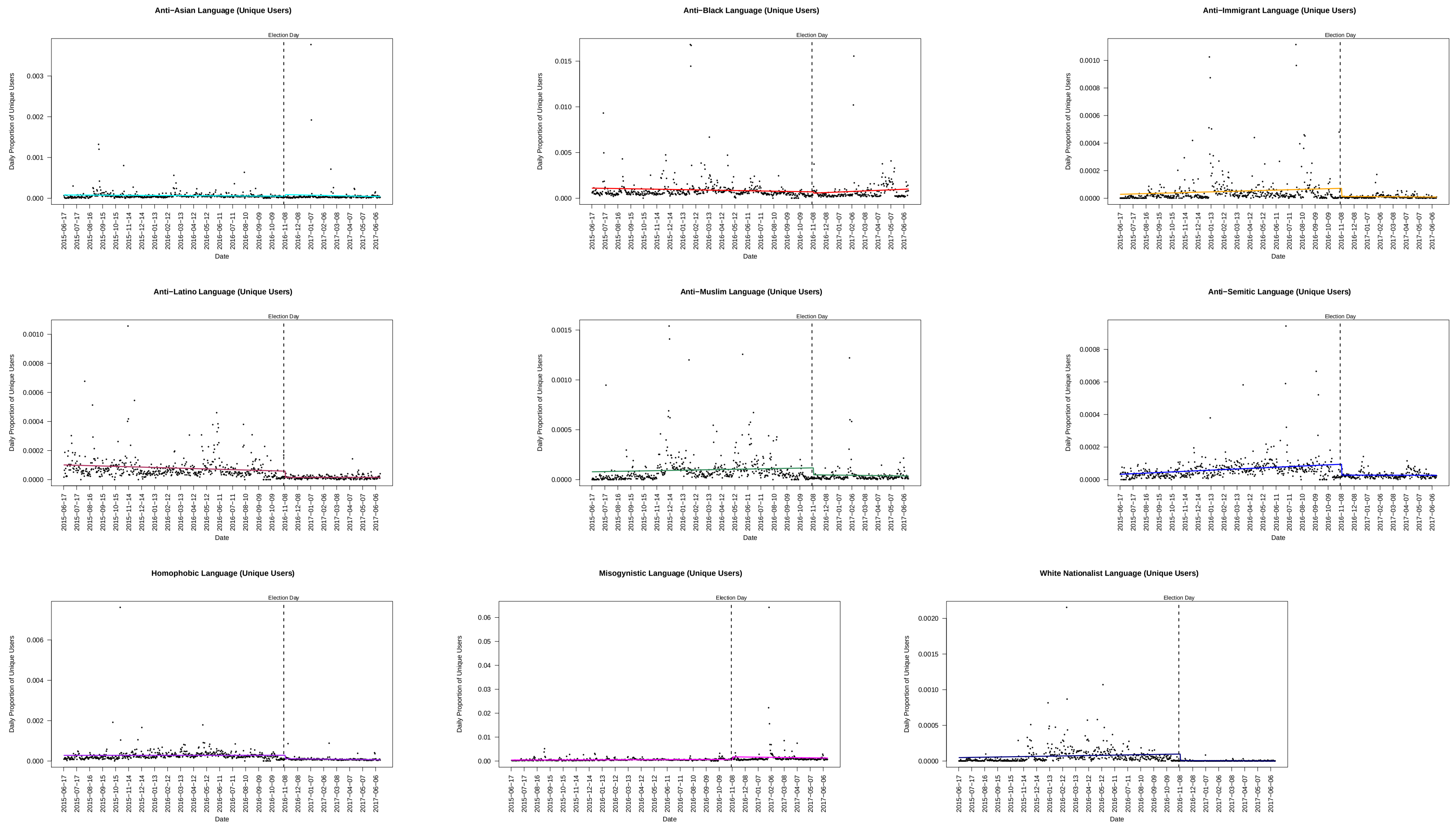
This plot shows the predicted trend lines based on our AR1 linear interrupted time series regression model plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in the dataset of over 600 million tweets referencing Donald Trump collected between June 17, 2015, and June 15, 2017.

Fig. A27: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language Including Quadratic Terms (Trump Dataset)



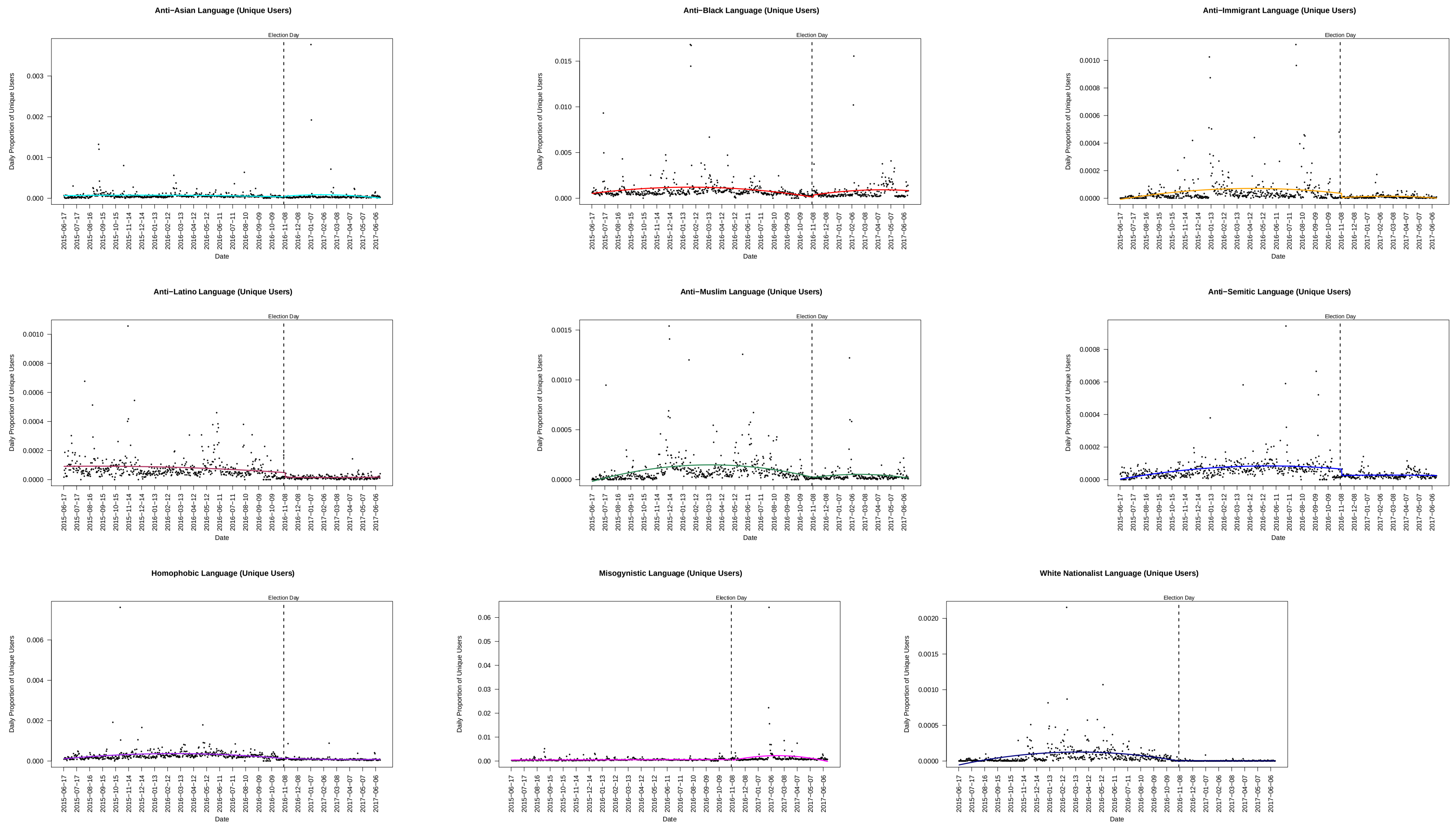
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model with quadratic terms, plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in the dataset of over 600 million tweets referencing Donald Trump collected between June 17, 2015, and June 15, 2017.hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A28: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Trump Dataset)



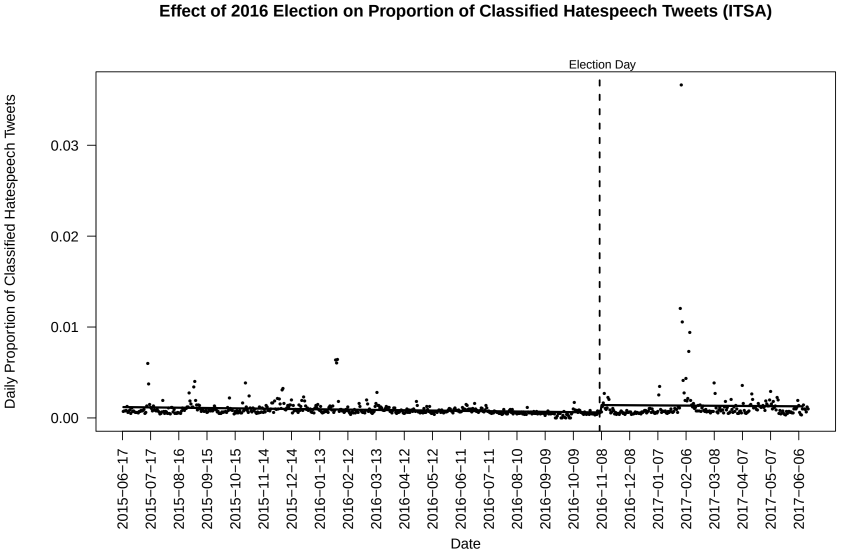
This plot shows the predicted trend lines based on our AR1 linear interrupted time series regression model plotted against the observed daily proportion of unique users producing hate speech and white nationalist rhetoric in a dataset of over 600 million tweets mentioning Donald Trump collected between June 17, 2015, and June 15, 2017.

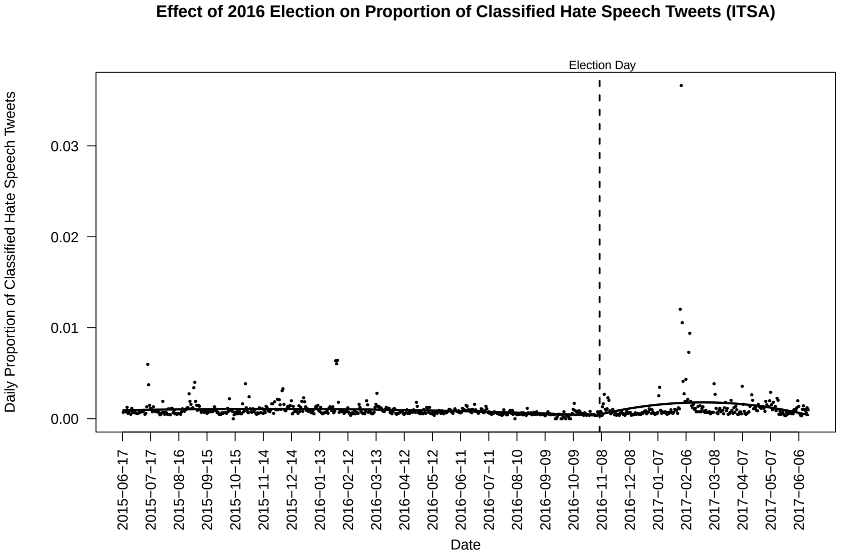
Fig. A29: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language Including Quadratic Terms (Trump Dataset)



This plot shows the predicted trend lines based on our AR1 interrupted time series regression model including quadratic terms plotted against the observed daily proportion of unique users producing hate speech and white nationalist rhetoric in a dataset of over 600 million tweets mentioning Donald Trump collected between June 17, 2015, and June 15, 2017.

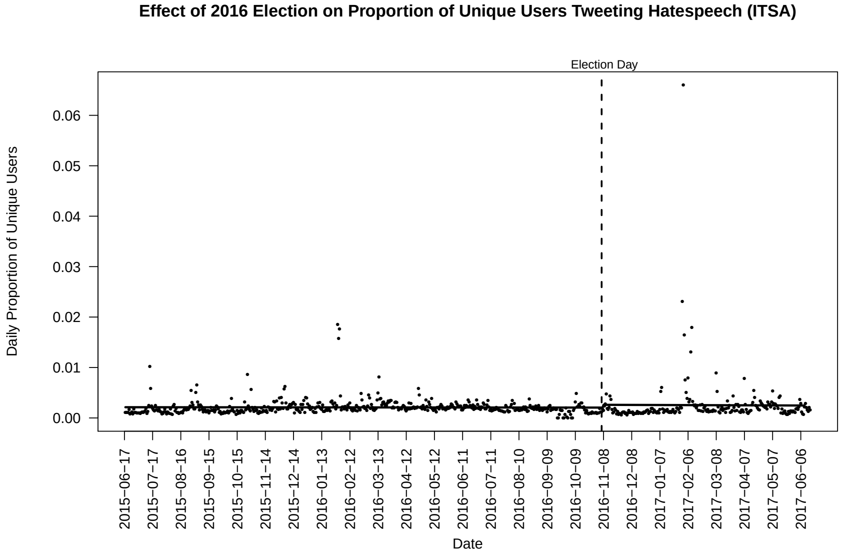
Fig. A30: Aggregate Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language
(Trump Dataset)

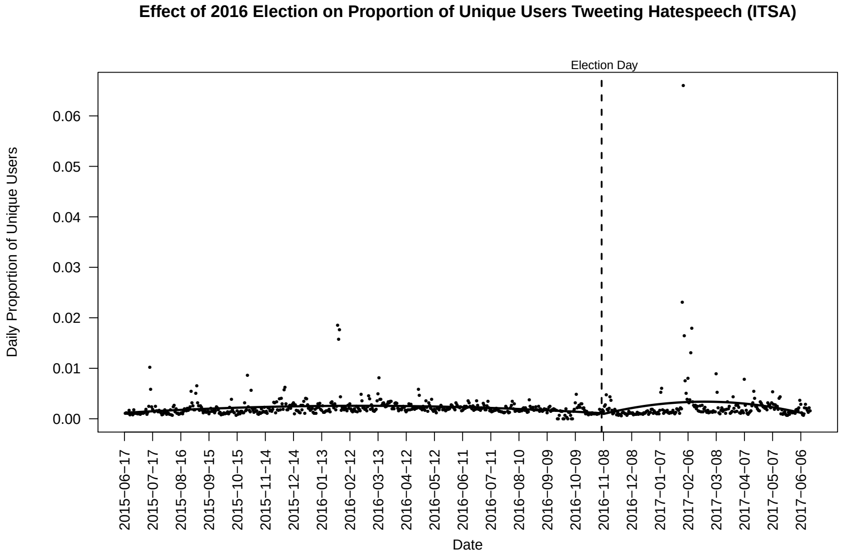




These plots show the predicted trend line of the proportion of hate speech tweets based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of hate speech tweets (black dots) in the dataset of over 600 million tweets referencing Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

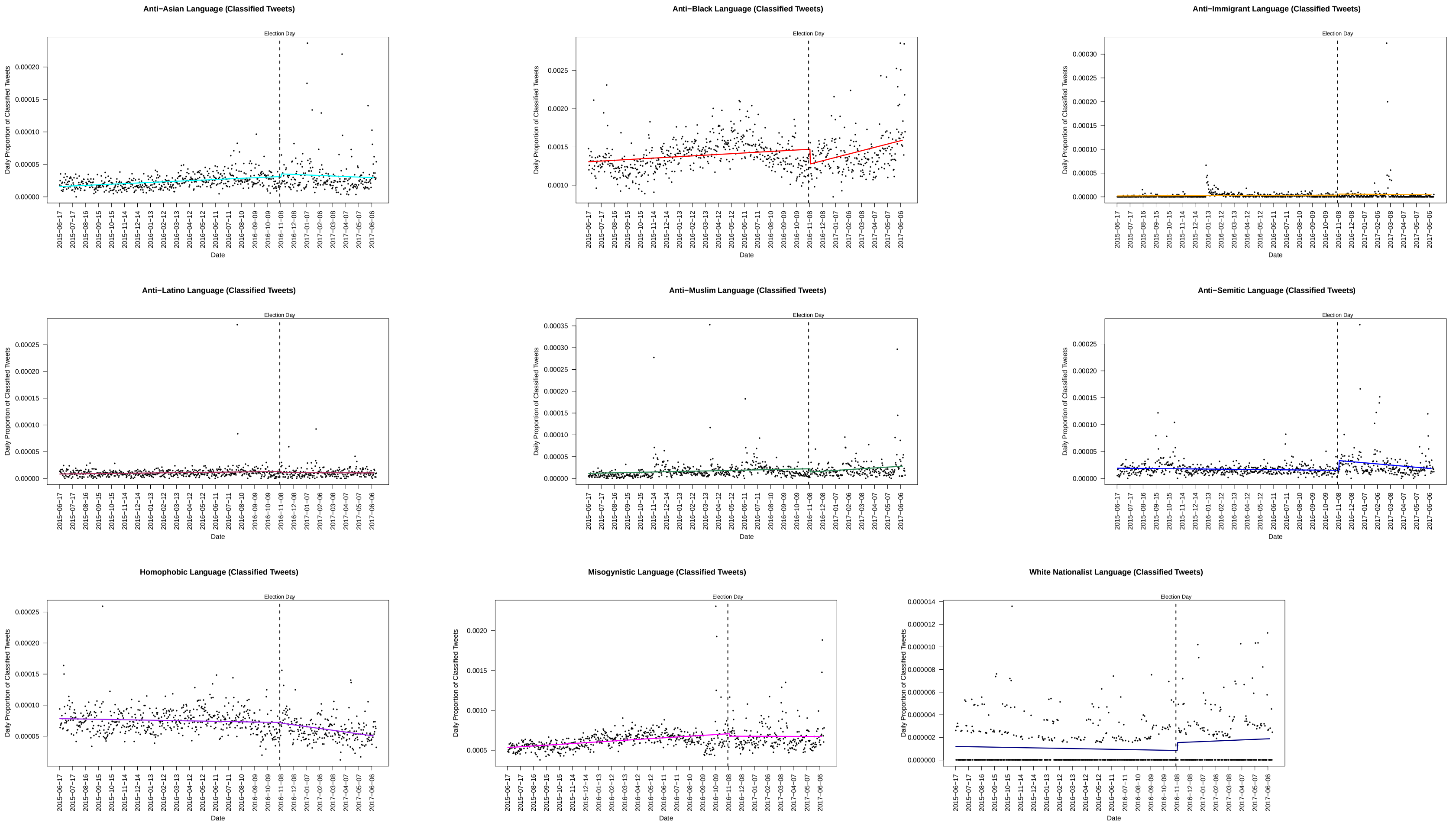
Fig. A31: Aggregate Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech
(Trump Dataset)





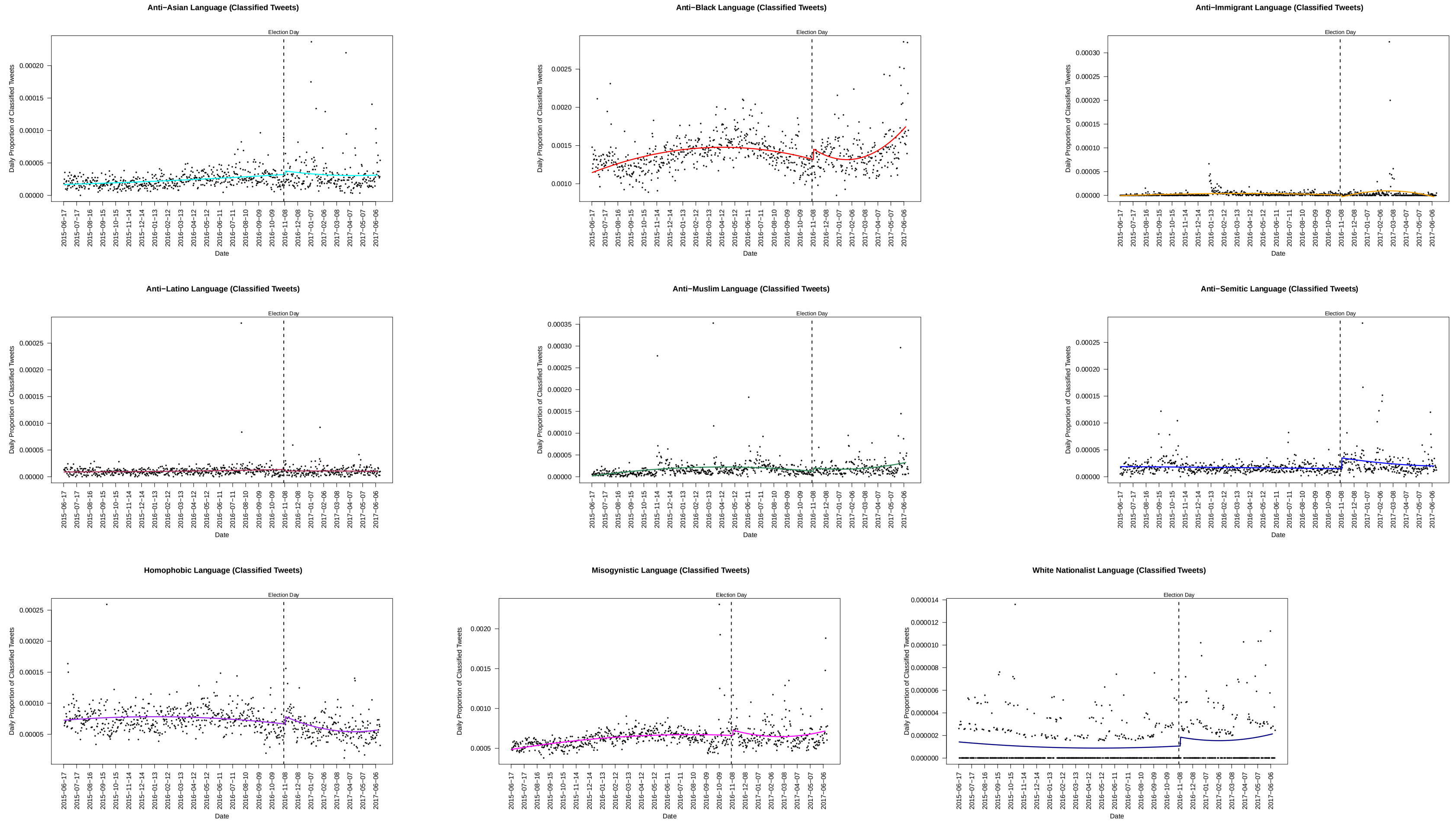
These plots show the predicted trend lines of the proportion of unique users tweeting hate speech based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of unique users (black dots) in the dataset tweeting hate speech in over 600 million tweets referencing Donald Trump collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A32: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Random Sample of Twitter Users)



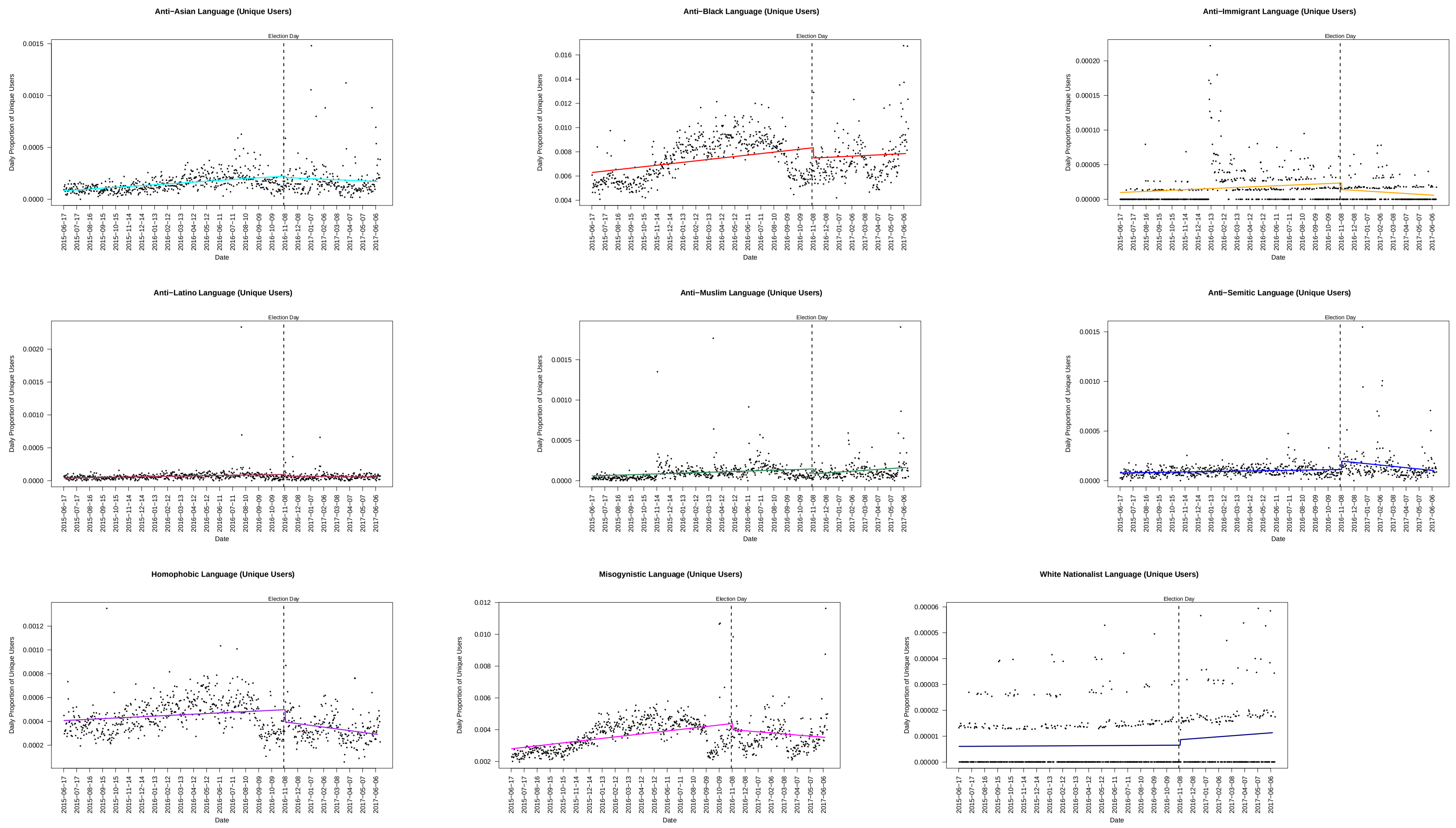
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in a dataset of over 400 million tweets collected from a random sample of 500,000 American Twitter users between June 17, 2015, and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A33: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language Including Quadratic Terms (Random Sample of Twitter Users)



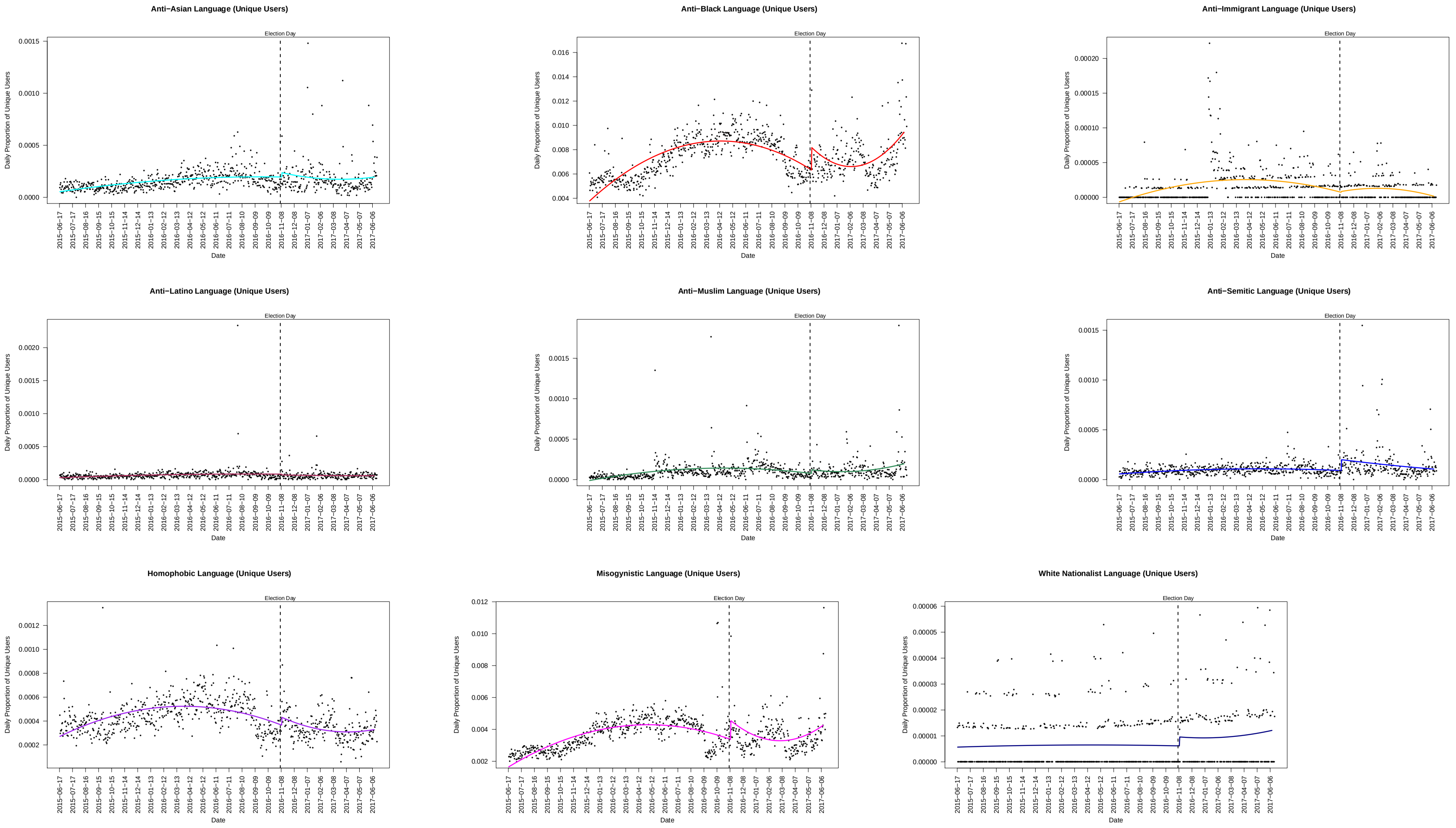
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model plotted against the observed daily proportion of each type of hate speech and white nationalist rhetoric in a dataset of over 400 million tweets collected from a random sample of 500,000 American Twitter users between June 17, 2015, and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A34: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Random Sample of Twitter Users)



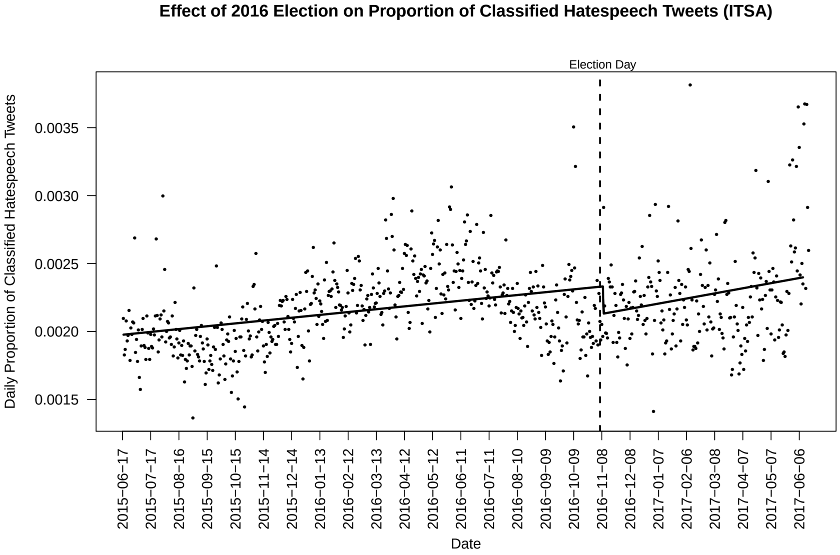
This plot shows the predicted trend lines based on our AR1 interrupted time series regression model plotted against the observed daily relative daily number of unique users tweeting hate speech and white nationalist rhetoric in a dataset of over 400 million tweets collected from a random sample of 500,000 American Twitter users between June 17, 2015, and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

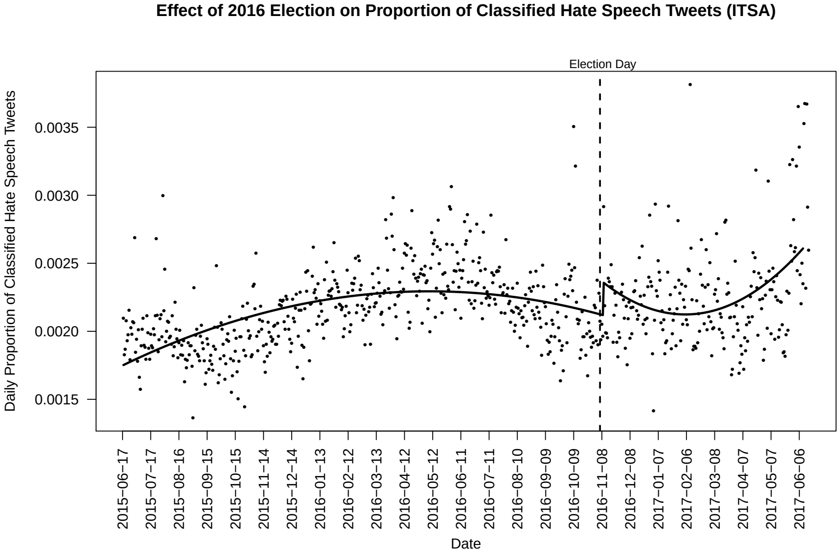
Fig. A35: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language Including Quadratic Terms (Random Sample of Twitter Users)



This plot shows the predicted trend lines based on our AR1 interrupted time series regression model plotted against the observed daily relative daily number of unique users tweeting hate speech and white nationalist rhetoric in a dataset of over 400 million tweets collected from a random sample of 500,000 American Twitter users between June 17, 2015, and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

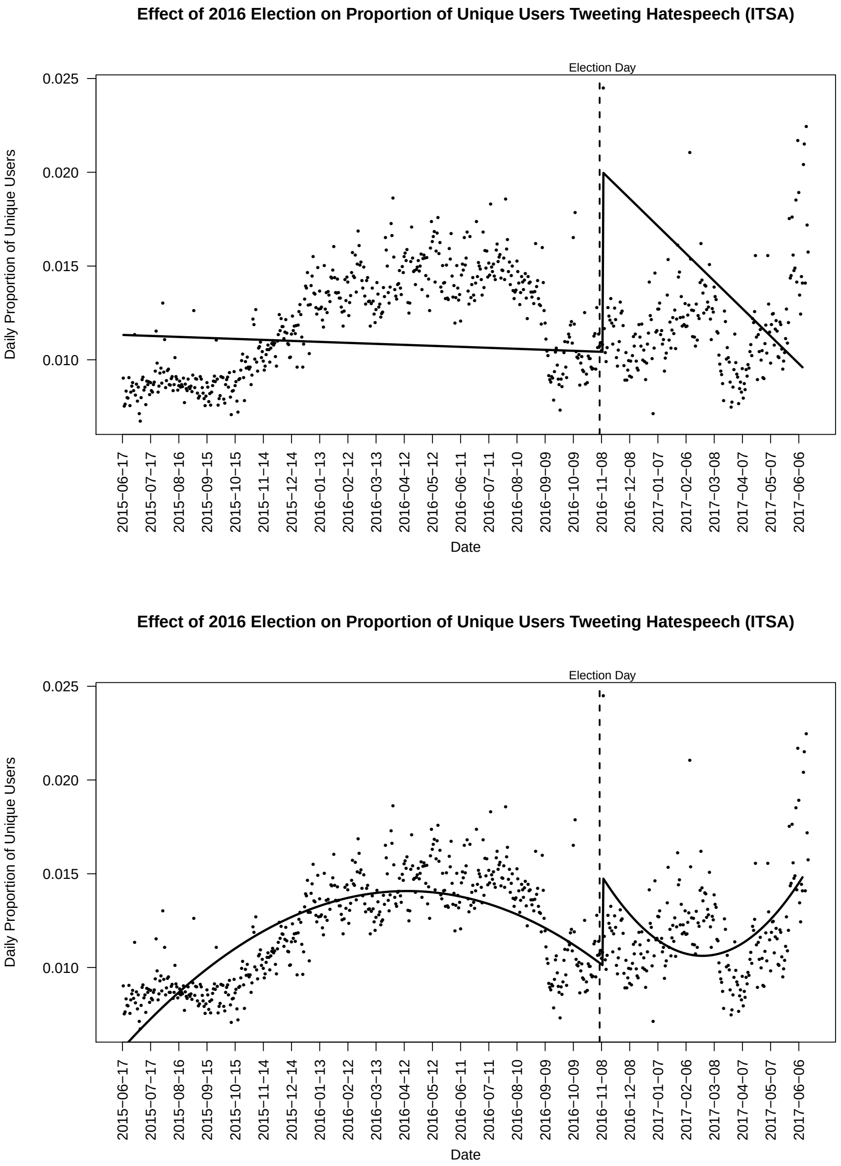
Fig. A36: Aggregate Effect of 2016 Election on Daily Proportion of Classified Hate Speech Tweets
(Random Sample Dataset)





These plots show the predicted trend line of the proportion of hate speech tweets based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of hate speech tweets (black dots) in the dataset of over 400 million tweets sent by a random sample of 500,000 American Twitter users collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

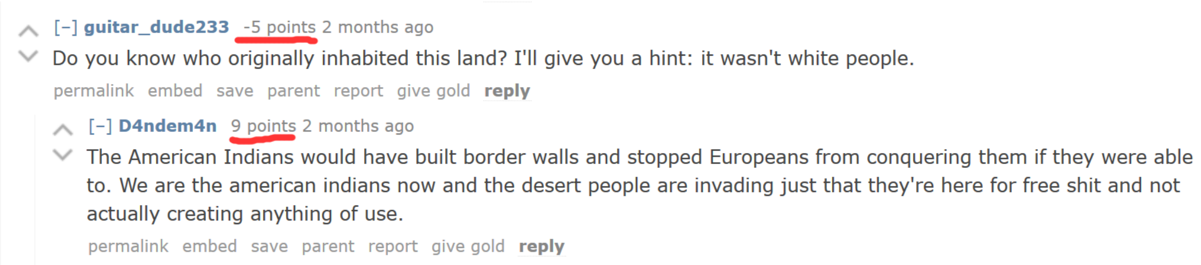
Fig. A37: Aggregate Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech
(Random Sample Dataset)


Effect of 2016 Election on Proportion of Unique Users Tweeting Hatespeech (ITSA)

Effect of 2016 Election on Proportion of Unique Users Tweeting Hatespeech (ITSA)

These plots show the predicted trend line of the proportion of unique users producing hate speech tweets based on our AR1 interrupted time series regression models. The top plot is the plain linear model (Model 1), while the bottom plots the model including quadratic terms (Model 2). Both plots also show the observed daily proportion of hate speech tweets (black dots) in the dataset of over 400 million tweets sent by a random sample of 500,000 American Twitter users collected using Twitter's Streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Fig. A38: Example of downvoted and upvoted comments



[–] guitar_dude233 -5 points 2 months ago

Do you know who originally inhabited this land? I'll give you a hint: it wasn't white people.

[–] D4ndem4n 9 points 2 months ago

The American Indians would have built border walls and stopped Europeans from conquering them if they were able to. We are the american indians now and the desert people are invading just that they're here for free shit and not actually creating anything of use.

Fig. A39: An architecture of fastText model

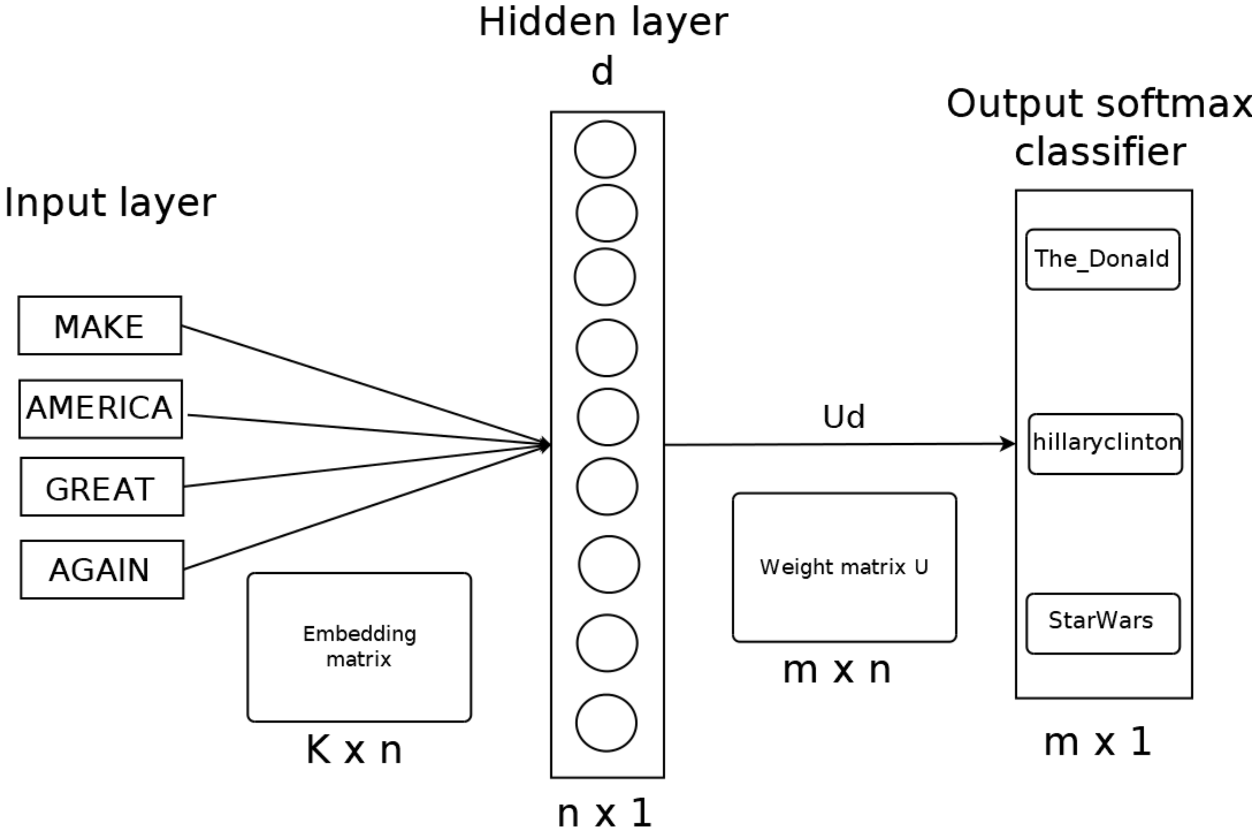


Fig. A40: Dendrogram of hierarchical clustering for individual subreddits

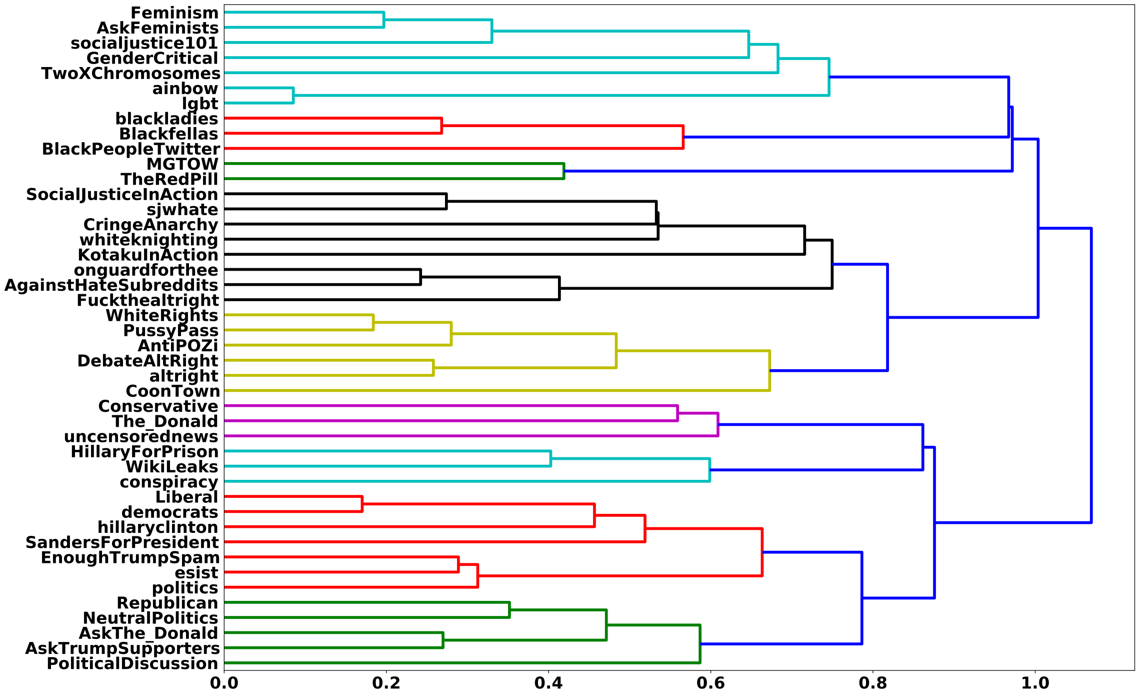


Fig. A41: t-SNE visualization of learned subreddit representations

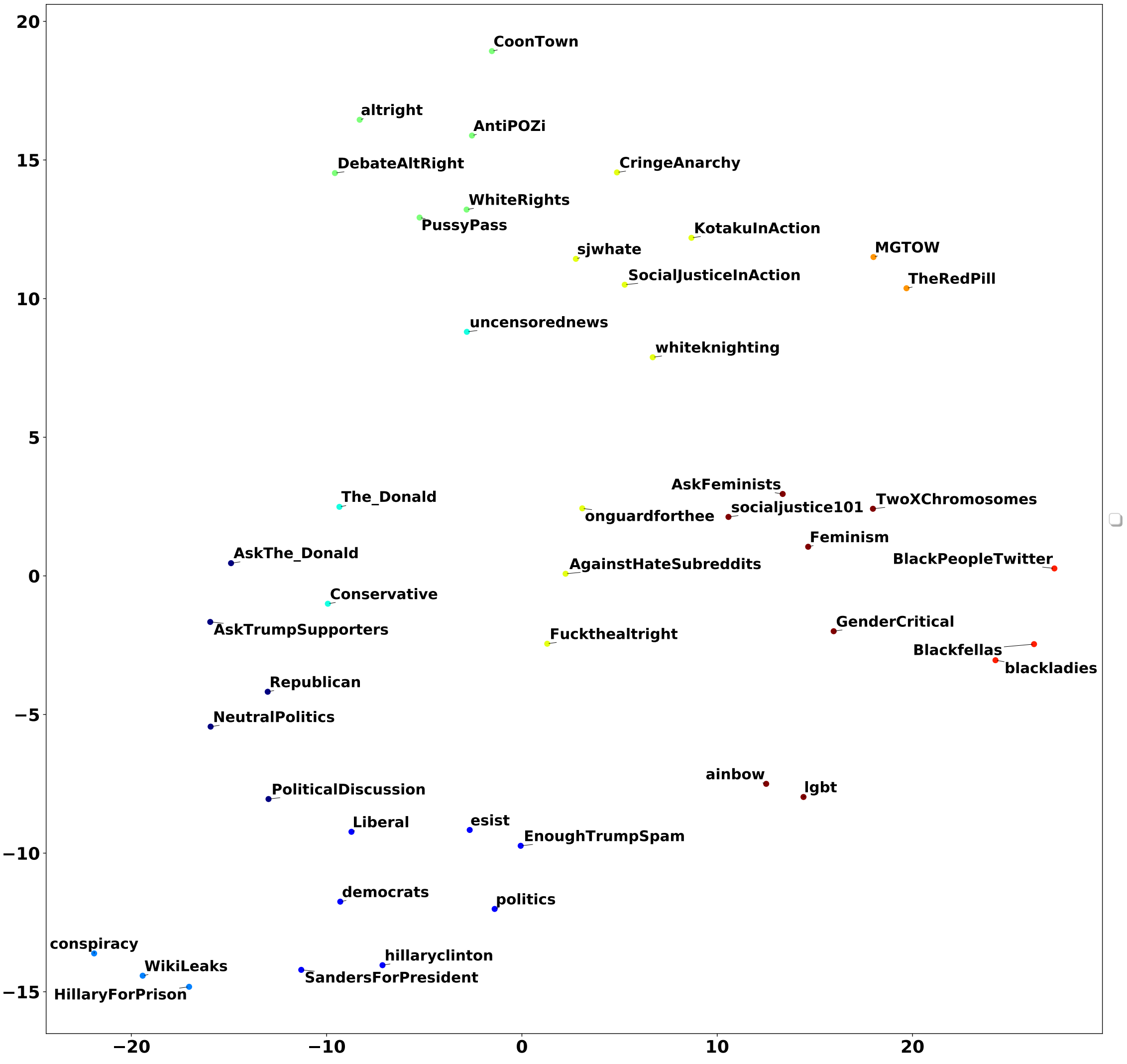


Fig. A42: Dendrogram of hierarchical clustering for subreddit groups

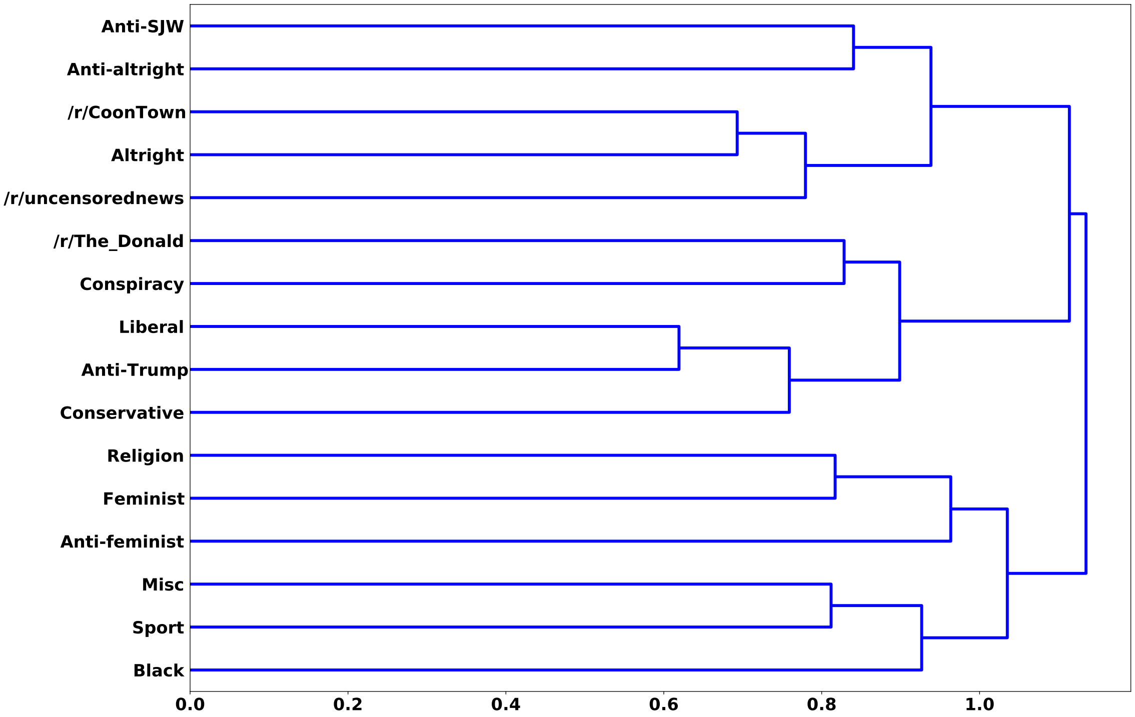
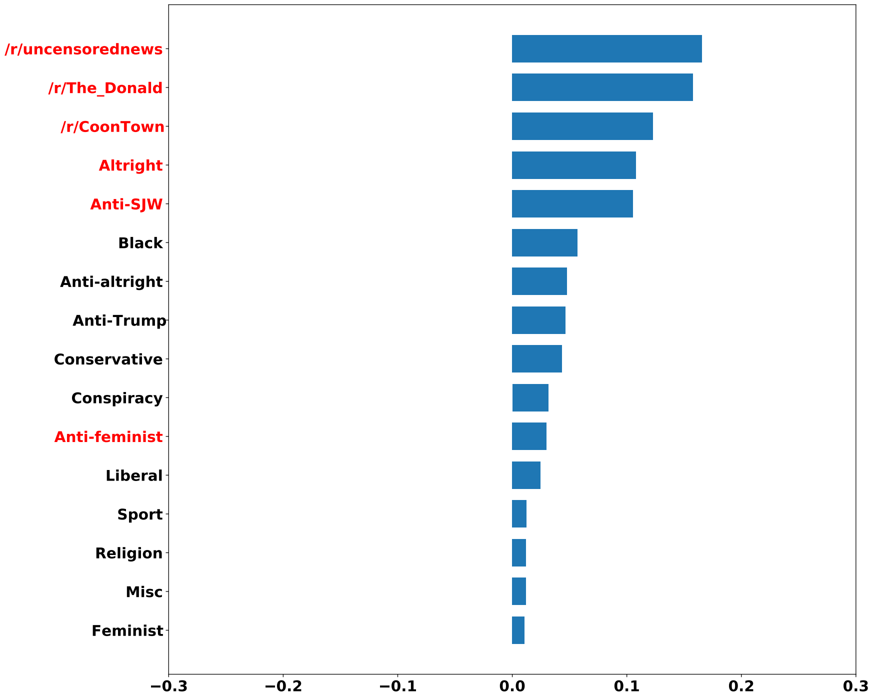
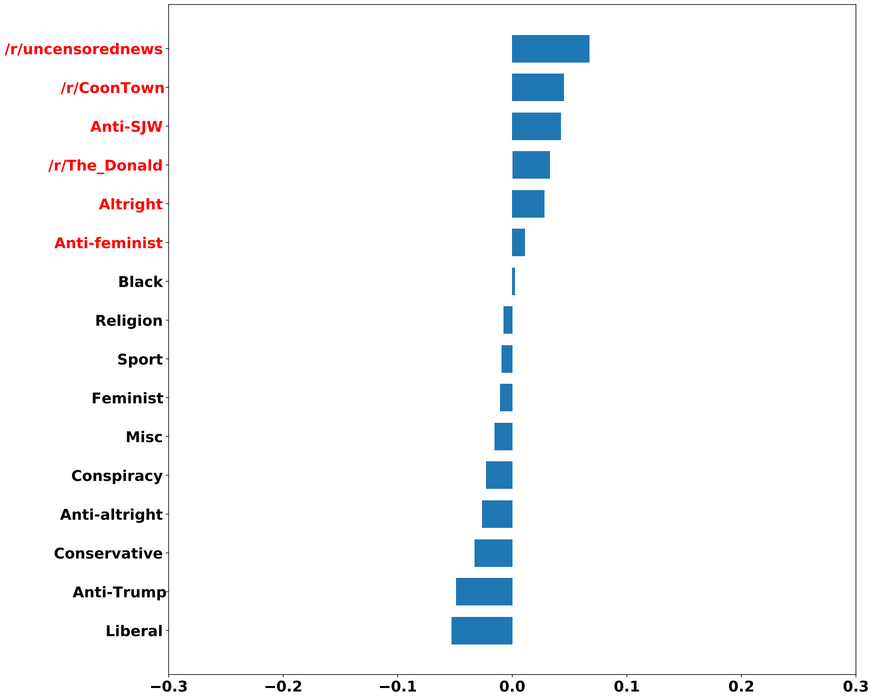


Fig. A43: Validation results for hate speech collection



(a) Average predicted probabilities for tweets classified as hate speech



(b) Difference in average predicted probabilities between tweets classified as hate speech and tweets coded as non-hate speech

Table A1: Hillary Clinton Collection Descriptive Statistics

	Tweet Type	**Mean**	**Median**	**SD**	**Min**	**Max**	**N**
Clinton Tweets Per Day	Anti Asian Clinton	185876.26	93854.50	233018.46	3602.00	1793464.00	724.00
Clinton Unique Users Per Day	Anti Asian Clinton	82656.10	55470.00	88195.99	3014.00	876630.00	724.00
Tweets Per Day 1	Anti Asian Clinton	16.94	6.00	33.44	0.00	421.00	730.00
Tweets Per Day 2	Anti Black Clinton	348.10	114.50	1333.04	0.00	33026.00	730.00
Tweets Per Day 3	Anti Immigrant Clinton	8.29	0.00	42.83	0.00	697.00	730.00
Tweets Per Day 4	Anti Latino Clinton	32.78	3.00	173.41	0.00	3146.00	730.00
Tweets Per Day 5	Anti Muslim Clinton	62.12	8.00	261.32	0.00	4946.00	730.00
Tweets Per Day 6	Anti Semitic Clinton	21.40	5.00	69.95	0.00	1077.00	730.00
Tweets Per Day 7	Homophobic Clinton	12.68	5.00	25.29	0.00	465.00	730.00
Tweets Per Day 8	Misogynistic Clinton	139.08	49.50	314.82	0.00	6395.00	730.00
Tweets Per Day 9	White Nationalist Clinton	46.35	12.00	168.32	0.00	3752.00	730.00
Tweets Per Day (Classified) 1	Anti Asian Clinton	4.48	2.00	12.42	0.00	207.00	730.00
Tweets Per Day (Classified) 2	Anti Black Clinton	26.04	7.00	70.82	0.00	836.00	730.00
Tweets Per Day (Classified) 3	Anti Immigrant Clinton	4.36	0.00	16.97	0.00	285.00	730.00
Tweets Per Day (Classified) 4	Anti Latino Clinton	1.18	0.00	3.53	0.00	68.00	730.00
Tweets Per Day (Classified) 5	Anti Muslim Clinton	11.48	2.00	57.56	0.00	1302.00	730.00
Tweets Per Day (Classified) 6	Anti Semitic Clinton	4.99	1.00	17.04	0.00	236.00	730.00
Tweets Per Day (Classified) 7	Homophobic Clinton	8.96	4.00	22.09	0.00	438.00	730.00
Tweets Per Day (Classified) 8	Misogynistic Clinton	95.77	34.00	195.21	0.00	2841.00	730.00
Tweets Per Day (Classified) 9	White Nationalist Clinton	1.57	0.00	8.25	0.00	184.00	730.00
Unique Users Per Day (Classified) 1	Anti Asian Clinton	4.39	2.00	12.35	0.00	207.00	730.00
Unique Users Per Day (Classified) 2	Anti Black Clinton	24.84	7.00	67.24	0.00	832.00	730.00
Unique Users Per Day (Classified) 3	Anti Immigrant Clinton	4.16	0.00	16.62	0.00	285.00	730.00
Unique Users Per Day (Classified) 4	Anti Latino Clinton	1.13	0.00	3.35	0.00	63.00	730.00
Unique Users Per Day (Classified) 5	Anti Muslim Clinton	10.23	2.00	49.32	0.00	1102.00	730.00
Unique Users Per Day (Classified) 6	Anti Semitic Clinton	4.68	1.00	16.13	0.00	233.00	730.00
Unique Users Per Day (Classified) 7	Homophobic Clinton	8.46	3.00	21.42	0.00	431.00	730.00
Unique Users Per Day (Classified) 8	Misogynistic Clinton	87.61	32.00	169.63	0.00	2479.00	730.00
Unique Users Per Day (Classified) 9	White Nationalist Clinton	1.46	0.00	8.11	0.00	183.00	730.00
Retweets (Classified) 1	Anti Asian Clinton	2.48	0.00	11.24	0.00	206.00	730.00
Retweets (Classified) 2	Anti Black Clinton	13.45	1.00	55.50	0.00	809.00	730.00
Retweets (Classified) 3	Anti Immigrant Clinton	3.41	0.00	15.99	0.00	280.00	730.00
Retweets (Classified) 4	Anti Latino Clinton	0.35	0.00	2.37	0.00	49.00	730.00
Retweets (Classified) 5	Anti Muslim Clinton	6.60	0.00	36.54	0.00	792.00	730.00
Retweets (Classified) 6	Anti Semitic Clinton	2.69	0.00	15.09	0.00	216.00	730.00
Retweets (Classified) 7	Homophobic Clinton	2.95	0.00	15.71	0.00	383.00	730.00
Retweets (Classified) 8	Misogynistic Clinton	37.41	6.00	130.91	0.00	2188.00	730.00
Retweets (Classified) 9	White Nationalist Clinton	1.23	0.00	8.05	0.00	183.00	730.00

Daily volume of tweets containing hate speech or white nationalist language in a collection of all tweets mentioning Hillary Clinton between June 17, 2015 and June 15, 2017. Tweets were collected using Twitter's Streaming API. Tweets were classified using a Naive Bayes Classifier to remove false positives.

Table A2: Donald Trump Collection Descriptive Statistics

	Tweet Type	Mean	Median	SD	Min	Max	N
Trump Tweets Per Day	Total Trump	828873.20	607988.00	646243.50	0.00	4671252.00	730.00
Trump Unique Users Per Day	Total Trump	324337.02	267243.00	226239.20	0.00	2724945.00	730.00
Tweets Per Day 1	Anti Asian Trump	165.81	79.00	400.29	0.00	5656.00	730.00
Tweets Per Day 2	Anti Black Trump	1912.25	1219.00	2425.29	0.00	31962.00	730.00
Tweets Per Day 3	Anti Immigrant Trump	49.39	14.00	161.48	0.00	2746.00	730.00
Tweets Per Day 4	Anti Latino Trump	62.56	34.50	89.14	0.00	984.00	730.00
Tweets Per Day 5	Anti Muslim Trump	178.07	69.00	381.59	0.00	4330.00	730.00
Tweets Per Day 6	Anti Semitic Trump	168.56	54.00	993.54	0.00	23075.00	730.00
Tweets Per Day 7	Homophobic Trump	110.11	63.00	122.61	0.00	801.00	730.00
Tweets Per Day 8	Misogynistic Trump	611.37	321.00	2100.39	0.00	50761.00	730.00
Tweets Per Day 9	White Nationalist Trump	327.19	143.00	961.45	0.00	22957.00	730.00
Tweets Per Day (Classified) 1	Anti Asian Trump	20.31	12.00	48.57	0.00	1103.00	730.00
Tweets Per Day (Classified) 2	Anti Black Trump	308.06	170.00	627.63	0.00	7493.00	730.00
Tweets Per Day (Classified) 3	Anti Immigrant Trump	13.84	4.00	44.86	0.00	785.00	730.00
Tweets Per Day (Classified) 4	Anti Latino Trump	19.12	11.00	26.97	0.00	243.00	730.00
Tweets Per Day (Classified) 5	Anti Muslim Trump	30.67	12.00	66.80	0.00	945.00	730.00
Tweets Per Day (Classified) 6	Anti Semitic Trump	19.72	11.00	27.56	0.00	325.00	730.00
Tweets Per Day (Classified) 7	Homophobic Trump	76.37	45.50	87.02	0.00	783.00	730.00
Tweets Per Day (Classified)8	Misogynistic Trump	367.55	161.50	2044.27	0.00	50477.00	730.00
Tweets Per Day (Classified)9	White Nationalist Trump	19.53	3.00	39.63	0.00	541.00	730.00
Unique Users Per Day (Classified) 1	Anti Asian Trump	19.44	11.00	43.35	0.00	941.00	730.00
Unique Users Per Day (Classified) 2	Anti Black Trump	298.13	165.00	620.58	0.00	7479.00	730.00
Unique Users Per Day (Classified) 3	Anti Immigrant Trump	13.35	4.00	44.30	0.00	785.00	730.00
Unique Users Per Day (Classified) 4	Anti Latino Trump	18.01	10.50	25.10	0.00	239.00	730.00
Unique Users Per Day (Classified) 5	Anti Muslim Trump	28.55	11.00	63.19	0.00	941.00	730.00
Unique Users Per Day (Classified) 6	Anti Semitic Trump	18.05	10.00	26.18	0.00	320.00	730.00
Unique Users Per Day (Classified) 7	Homophobic Trump	71.89	43.00	82.50	0.00	782.00	730.00
Unique Users Per Day (Classified) 8	Misogynistic Trump	352.36	153.00	2005.94	0.00	49503.00	730.00
Unique Users Per Day (Classified) 9	White Nationalist Trump	16.74	3.00	35.43	0.00	537.00	730.00
Retweets (Classified) 1	Anti Asian Trump	9.52	2.00	41.45	0.00	944.00	730.00
Retweets (Classified) 2	Anti Black Trump	185.94	59.00	561.27	0.00	7319.00	730.00
Retweets (Classified) 3	Anti Immigrant Trump	8.98	0.00	43.14	0.00	780.00	730.00
Retweets (Classified) 4	Anti Latino Trump	7.39	2.00	17.36	0.00	199.00	730.00
Retweets (Classified) 5	Anti Muslim Trump	16.58	3.00	54.24	0.00	908.00	730.00
Retweets (Classified) 6	Anti Semitic Trump	7.06	1.00	19.89	0.00	262.00	730.00
Retweets (Classified) 7	Homophobic Trump	32.48	12.00	59.36	0.00	764.00	730.00
Retweets (Classified) 8	Misogynistic Trump	215.70	34.50	1999.22	0.00	49746.00	730.00
Retweets (Classified) 9	White Nationalist Trump	15.93	1.00	36.89	0.00	534.00	730.00

Daily volume of tweets containing hate speech or white nationalist language in a collection of all tweets mentioning Donald Trump between June 17, 2015 and June 15, 2017. Tweets were collected using Twitter's Streaming API. Tweets were classified using a Naive Bayes Classifier to remove false positives.

Table A3: Random Sample of Twitter Users Descriptive Statistics

	Tweet Type	Mean	Median	SD	Min	Max	N
Random Sample Tweets Per Day	Total Random	443235.71	427582.50	101499.85	247568.00	777732.00	730.00
Random Sample Unique Users Per Day	Total Random	67259.48	68389.50	7897.66	48835.00	81897.00	730.00
Tweets Per Day 1	Anti Asian Random	331.92	318.00	96.22	165.00	1039.00	730.00
Tweets Per Day 2	Anti Black Random	3022.57	2853.00	1040.62	1410.00	18545.00	730.00
Tweets Per Day 3	Anti Immigrant Random	7.94	6.00	10.39	0.00	159.00	730.00
Tweets Per Day 4	Anti Latino Random	80.57	68.00	157.82	28.00	4097.00	730.00
Tweets Per Day 5	Anti Muslim Random	94.52	82.50	58.46	31.00	698.00	730.00
Tweets Per Day 6	Anti Semitic Random	119.13	114.00	46.49	51.00	839.00	730.00
Tweets Per Day 7	Homophobic Random	81.72	76.00	34.76	23.00	308.00	730.00
Tweets Per Day 8	Misogynistic Random	790.33	758.50	248.33	336.00	1986.00	730.00
Tweets Per Day 9	White Nationalist Random	125.93	114.00	63.15	41.00	949.00	730.00
Tweets Per Day (Classified) 1	Anti Asian Random	11.52	10.00	7.72	0.00	83.00	730.00
Tweets Per Day (Classified) 2	Anti Black Random	626.97	586.50	196.61	272.00	1262.00	730.00
Tweets Per Day (Classified) 3	Anti Immigrant Random	1.66	0.00	6.83	0.00	138.00	730.00
Tweets Per Day (Classified) 4	Anti Latino Random	4.77	4.00	6.43	0.00	150.00	730.00
Tweets Per Day (Classified) 5	Anti Muslim Random	8.33	6.00	11.61	0.00	192.00	730.00
Tweets Per Day (Classified) 6	Anti Semitic Random	8.51	7.00	7.32	0.00	84.00	730.00
Tweets Per Day (Classified) 7	Homophobic Random	32.04	31.00	13.37	3.00	107.00	730.00
Tweets Per Day (Classified) 8	Misogynistic Random	284.06	269.50	95.53	118.00	906.00	730.00
Tweets Per Day (Classified) 9	White Nationalist Random	0.53	0.00	0.82	0.00	6.00	730.00
Unique Users Per Day (Classified) 1	Anti Asian Random	10.91	9.00	7.38	0.00	83.00	730.00
Unique Users Per Day (Classified) 2	Anti Black Random	499.79	477.50	134.38	240.00	1032.00	730.00
Unique Users Per Day (Classified) 3	Anti Immigrant Random	1.02	0.00	1.73	0.00	16.00	730.00
Unique Users Per Day (Classified) 4	Anti Latino Random	4.47	4.00	6.29	0.00	149.00	730.00
Unique Users Per Day (Classified) 5	Anti Muslim Random	7.27	5.00	9.00	0.00	129.00	730.00
Unique Users Per Day (Classified) 6	Anti Semitic Random	7.38	6.00	6.26	0.00	82.00	730.00
Unique Users Per Day (Classified) 7	Homophobic Random	28.62	28.00	11.20	3.00	104.00	730.00
Unique Users Per Day (Classified) 8	Misogynistic Random	244.97	235.00	75.47	108.00	788.00	730.00
Unique Users Per Day (Classified) 9	White Nationalist Random	0.49	0.00	0.75	0.00	4.00	730.00
Retweets (Classified) 1	Anti Asian Random	0.67	0.00	0.87	0.00	6.00	730.00
Retweets (Classified) 2	Anti Black Random	282.24	269.00	113.40	98.00	906.00	730.00
Retweets (Classified) 3	Anti Immigrant Random	0.22	0.00	2.46	0.00	56.00	730.00
Retweets (Classified) 4	Anti Latino Random	0.35	0.00	0.65	0.00	8.00	730.00
Retweets (Classified) 5	Anti Muslim Random	0.40	0.00	0.87	0.00	14.00	730.00
Retweets (Classified) 6	Anti Semitic Random	0.41	0.00	0.71	0.00	6.00	730.00
Retweets (Classified) 7	Homophobic Random	2.21	2.00	2.07	0.00	18.00	730.00
Retweets (Classified) 8	Misogynistic Random	15.16	14.00	7.04	2.00	42.00	730.00
Retweets (Classified) 9	White Nationalist Random	0.01	0.00	0.10	0.00	1.00	730.00

Daily volume of tweets containing hate speech or white nationalist language in a collection of tweets by a random sample of American Twitter users tweeting between June 17, 2015 and June 15, 2017. Tweets were collected using Twitter's Streaming API. Tweets are grouped by whether they contain white nationalist or derogatory keywords directed at a particular group. Categories are not mutually exclusive. Classified tweets use a Naive Bayes Classifier to remove false positives.

Table A4: Top Dates Containing Hate Speech or White Nationalist Language (Clinton Dataset)

Top Dates	Proportion of Hate Speech Tweets
2016-04-14	0.0154
2016-12-16	0.0096
2016-12-24	0.0067
2017-01-03	0.0052
2015-11-22	0.0048
2016-10-09	0.0046
2016-10-28	0.0043
2016-09-25	0.0043
2015-12-10	0.0034
2016-10-01	0.0033

Table A5: Top Dates Containing Hate Speech or White Nationalist Language (Trump Dataset)

Top Dates	Proportion of Hate Speech Tweets
2017-01-31	0.0366
2017-01-30	0.0120
2017-02-01	0.0106
2017-02-09	0.0094
2017-02-08	0.0073
2016-01-29	0.0064
2016-01-31	0.0064
2016-01-30	0.0061
2015-07-13	0.0060
2017-02-05	0.0043

Table A6: Top Dates Containing Hate Speech or White Nationalist Language (Random Sample Dataset)

Top Dates	Proportion of Hate Speech Tweets
2017-02-09	0.0038
2017-06-13	0.0037
2017-06-11	0.0037
2017-06-04	0.0037
2017-06-10	0.0035
2016-10-08	0.0035
2017-06-05	0.0034
2017-05-29	0.0033
2017-05-26	0.0032
2017-06-02	0.0032

Table A7: Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Clinton Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$3.35 \times 10^{-5***}$	$5.93 \times 10^{-5**}$	1.01×10^{-6}	$7.82 \times 10^{-6***}$	$9.25 \times 10^{-5**}$	7.60×10^{-6}	$3.57 \times 10^{-5***}$	$3.21 \times 10^{-4***}$	-8.67×10^{-7}
	(1.01×10^{-5})	(2.16×10^{-5})	(7.29×10^{-6})	(1.73×10^{-6})	(2.96×10^{-5})	(6.65×10^{-6})	(7.01×10^{-6})	(6.40×10^{-5})	(5.70×10^{-6})
Pre-Election Trend	-1.81×10^{-8}	$2.00 \times 10^{-7**}$	$7.65 \times 10^{-8**}$	-2.99×10^{-9}	-4.26×10^{-8}	$6.75 \times 10^{-8**}$	3.74×10^{-8}	$6.98 \times 10^{-7**}$	1.29×10^{-8}
	(3.43×10^{-8})	(7.32×10^{-8})	(2.47×10^{-8})	(5.86×10^{-9})	(1.00×10^{-7})	(2.25×10^{-8})	(2.37×10^{-8})	(2.17×10^{-7})	(1.93×10^{-8})
Election Level Change	3.65×10^{-6}	3.55×10^{-5}	$-3.38 \times 10^{-5*}$	5.28×10^{-6}	-3.26×10^{-5}	-1.64×10^{-5}	$6.47 \times 10^{-5***}$	-1.60×10^{-4}	$5.28 \times 10^{-5***}$
	(1.85×10^{-5})	(3.96×10^{-5})	(1.33×10^{-5})	(3.17×10^{-6})	(5.39×10^{-5})	(1.21×10^{-5})	(1.29×10^{-5})	(1.17×10^{-4})	(1.04×10^{-5})
Election Slope Change	6.49×10^{-8}	$-9.03 \times 10^{-7***}$	-9.16×10^{-8}	-2.97×10^{-8}	-8.83×10^{-8}	-1.57×10^{-7}	$-5.42 \times 10^{-7***}$	$-1.92 \times 10^{-6*}$	$-1.61 \times 10^{-7*}$
	(1.27×10^{-7})	(2.71×10^{-7})	(9.12×10^{-8})	(2.17×10^{-8})	(3.70×10^{-7})	(8.31×10^{-8})	(8.79×10^{-8})	(8.02×10^{-7})	(7.13×10^{-8})
AIC	-1.13×10^{4}	-1.02×10^{4}	-1.19×10^{4}	-1.37×10^{4}	-1.01×10^{4}	-1.20×10^{4}	-1.16×10^{4}	-8.49×10^{3}	-1.21×10^{4}
BIC	-1.13×10^{4}	-1.01×10^{4}	-1.19×10^{4}	-1.36×10^{4}	-1.01×10^{4}	-1.20×10^{4}	-1.15×10^{4}	-8.46×10^{3}	-1.21×10^{4}
Log Likelihood	5.66×10^{3}	5.09×10^{3}	5.95×10^{3}	6.83×10^{3}	5.05×10^{3}	6.01×10^{3}	5.79×10^{3}	4.25×10^{3}	6.05×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A8: Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language Including Quadratic Terms (Clinton Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	6.57×10^{-6}	$6.92 \times 10^{-5*}$	-3.30×10^{-6}	$8.20 \times 10^{-6**}$	3.88×10^{-5}	1.27×10^{-5}	$3.24 \times 10^{-5**}$	$2.10 \times 10^{-4*}$	3.77×10^{-6}
	(1.50×10^{-5})	(3.26×10^{-5})	(1.10×10^{-5})	(2.61×10^{-6})	(4.44×10^{-5})	(1.00×10^{-5})	(1.06×10^{-5})	(9.60×10^{-5})	(8.59×10^{-6})
Pre-Election Trend	$2.97 \times 10^{-7*}$	8.46×10^{-8}	1.27×10^{-7}	-7.36×10^{-9}	5.87×10^{-7}	8.06×10^{-9}	7.63×10^{-8}	$2.00 \times 10^{-6*}$	-4.14×10^{-8}
	(1.35×10^{-7})	(2.94×10^{-7})	(9.92×10^{-8})	(2.36×10^{-8})	(4.00×10^{-7})	(9.04×10^{-8})	(9.55×10^{-8})	(8.66×10^{-7})	(7.75×10^{-8})
Election Level Change	7.05×10^{-6}	3.79×10^{-5}	-3.27×10^{-5}	6.04×10^{-6}	1.71×10^{-5}	-2.38×10^{-5}	$6.53 \times 10^{-5***}$	-1.24×10^{-4}	$4.83 \times 10^{-5**}$
	(2.76×10^{-5})	(6.01×10^{-5})	(2.02×10^{-5})	(4.83×10^{-6})	(8.05×10^{-5})	(1.84×10^{-5})	(1.96×10^{-5})	(1.77×10^{-4})	(1.58×10^{-5})
Election Slope Change	9.88×10^{-7}	-1.34×10^{-6}	4.27×10^{-8}	-6.41×10^{-8}	5.94×10^{-7}	-1.48×10^{-7}	-4.32×10^{-7}	1.32×10^{-6}	-2.15×10^{-7}
	(5.05×10^{-7})	(1.10×10^{-6})	(3.70×10^{-7})	(8.81×10^{-8})	(1.49×10^{-6})	(3.37×10^{-7})	(3.57×10^{-7})	(3.24×10^{-6})	(2.89×10^{-7})
Pre-Elelction Trend^2	$-6.15 \times 10^{-10*}$	2.25×10^{-10}	-9.90×10^{-11}	9.00×10^{-12}	-1.23×10^{-9}	1.16×10^{-10}	-7.60×10^{-11}	-2.55×10^{-9}	1.06×10^{-10}
	(2.56×10^{-10})	(5.57×10^{-10})	(1.88×10^{-10})	(4.50×10^{-11})	(7.56×10^{-10})	(1.71×10^{-10})	(1.81×10^{-10})	(1.64×10^{-9})	(1.47×10^{-10})
Election Slope Change^2	-2.13×10^{-9}	1.22×10^{-9}	-2.80×10^{-10}	1.27×10^{-10}	9.89×10^{-10}	-4.21×10^{-10}	-2.43×10^{-10}	-6.22×10^{-9}	-1.04×10^{-10}
	(2.14×10^{-9})	(4.65×10^{-9})	(1.56×10^{-9})	(3.73×10^{-10})	(6.28×10^{-9})	(1.43×10^{-9})	(1.51×10^{-9})	(1.37×10^{-8})	(1.23×10^{-9})
AIC	-1.12×10^{4}	-1.01×10^{4}	-1.18×10^{4}	-1.36×10^{4}	-1.00×10^{4}	-1.19×10^{4}	-1.15×10^{4}	-8.42×10^{3}	-1.20×10^{4}
BIC	-1.12×10^{4}	-1.00×10^{4}	-1.18×10^{4}	-1.35×10^{4}	-9.98×10^{3}	-1.19×10^{4}	-1.14×10^{4}	-8.38×10^{3}	-1.20×10^{4}
Log Likelihood	5.62×10^{3}	5.05×10^{3}	5.91×10^{3}	6.79×10^{3}	5.02×10^{3}	5.97×10^{3}	5.75×10^{3}	4.22×10^{3}	6.01×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model including quadratic terms, demonstrating the effect of Trump's election on the daily proportion of hate speech or white nationalist language tweets in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A9: Aggregate Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Clinton Data)

	Model 1	Model 2
Baseline	$5.59 \times 10^{-4***}$	$3.75 \times 10^{-4**}$
	(8.26×10^{-5})	(1.23×10^{-4})
Pre-Election Trend	$1.01 \times 10^{-6***}$	$3.18 \times 10^{-6**}$
	(2.79×10^{-7})	(1.11×10^{-6})
Election Level Change	-1.37×10^{-4}	-4.45×10^{-5}
	(1.51×10^{-4})	(2.25×10^{-4})
Election Slope Change	$-3.67 \times 10^{-6***}$	9.27×10^{-7}
	(1.03×10^{-6})	(4.13×10^{-6})
Pre-Election Trend^2		$-4.00 \times 10^{-9*}$
		(2.00×10^{-9})
Election Slope Change^2		-7.00×10^{-9}
		(1.80×10^{-8})
AIC	-8.13×10^{3}	-8.06×10^{3}
BIC	-8.11×10^{3}	-8.03×10^{3}
Log Likelihood	4.07×10^{3}	4.04×10^{3}
Num. obs.	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion hate speech tweets in a dataset of over 125 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A10: Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language (Clinton Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$5.56 \times 10^{-5**}$	5.67×10^{-5}	-6.48×10^{-6}	$1.13 \times 10^{-5***}$	$1.07 \times 10^{-4*}$	5.97×10^{-7}	$4.65 \times 10^{-5***}$	$3.72 \times 10^{-4**}$	-3.52×10^{-6}
	(1.95×10^{-5})	(4.11×10^{-5})	(1.41×10^{-5})	(2.85×10^{-6})	(4.51×10^{-5})	(1.61×10^{-5})	(1.13×10^{-5})	(1.17×10^{-4})	(6.72×10^{-6})
Pre-Election Trend	7.90×10^{-9}	$7.12 \times 10^{-7***}$	$2.02 \times 10^{-7***}$	1.16×10^{-8}	1.34×10^{-7}	$2.11 \times 10^{-7***}$	$1.79 \times 10^{-7***}$	$2.50 \times 10^{-6***}$	3.73×10^{-8}
	(6.59×10^{-8})	(1.39×10^{-7})	(4.78×10^{-8})	(9.65×10^{-9})	(1.52×10^{-7})	(5.44×10^{-8})	(3.84×10^{-8})	(3.97×10^{-7})	(2.27×10^{-8})
Election Level Change	-1.97×10^{-5}	-1.40×10^{-4}	$-8.70 \times 10^{-5***}$	-1.51×10^{-6}	-1.29×10^{-4}	$-7.35 \times 10^{-5*}$	1.70×10^{-5}	$-9.51 \times 10^{-4***}$	$5.22 \times 10^{-5***}$
	(3.56×10^{-5})	(7.52×10^{-5})	(2.58×10^{-5})	(5.23×10^{-6})	(8.11×10^{-5})	(2.93×10^{-5})	(2.08×10^{-5})	(2.15×10^{-4})	(1.23×10^{-5})
Election Slope Change	5.21×10^{-8}	$-1.70 \times 10^{-6**}$	-2.27×10^{-7}	-5.77×10^{-8}	-2.87×10^{-7}	-3.25×10^{-7}	$-8.23 \times 10^{-7***}$	$-4.03 \times 10^{-6**}$	$-1.90 \times 10^{-7*}$
	(2.44×10^{-7})	(5.15×10^{-7})	(1.77×10^{-7})	(3.57×10^{-8})	(5.60×10^{-7})	(2.01×10^{-7})	(1.42×10^{-7})	(1.47×10^{-6})	(8.41×10^{-8})
AIC	-1.04×10^{4}	-9.28×10^{3}	-1.09×10^{4}	-1.30×10^{4}	-9.84×10^{3}	-1.09×10^{4}	-1.09×10^{4}	-7.70×10^{3}	-1.19×10^{4}
BIC	-1.04×10^{4}	-9.25×10^{3}	-1.09×10^{4}	-1.30×10^{4}	-9.81×10^{3}	-1.09×10^{4}	-1.09×10^{4}	-7.67×10^{3}	-1.19×10^{4}
Log Likelihood	5.20×10^{3}	4.64×10^{3}	5.45×10^{3}	6.50×10^{3}	4.93×10^{3}	5.45×10^{3}	5.47×10^{3}	3.85×10^{3}	5.97×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of a linear AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech or white nationalist language in a dataset of over 125 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech and white nationalist tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A11: Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech or White Nationalist Language Including Quadratic Terms (Clinton Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	3.85×10^{-6}	1.13×10^{-4}	-2.30×10^{-6}	$1.29 \times 10^{-5**}$	6.31×10^{-5}	1.64×10^{-5}	$4.27 \times 10^{-5*}$	2.46×10^{-4}	8.45×10^{-6}
	(2.89×10^{-5})	(6.18×10^{-5})	(2.13×10^{-5})	(4.30×10^{-6})	(6.79×10^{-5})	(2.42×10^{-5})	(1.71×10^{-5})	(1.77×10^{-4})	(1.01×10^{-5})
Pre-Election Trend	$6.14 \times 10^{-7*}$	5.66×10^{-8}	1.53×10^{-7}	-7.14×10^{-9}	6.52×10^{-7}	2.51×10^{-8}	2.24×10^{-7}	$3.98 \times 10^{-6*}$	-1.03×10^{-7}
	(2.61×10^{-7})	(5.58×10^{-7})	(1.92×10^{-7})	(3.88×10^{-8})	(6.12×10^{-7})	(2.19×10^{-7})	(1.54×10^{-7})	(1.60×10^{-6})	(9.10×10^{-8})
Election Level Change	1.54×10^{-6}	-1.71×10^{-4}	$-9.47 \times 10^{-5*}$	-1.13×10^{-6}	-8.44×10^{-5}	$-9.14 \times 10^{-5*}$	2.39×10^{-5}	$-9.13 \times 10^{-4**}$	$3.87 \times 10^{-5*}$
	(5.31×10^{-5})	(1.14×10^{-4})	(3.92×10^{-5})	(7.93×10^{-6})	(1.20×10^{-4})	(4.42×10^{-5})	(3.16×10^{-5})	(3.25×10^{-4})	(1.85×10^{-5})
Election Slope Change	1.44×10^{-6}	-2.98×10^{-6}	-1.75×10^{-7}	-1.28×10^{-7}	1.59×10^{-7}	-4.41×10^{-7}	-8.61×10^{-7}	-3.01×10^{-7}	-2.80×10^{-7}
	(9.73×10^{-7})	(2.08×10^{-6})	(7.18×10^{-7})	(1.45×10^{-7})	(2.26×10^{-6})	(8.14×10^{-7})	(5.77×10^{-7})	(5.96×10^{-6})	(3.40×10^{-7})
Pre-Elelction Trend^2	$-1.18 \times 10^{-9*}$	1.28×10^{-9}	9.50×10^{-11}	3.70×10^{-11}	-1.01×10^{-9}	3.62×10^{-10}	-8.80×10^{-11}	-2.89×10^{-9}	2.74×10^{-10}
	(4.93×10^{-10})	(1.06×10^{-9})	(3.64×10^{-10})	(7.30×10^{-11})	(1.16×10^{-9})	(4.13×10^{-10})	(2.92×10^{-10})	(3.02×10^{-9})	(1.72×10^{-10})
Election Slope Change^2	-2.33×10^{-9}	1.58×10^{-9}	-5.48×10^{-10}	1.97×10^{-10}	1.35×10^{-9}	-6.76×10^{-10}	4.56×10^{-10}	-7.31×10^{-9}	-4.97×10^{-10}
	(4.12×10^{-9})	(8.81×10^{-9})	(3.04×10^{-9})	(6.14×10^{-10})	(9.48×10^{-9})	(3.44×10^{-9})	(2.45×10^{-9})	(2.52×10^{-8})	(1.44×10^{-9})
AIC	-1.03×10^{4}	-9.20×10^{3}	-1.08×10^{4}	-1.29×10^{4}	-9.76×10^{3}	-1.08×10^{4}	-1.08×10^{4}	-7.62×10^{3}	-1.18×10^{4}
BIC	-1.03×10^{4}	-9.16×10^{3}	-1.08×10^{4}	-1.29×10^{4}	-9.73×10^{3}	-1.08×10^{4}	-1.08×10^{4}	-7.59×10^{3}	-1.18×10^{4}
Log Likelihood	5.17×10^{3}	4.61×10^{3}	5.41×10^{3}	6.46×10^{3}	4.89×10^{3}	5.41×10^{3}	5.43×10^{3}	3.82×10^{3}	5.93×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model including quadratic terms, demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 150 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A12: Aggregate Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech (Clinton Data)

	Model 1	Model 2
Baseline	$6.44 \times 10^{-4***}$	$4.96 \times 10^{-4*}$
	(1.47×10^{-4})	(2.22×10^{-4})
Pre-Election Trend	$3.96 \times 10^{-6***}$	$5.69 \times 10^{-6**}$
	(4.98×10^{-7})	(2.00×10^{-6})
Election Level Change	$-1.40 \times 10^{-3***}$	$-1.34 \times 10^{-3***}$
	(2.68×10^{-4})	(4.04×10^{-4})
Election Slope Change	$-7.38 \times 10^{-6***}$	-3.25×10^{-6}
	(1.84×10^{-6})	(7.41×10^{-6})
Pre-Election Trend^2		-3.00×10^{-9}
		(4.00×10^{-9})
Election Slope Change^2		-8.00×10^{-9}
		(3.20×10^{-8})
AIC	-7.38×10^{3}	-7.31×10^{3}
BIC	-7.35×10^{3}	-7.27×10^{3}
Log Likelihood	3.70×10^{3}	3.66×10^{3}
Num. obs.	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech or white nationalist language in a dataset of almost 150 million tweets mentioning Hillary Clinton collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech and white nationalist tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A13: Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Rhetoric (Trump Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$4.55 \times 10^{-5***}$	$5.90 \times 10^{-4***}$	$1.69 \times 10^{-5**}$	$6.06 \times 10^{-5***}$	$4.35 \times 10^{-5***}$	$2.47 \times 10^{-5***}$	$1.55 \times 10^{-4***}$	2.55×10^{-4}	$2.35 \times 10^{-5***}$
	(1.35×10^{-5})	(1.07×10^{-4})	(5.90×10^{-6})	(4.73×10^{-6})	(1.20×10^{-5})	(3.18×10^{-6})	(1.50×10^{-5})	(2.27×10^{-4})	(6.45×10^{-6})
Pre-Election Trend	-6.48×10^{-8}	$-7.93 \times 10^{-7*}$	1.27×10^{-8}	$-9.49 \times 10^{-8***}$	-3.70×10^{-9}	1.00×10^{-8}	$-1.35 \times 10^{-7**}$	-9.31×10^{-8}	2.80×10^{-8}
	(4.57×10^{-8})	(3.60×10^{-7})	(1.99×10^{-8})	(1.60×10^{-8})	(4.03×10^{-8})	(1.08×10^{-8})	(5.06×10^{-8})	(7.68×10^{-7})	(2.18×10^{-8})
Election Level Change	3.62×10^{-5}	1.10×10^{-4}	-1.63×10^{-5}	-3.99×10^{-6}	-1.52×10^{-5}	$-1.48 \times 10^{-5*}$	-2.41×10^{-5}	7.50×10^{-4}	$-3.54 \times 10^{-5**}$
	(2.44×10^{-5})	(1.89×10^{-4})	(1.06×10^{-5})	(8.57×10^{-6})	(2.12×10^{-5})	(5.77×10^{-6})	(2.74×10^{-5})	(4.07×10^{-4})	(1.17×10^{-5})
Election Slope Change	-7.01×10^{-8}	1.89×10^{-6}	-2.50×10^{-8}	9.26×10^{-8}	-2.73×10^{-8}	-1.62×10^{-8}	-6.07×10^{-9}	-1.35×10^{-6}	-3.60×10^{-8}
	(1.68×10^{-7})	(1.32×10^{-6})	(7.33×10^{-8})	(5.89×10^{-8})	(1.48×10^{-7})	(3.97×10^{-8})	(1.87×10^{-7})	(2.82×10^{-6})	(8.05×10^{-8})
AIC	-1.15×10^{4}	-8.92×10^{3}	-1.28×10^{4}	-1.29×10^{4}	-1.20×10^{4}	-1.35×10^{4}	-1.07×10^{4}	-7.61×10^{3}	-1.24×10^{4}
BIC	-1.15×10^{4}	-8.89×10^{3}	-1.28×10^{4}	-1.29×10^{4}	-1.20×10^{4}	-1.35×10^{4}	-1.07×10^{4}	-7.58×10^{3}	-1.24×10^{4}
Log Likelihood	5.78×10^{3}	4.47×10^{3}	6.42×10^{3}	6.46×10^{3}	6.03×10^{3}	6.75×10^{3}	5.38×10^{3}	3.81×10^{3}	6.23×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A14: Effect of Election on Daily Proportion Tweets Containing Hate Speeech or White Nationalist Rhetoric Including Quadratic Terms (Trump Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$4.37 \times 10^{-5*}$	$4.41 \times 10^{-4**}$	-9.00×10^{-8}	$6.63 \times 10^{-5***}$	1.97×10^{-7}	$1.30 \times 10^{-5**}$	$9.42 \times 10^{-5***}$	2.80×10^{-4}	$-1.89 \times 10^{-5*}$
	(2.04×10^{-5})	(1.60×10^{-4})	(8.62×10^{-6})	(7.10×10^{-6})	(1.69×10^{-5})	(4.60×10^{-6})	(2.18×10^{-5})	(3.37×10^{-4})	(8.45×10^{-6})
Pre-Election Trend	-4.30×10^{-8}	9.69×10^{-7}	$2.12 \times 10^{-7**}$	$-1.62 \times 10^{-7*}$	$5.04 \times 10^{-7***}$	$1.48 \times 10^{-7***}$	$5.78 \times 10^{-7**}$	-3.48×10^{-7}	$5.25 \times 10^{-7***}$
	(1.84×10^{-7})	(1.44×10^{-6})	(7.77×10^{-8})	(6.40×10^{-8})	(1.53×10^{-7})	(4.15×10^{-8})	(1.97×10^{-7})	(3.03×10^{-6})	(7.62×10^{-8})
Election Level Change	1.77×10^{-5}	1.70×10^{-4}	-5.19×10^{-6}	-8.16×10^{-6}	1.13×10^{-5}	-4.66×10^{-6}	4.68×10^{-5}	-3.30×10^{-5}	5.58×10^{-6}
	(3.62×10^{-5})	(2.73×10^{-4})	(1.53×10^{-5})	(1.28×10^{-5})	(2.94×10^{-5})	(8.31×10^{-6})	(4.01×10^{-5})	(5.90×10^{-4})	(1.54×10^{-5})
Election Slope Change	5.10×10^{-7}	5.85×10^{-6}	3.08×10^{-7}	-9.80×10^{-9}	8.56×10^{-7}	1.50×10^{-7}	3.92×10^{-7}	1.98×10^{-5}	4.45×10^{-7}
	(6.78×10^{-7})	(5.26×10^{-6})	(2.87×10^{-7})	(2.37×10^{-7})	(5.60×10^{-7})	(1.54×10^{-7})	(7.34×10^{-7})	(1.12×10^{-5})	(2.83×10^{-7})
Pre-Elelction Trend^2	-4.40×10^{-11}	-3.44×10^{-9}	$-3.89 \times 10^{-10**}$	1.31×10^{-10}	$-9.91 \times 10^{-10***}$	$-2.68 \times 10^{-10***}$	$-1.39 \times 10^{-9***}$	4.22×10^{-10}	$-9.69 \times 10^{-10***}$
	(3.47×10^{-10})	(2.71×10^{-9})	(1.47×10^{-10})	(1.21×10^{-10})	(2.88×10^{-10})	(7.80×10^{-11})	(3.72×10^{-10})	(5.72×10^{-9})	(1.44×10^{-10})
Election Slope Change^2	-2.47×10^{-9}	-6.32×10^{-9}	-2.08×10^{-10}	2.70×10^{-11}	-6.65×10^{-10}	1.41×10^{-10}	2.79×10^{-9}	$-9.68 \times 10^{-8*}$	1.04×10^{-9}
	(2.85×10^{-9})	(2.19×10^{-8})	(1.20×10^{-9})	(1.00×10^{-9})	(2.34×10^{-9})	(6.49×10^{-10})	(3.11×10^{-9})	(4.68×10^{-8})	(1.20×10^{-9})
AIC	-1.15×10^{4}	-8.85×10^{3}	-1.28×10^{4}	-1.28×10^{4}	-1.20×10^{4}	-1.34×10^{4}	-1.07×10^{4}	-7.54×10^{3}	-1.24×10^{4}
BIC	-1.14×10^{4}	-8.81×10^{3}	-1.27×10^{4}	-1.28×10^{4}	-1.19×10^{4}	-1.34×10^{4}	-1.06×10^{4}	-7.50×10^{3}	-1.24×10^{4}
Log Likelihood	5.74×10^{3}	4.43×10^{3}	6.38×10^{3}	6.42×10^{3}	5.99×10^{3}	6.71×10^{3}	5.35×10^{3}	3.78×10^{3}	6.21×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model including quadratic terms, demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A15: Aggregate Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Trump Data)

	Model 1	Model 2
Baseline	$1.19 \times 10^{-3***}$	$9.39 \times 10^{-4*}$
	(2.55×10^{-4})	(3.76×10^{-4})
Pre-Election Trend	-1.16×10^{-6}	1.84×10^{-6}
	(8.61×10^{-7})	(3.38×10^{-6})
Election Level Change	$8.16 \times 10^{-4\cdot}$	2.08×10^{-4}
	(4.55×10^{-4})	(6.52×10^{-4})
Election Slope Change	5.05×10^{-7}	$2.77 \times 10^{-5*}$
	(3.17×10^{-6})	(1.24×10^{-5})
Pre-Election Trend^2		-6.00×10^{-9}
		(6.00×10^{-9})
Election Slope Change^2		$-1.04 \times 10^{-7*}$
		(5.20×10^{-8})
AIC	-7.46×10^3	-7.39×10^3
BIC	-7.43×10^3	-7.35×10^3
Log Likelihood	3.73×10^3	3.70×10^3
Num. obs.	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion hate speech tweets in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A16: Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech (Trump Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$8.06 \times 10^{-5**}$	$1.11 \times 10^{-3***}$	$2.86 \times 10^{-5*}$	$1.01 \times 10^{-4***}$	$7.87 \times 10^{-5***}$	$3.30 \times 10^{-5***}$	$2.80 \times 10^{-4***}$	4.06×10^{-4}	$4.79 \times 10^{-5**}$
	(2.49×10^{-5})	(2.48×10^{-4})	(1.34×10^{-5})	(1.00×10^{-5})	(2.33×10^{-5})	(8.51×10^{-6})	(3.39×10^{-5})	(4.02×10^{-4})	(1.65×10^{-5})
Pre-Election Trend	-5.77×10^{-8}	-8.00×10^{-7}	8.50×10^{-8}	$-8.39 \times 10^{-8*}$	7.67×10^{-8}	$1.20 \times 10^{-7***}$	9.58×10^{-9}	4.92×10^{-7}	9.73×10^{-8}
	(8.40×10^{-8})	(8.36×10^{-7})	(4.51×10^{-8})	(3.39×10^{-8})	(7.85×10^{-8})	(2.88×10^{-8})	(1.15×10^{-7})	(1.36×10^{-6})	(5.58×10^{-8})
Election Level Change	3.85×10^{-5}	-1.39×10^{-4}	$-5.88 \times 10^{-5*}$	$-4.35 \times 10^{-5*}$	-6.95×10^{-5}	$-6.66 \times 10^{-5***}$	$-1.71 \times 10^{-4**}$	1.09×10^{-3}	$-9.19 \times 10^{-5**}$
	(4.50×10^{-5})	(4.37×10^{-4})	(2.41×10^{-5})	(1.82×10^{-5})	(4.15×10^{-5})	(1.54×10^{-5})	(6.20×10^{-5})	(7.21×10^{-4})	(2.99×10^{-5})
Election Slope Change	-1.69×10^{-7}	2.81×10^{-6}	-1.06×10^{-7}	8.76×10^{-8}	-1.29×10^{-7}	-1.26×10^{-7}	-2.44×10^{-7}	-2.70×10^{-6}	-1.22×10^{-7}
	(3.10×10^{-7})	(3.06×10^{-6})	(1.66×10^{-7})	(1.25×10^{-7})	(2.88×10^{-7})	(1.06×10^{-7})	(4.24×10^{-7})	(4.99×10^{-6})	(2.06×10^{-7})
AIC	-1.06×10^{4}	-7.76×10^{3}	-1.15×10^{4}	-1.19×10^{4}	-1.10×10^{4}	-1.21×10^{4}	-9.59×10^{3}	-6.73×10^{3}	-1.11×10^{4}
BIC	-1.05×10^{4}	-7.73×10^{3}	-1.15×10^{4}	-1.18×10^{4}	-1.09×10^{4}	-1.21×10^{4}	-9.57×10^{3}	-6.70×10^{3}	-1.11×10^{4}
Log Likelihood	5.28×10^{3}	3.89×10^{3}	5.77×10^{3}	5.93×10^{3}	5.49×10^{3}	6.05×10^{3}	4.80×10^{3}	3.37×10^{3}	5.57×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech or white nationalist language in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech and white nationalist tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A17: Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech Including Quadratic Terms (Trump Dataset)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	6.42×10^{-5}	5.23×10^{-4}	-8.62×10^{-6}	$9.13 \times 10^{-5***}$	-1.89×10^{-5}	2.94×10^{-6}	$1.10 \times 10^{-4*}$	4.00×10^{-4}	$-5.61 \times 10^{-5*}$
	(3.74×10^{-5})	(3.65×10^{-4})	(1.96×10^{-5})	(1.51×10^{-5})	(3.25×10^{-5})	(1.24×10^{-5})	(4.84×10^{-5})	(5.94×10^{-4})	(2.18×10^{-5})
Pre-Election Trend	1.35×10^{-7}	6.08×10^{-6}	$5.21 \times 10^{-7**}$	3.42×10^{-8}	$1.22 \times 10^{-6***}$	$4.72 \times 10^{-7***}$	$2.01 \times 10^{-6***}$	6.34×10^{-7}	$1.32 \times 10^{-6***}$
	(3.37×10^{-7})	(3.28×10^{-6})	(1.77×10^{-7})	(1.36×10^{-7})	(2.93×10^{-7})	(1.11×10^{-7})	(4.37×10^{-7})	(5.35×10^{-6})	(1.96×10^{-7})
Election Level Change	1.63×10^{-5}	2.07×10^{-4}	-3.16×10^{-5}	-3.19×10^{-5}	-1.21×10^{-6}	-3.97×10^{-5}	1.39×10^{-5}	-3.34×10^{-4}	8.00×10^{-6}
	(6.70×10^{-5})	(6.18×10^{-4})	(3.51×10^{-5})	(2.71×10^{-5})	(5.71×10^{-5})	(2.22×10^{-5})	(8.91×10^{-5})	(1.05×10^{-3})	(3.94×10^{-5})
Election Slope Change	1.06×10^{-6}	1.49×10^{-5}	5.53×10^{-7}	1.51×10^{-7}	1.64×10^{-6}	2.71×10^{-7}	1.23×10^{-6}	3.73×10^{-5}	1.06×10^{-6}
	(1.25×10^{-6})	(1.20×10^{-5})	(6.55×10^{-7})	(5.05×10^{-7})	(1.08×10^{-6})	(4.13×10^{-7})	(1.63×10^{-6})	(1.97×10^{-5})	(7.29×10^{-7})
Pre-Elelction Trend^2	-3.78×10^{-10}	$-1.34 \times 10^{-8*}$	$-8.52 \times 10^{-10*}$	-2.30×10^{-10}	$-2.23 \times 10^{-9***}$	$-6.87 \times 10^{-10**}$	$-3.90 \times 10^{-9***}$	-4.04×10^{-10}	$-2.38 \times 10^{-9***}$
	(6.37×10^{-10})	(6.18×10^{-9})	(3.34×10^{-10})	(2.57×10^{-10})	(5.53×10^{-10})	(2.11×10^{-10})	(8.26×10^{-10})	(1.01×10^{-8})	(3.71×10^{-10})
Election Slope Change^2	-4.29×10^{-9}	-9.58×10^{-9}	-1.40×10^{-10}	4.78×10^{-10}	-4.95×10^{-10}	4.89×10^{-10}	6.22×10^{-9}	$-1.80 \times 10^{-7*}$	2.54×10^{-9}
	(5.25×10^{-9})	(4.98×10^{-8})	(2.75×10^{-9})	(2.13×10^{-9})	(4.52×10^{-9})	(1.74×10^{-9})	(6.90×10^{-9})	(8.28×10^{-8})	(3.07×10^{-9})
AIC	-1.05×10^{4}	-7.69×10^{3}	-1.14×10^{4}	-1.18×10^{4}	-1.09×10^{4}	-1.20×10^{4}	-9.54×10^{3}	-6.66×10^{3}	-1.11×10^{4}
BIC	-1.04×10^{4}	-7.66×10^{3}	-1.14×10^{4}	-1.17×10^{4}	-1.09×10^{4}	-1.20×10^{4}	-9.50×10^{3}	-6.63×10^{3}	-1.11×10^{4}
Log Likelihood	5.25×10^{3}	3.85×10^{3}	5.73×10^{3}	5.89×10^{3}	5.46×10^{3}	6.01×10^{3}	4.78×10^{3}	3.34×10^{3}	5.55×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model including quadratic terms, demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech or white nationalist language in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech and white nationalist tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A18: Aggregate Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech (Trump Data)

	Model 1	Model 2
Baseline	$2.11 \times 10^{-3***}$	$1.16 \times 10^{-3\cdot}$
	(4.79×10^{-4})	(6.96×10^{-4})
Pre-Election Trend	-1.42×10^{-7}	$1.11 \times 10^{-5\cdot}$
	(1.62×10^{-6})	(6.27×10^{-6})
Election Level Change	5.57×10^{-4}	-1.65×10^{-4}
	(8.54×10^{-4})	(1.21×10^{-3})
Election Slope Change	-5.00×10^{-7}	$5.70 \times 10^{-5*}$
	(5.94×10^{-6})	(2.30×10^{-5})
Pre-Election Trend^2		$-2.20 \times 10^{-8\cdot}$
		(1.20×10^{-8})
Election Slope Change^2		$-1.88 \times 10^{-7\cdot}$
		(9.70×10^{-8})
AIC	-6.53×10^{3}	-6.46×10^{3}
BIC	-6.50×10^{3}	-6.43×10^{3}
Log Likelihood	3.27×10^{3}	3.24×10^{3}
Num. obs.	730	730

$***p < 0.001$, $**p < 0.01$, $*p < 0.05$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech in a dataset of over 600 million tweets mentioning Donald Trump collected using Twitter's streaming API between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A19: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Random Sample of Twitter Users)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$1.58 \times 10^{-5***}$	$1.31 \times 10^{-3***}$	1.93×10^{-6}	$8.30 \times 10^{-6***}$	$1.16 \times 10^{-5***}$	$1.90 \times 10^{-5***}$	$7.83 \times 10^{-5***}$	$5.33 \times 10^{-4***}$	$1.20 \times 10^{-6***}$
	(2.12×10^{-6})	(3.45×10^{-5})	(1.66×10^{-6})	(1.46×10^{-6})	(3.17×10^{-6})	(2.51×10^{-6})	(2.26×10^{-6})	(2.05×10^{-5})	(1.70×10^{-7})
Pre-Election Trend	$3.00 \times 10^{-8***}$	$3.20 \times 10^{-7**}$	0.00	1.00×10^{-8}	2.00×10^{-8}	-1.00×10^{-8}	-1.00×10^{-8}	$3.40 \times 10^{-7***}$	-1.00×10^{-9}
	(1.00×10^{-8})	(1.20×10^{-7})	(1.00×10^{-8})	(1.00×10^{-9})	(1.00×10^{-8})	(1.00×10^{-8})	(1.00×10^{-8})	(7.00×10^{-8})	(1.00×10^{-9})
Election Level Change	4.09×10^{-6}	$-1.94 \times 10^{-4**}$	2.27×10^{-6}	-1.99×10^{-6}	-7.44×10^{-6}	$1.82 \times 10^{-5***}$	-1.00×10^{-6}	-3.09×10^{-5}	$6.90 \times 10^{-7*}$
	(3.85×10^{-6})	(6.17×10^{-5})	(3.02×10^{-6})	(2.65×10^{-6})	(5.72×10^{-6})	(4.53×10^{-6})	(4.11×10^{-6})	(3.69×10^{-5})	(3.10×10^{-7})
Election Slope Change	$-6.00 \times 10^{-8*}$	$1.14 \times 10^{-6**}$	-1.00×10^{-8}	-1.00×10^{-8}	4.00×10^{-8}	$-6.00 \times 10^{-8*}$	$-8.00 \times 10^{-8**}$	-3.60×10^{-7}	1.00×10^{-9}
	(3.00×10^{-8})	(4.30×10^{-7})	(2.00×10^{-8})	(2.00×10^{-8})	(4.00×10^{-8})	(3.00×10^{-8})	(3.00×10^{-8})	(2.50×10^{-7})	(1.00×10^{-9})
AIC	-1.38×10^{4}	-1.02×10^{4}	-1.40×10^{4}	-1.43×10^{4}	-1.34×10^{4}	-1.39×10^{4}	-1.36×10^{4}	-1.09×10^{4}	-1.70×10^{4}
BIC	-1.37×10^{4}	-1.02×10^{4}	-1.40×10^{4}	-1.43×10^{4}	-1.34×10^{4}	-1.38×10^{4}	-1.36×10^{4}	-1.09×10^{4}	-1.69×10^{4}
Log Likelihood	6.89×10^{3}	5.12×10^{3}	7.02×10^{3}	7.16×10^{3}	6.72×10^{3}	6.94×10^{3}	6.82×10^{3}	5.47×10^{3}	8.49×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 400 million tweets produced by a random sample of 500,000 American Twitter Users between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A20: Effect of 2016 Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language Including Quadratic Terms (Random Sample of Twitter Users)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$1.68 \times 10^{-5***}$	$1.14 \times 10^{-3***}$	-8.07×10^{-7}	$8.92 \times 10^{-6***}$	1.41×10^{-6}	$1.88 \times 10^{-5***}$	$7.25 \times 10^{-5***}$	$4.84 \times 10^{-4***}$	$1.44 \times 10^{-6***}$
	(3.23×10^{-6})	(4.62×10^{-5})	(2.45×10^{-6})	(2.23×10^{-6})	(4.68×10^{-6})	(3.83×10^{-6})	(3.35×10^{-6})	(3.10×10^{-5})	(2.57×10^{-7})
Pre-Election Trend	1.87×10^{-8}	$2.22 \times 10^{-6***}$	3.48×10^{-8}	1.80×10^{-9}	$1.38 \times 10^{-7**}$	-5.30×10^{-9}	5.47×10^{-8}	$9.00 \times 10^{-7**}$	-3.40×10^{-9}
	(2.89×10^{-8})	(4.13×10^{-7})	(2.19×10^{-8})	(1.99×10^{-8})	(4.18×10^{-8})	(3.43×10^{-8})	(4.00×10^{-8})	(2.77×10^{-7})	(2.30×10^{-9})
Election Level Change	5.46×10^{-6}	1.41×10^{-4}	-4.49×10^{-6}	-2.77×10^{-6}	7.52×10^{-6}	$2.05 \times 10^{-5**}$	$1.24 \times 10^{-5*}$	6.76×10^{-5}	7.91×10^{-7}
	(5.81×10^{-6})	(8.15×10^{-5})	(4.42×10^{-6})	(4.01×10^{-6})	(8.33×10^{-6})	(6.78×10^{-6})	(6.05×10^{-6})	(5.43×10^{-5})	(4.69×10^{-7})
Election Slope Change	-1.31×10^{-7}	-2.14×10^{-6}	$2.73 \times 10^{-7***}$	-1.37×10^{-8}	1.08×10^{-8}	-1.18×10^{-7}	$-2.31 \times 10^{-7*}$	-1.28×10^{-6}	-8.90×10^{-9}
	(1.07×10^{-7})	(1.52×10^{-6})	(8.11×10^{-8})	(7.37×10^{-8})	(1.54×10^{-7})	(1.27×10^{-7})	(1.11×10^{-7})	(1.02×10^{-6})	(8.60×10^{-9})
Pre-Election Trend^2	0.00	$-3.70 \times 10^{-9***}$	-1.00×10^{-10}	1.00×10^{-11}	$-2.00 \times 10^{-10**}$	-1.00×10^{-11}	$-1.00 \times 10^{-10*}$	$-1.10 \times 10^{-9*}$	1.00×10^{-11}
	(1.00×10^{-10})	(8.00×10^{-10})	(1.00×10^{-11})	(0.00)	(1.00×10^{-10})	(1.00×10^{-10})	(1.00×10^{-10})	(5.00×10^{-10})	(0.00)
Election Slope Change^2	3.00×10^{-10}	$2.70 \times 10^{-8***}$	$-1.10 \times 10^{-9**}$	-1.00×10^{-11}	9.00×10^{-10}	3.00×10^{-10}	$1.10 \times 10^{-9*}$	7.80×10^{-9}	1.00×10^{-11}
	(5.00×10^{-10})	(6.40×10^{-9})	(3.00×10^{-10})	(3.00×10^{-10})	(6.00×10^{-10})	(5.00×10^{-10})	(5.00×10^{-10})	(4.30×10^{-9})	(1.00×10^{-11})
AIC	-1.37×10^{4}	-1.02×10^{4}	-1.40×10^{4}	-1.42×10^{4}	-1.33×10^{4}	-1.38×10^{4}	-1.36×10^{4}	-1.08×10^{4}	-1.68×10^{4}
BIC	-1.36×10^{4}	-1.01×10^{4}	-1.39×10^{4}	-1.42×10^{4}	-1.33×10^{4}	-1.37×10^{4}	-1.35×10^{4}	-1.08×10^{4}	-1.68×10^{4}
Log Likelihood	6.85×10^{3}	5.10×10^{3}	6.99×10^{3}	7.12×10^{3}	6.68×10^{3}	6.89×10^{3}	6.78×10^{3}	5.43×10^{3}	8.43×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model including quadratic terms, demonstrating the effect of Trump's election on the daily proportion hate speech or white nationalist language tweets in a dataset of over 400 million tweets produced by a random sample of 500,000 American Twitter Users between June 17, 2015 and June 15, 2017. Hate speech or white nationalist language tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A21: Aggregate Effect of Election on Daily Proportion of Tweets Containing Hate Speech or White Nationalist Language (Random Sample Data)

	Model 1	Model 2
Baseline	$1.98 \times 10^{-3***}$	$1.75 \times 10^{-3***}$
	(4.52×10^{-5})	(6.00×10^{-5})
Pre-Election Trend	$6.96 \times 10^{-7***}$	$3.34 \times 10^{-6***}$
	(1.53×10^{-7})	(5.41×10^{-7})
Election Level Change	$-2.01 \times 10^{-4*}$	$2.43 \times 10^{-4*}$
	(8.09×10^{-5})	(1.06×10^{-4})
Election Slope Change	5.61×10^{-7}	$-3.53 \times 10^{-6\cdot}$
	(5.62×10^{-7})	(1.99×10^{-6})
Pre-Election Trend^2		$-5.00 \times 10^{-9***}$
		(1.00×10^{-9})
Election Slope Change^2		$3.60 \times 10^{-8***}$
		(8.00×10^{-9})
AIC	-9.87×10^{3}	-9.83×10^{3}
BIC	-9.85×10^{3}	-9.79×10^{3}
Log Likelihood	4.94×10^{3}	4.92×10^{3}
Num. obs.	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion hate speech tweets in a dataset of over 400 million tweets produced by a random sample of 500,000 American Twitter Users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A22: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech (Random Sample of Twitter Users)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$8.21 \times 10^{-5***}$	$6.28 \times 10^{-3***}$	$9.79 \times 10^{-6**}$	$4.12 \times 10^{-5***}$	$5.84 \times 10^{-5**}$	$7.79 \times 10^{-5***}$	$4.06 \times 10^{-4***}$	$2.79 \times 10^{-3***}$	$6.06 \times 10^{-6***}$
	(1.37×10^{-5})	(3.52×10^{-4})	(3.62×10^{-6})	(1.19×10^{-5})	(1.88×10^{-5})	(1.45×10^{-5})	(1.95×10^{-5})	(1.71×10^{-4})	(9.90×10^{-7})
Pre-Election Trend	$2.80 \times 10^{-7***}$	$4.00 \times 10^{-6***}$	$3.00 \times 10^{-8*}$	$1.10 \times 10^{-7**}$	$1.70 \times 10^{-7**}$	7.00×10^{-8}	$1.80 \times 10^{-7**}$	$3.12 \times 10^{-6***}$	1.00×10^{-9}
	(5.00×10^{-8})	(1.17×10^{-6})	(1.00×10^{-8})	(4.00×10^{-8})	(6.00×10^{-8})	(5.00×10^{-8})	(7.00×10^{-8})	(5.70×10^{-7})	(1.00×10^{-9})
Election Level Change	-1.40×10^{-5}	-8.49×10^{-4}	-9.41×10^{-6}	-2.73×10^{-5}	-6.06×10^{-5}	$8.54 \times 10^{-5**}$	$-1.02 \times 10^{-4**}$	-3.90×10^{-4}	2.12×10^{-6}
	(2.49×10^{-5})	(5.87×10^{-4})	(6.46×10^{-6})	(2.15×10^{-5})	(3.39×10^{-5})	(2.60×10^{-5})	(3.48×10^{-5})	(2.98×10^{-4})	(1.80×10^{-6})
Election Slope Change	$-4.60 \times 10^{-7**}$	-2.26×10^{-6}	-7.00×10^{-8}	-1.60×10^{-7}	2.00×10^{-7}	$-5.10 \times 10^{-7**}$	$-6.70 \times 10^{-7**}$	$-5.35 \times 10^{-6*}$	1.00×10^{-8}
	(1.70×10^{-7})	(4.25×10^{-6})	(4.00×10^{-8})	(1.50×10^{-7})	(2.30×10^{-7})	(1.80×10^{-7})	(2.40×10^{-7})	(2.09×10^{-6})	(1.00×10^{-8})
AIC	-1.11×10^{4}	-7.70×10^{3}	-1.36×10^{4}	-1.14×10^{4}	-1.10×10^{4}	-1.14×10^{4}	-1.11×10^{4}	-8.31×10^{3}	-1.44×10^{4}
BIC	-1.11×10^{4}	-7.68×10^{3}	-1.36×10^{4}	-1.13×10^{4}	-1.09×10^{4}	-1.14×10^{4}	-1.10×10^{4}	-8.28×10^{3}	-1.44×10^{4}
Log Likelihood	5.57×10^{3}	3.86×10^{3}	6.80×10^{3}	5.69×10^{3}	5.48×10^{3}	5.72×10^{3}	5.53×10^{3}	4.16×10^{3}	7.22×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$***p < 0.001, **p < 0.01, *p < 0.05$

This regression table shows results of an AR1 Interrupted Time Series Analysis (ITSA) model demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech or white nationalist language in a dataset of over 400 million tweets produced by a random sample of 500,000 American Twitter users between June 17, 2015 and June 15, 2017. Hate speech and white nationalist tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A23: Effect of 2016 Election on Daily Proportion of Unique Users Tweeting Hate Speech (Random Sample of Twitter Users)

	Anti-Asian	**Anti-Black**	**Anti-Immigrant**	**Anti-Latino**	**Anti-Muslim**	**Anti-Semitic**	**Homophobic**	**Misogynistic**	**White Nationalist**
Baseline	$5.22 \times 10^{-5*}$	$3.69 \times 10^{-3***}$	-7.13×10^{-6}	2.83×10^{-5}	-1.44×10^{-5}	$5.59 \times 10^{-5*}$	$2.66 \times 10^{-4***}$	$1.60 \times 10^{-3***}$	$5.68 \times 10^{-6***}$
	(2.07×10^{-5})	(3.41×10^{-4})	(5.03×10^{-6})	(1.80×10^{-5})	(2.73×10^{-5})	(2.20×10^{-5})	(2.42×10^{-5})	(2.12×10^{-4})	(1.50×10^{-6})
Pre-Election Trend	$6.20 \times 10^{-7***}$	$3.31 \times 10^{-5***}$	$2.20 \times 10^{-7***}$	2.50×10^{-7}	$1.01 \times 10^{-6***}$	3.20×10^{-7}	$1.79 \times 10^{-6***}$	$1.66 \times 10^{-5***}$	1.00×10^{-8}
	(1.90×10^{-7})	(3.04×10^{-6})	(4.00×10^{-8})	(1.60×10^{-7})	(2.40×10^{-7})	(2.00×10^{-7})	(2.20×10^{-7})	(1.89×10^{-6})	(1.00×10^{-8})
Election Level Change	4.28×10^{-5}	$1.88 \times 10^{-3**}$	8.80×10^{-7}	-1.25×10^{-5}	4.49×10^{-5}	$1.10 \times 10^{-4**}$	6.64×10^{-5}	$1.20 \times 10^{-3***}$	3.38×10^{-6}
	(3.72×10^{-5})	(5.77×10^{-4})	(8.80×10^{-6})	(3.23×10^{-5})	(4.83×10^{-5})	(3.87×10^{-5})	(4.29×10^{-5})	(3.64×10^{-4})	(2.74×10^{-6})
Election Slope Change	-8.90×10^{-7}	-1.26×10^{-5}	2.80×10^{-7}	-8.00×10^{-8}	-2.00×10^{-8}	-3.80×10^{-7}	-2.40×10^{-7}	-1.30×10^{-5}	-1.00×10^{-8}
	(6.90×10^{-7})	(1.11×10^{-5})	(1.70×10^{-7})	(6.00×10^{-7})	(9.00×10^{-7})	(7.30×10^{-7})	(8.00×10^{-7})	(6.92×10^{-6})	(5.00×10^{-8})
Pre-Elelction Trend2	-1.00×10^{-9}	$-5.00 \times 10^{-8***}$	$-1.00 \times 10^{-9***}$	-1.00×10^{-9}	$-1.00 \times 10^{-9***}$	-1.00×10^{-9}	$-1.00 \times 10^{-9***}$	$-3.00 \times 10^{-8***}$	-1.00×10^{-9}
	(1.00×10^{-9})	(1.00×10^{-8})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})
Election Slope Change2	1.00×10^{-9}	$2.50 \times 10^{-7***}$	-1.00×10^{-9}	1.10×10^{-9}	1.00×10^{-8}	1.00×10^{-9}	$1.00 \times 10^{-8*}$	$1.20 \times 10^{-7***}$	1.00×10^{-9}
	(1.00×10^{-9})	(5.00×10^{-8})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(1.00×10^{-9})	(3.00×10^{-8})	(1.00×10^{-9})
AIC	-1.11×10^{4}	-7.71×10^{3}	-1.35×10^{4}	-1.13×10^{4}	-1.09×10^{4}	-1.13×10^{4}	-1.10×10^{4}	-8.29×10^{3}	-1.43×10^{4}
BIC	-1.10×10^{4}	-7.68×10^{3}	-1.35×10^{4}	-1.13×10^{4}	-1.08×10^{4}	-1.13×10^{4}	-1.10×10^{4}	-8.26×10^{3}	-1.43×10^{4}
Log Likelihood	5.53×10^{3}	3.86×10^{3}	6.76×10^{3}	5.65×10^{3}	5.45×10^{3}	5.67×10^{3}	5.52×10^{3}	4.15×10^{3}	7.17×10^{3}
Num. obs.	730	730	730	730	730	730	730	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A24: Aggregate Effect of Election on Daily Proportion of Unique Users Tweeting Hate Speech
(Random Sample Data)

	Model 1	Model 2
Baseline	0.011326686***	0.005680348***
	(0.001612345)	(0.000591132)
Pre-Election Trend	-0.000001764	0.000055408***
	(0.000005145)	(0.000005309)
Election Level Change	0.009589787***	0.004686698***
	(0.001505211)	(0.000955332)
Election Slope Change	-0.000047131**	-0.000040721*
	(0.000017605)	(0.000019121)
Pre-Election Trend^2		-0.000000091***
		(0.000000010)
Election Slope Change^2		0.000000461***
		(0.000000080)
AIC	-7211.280866325	-7218.615716584
BIC	-7183.755566236	-7181.937385445
Log Likelihood	3611.640433163	3617.307858292
Num. obs.	730	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

This regression table shows results of AR1 Interrupted Time Series Analysis (ITSA) models demonstrating the effect of Trump's election on the daily proportion of unique users tweeting hate speech in a dataset of over 400 million tweets produced by a random sample of 500,000 American Twitter Users between June 17, 2015 and June 15, 2017. Hate speech tweets were identified both using dictionaries of slurs and a Naive Bayes Classifier trained to remove false positives from the data.

Table A25: Descriptions for popular Twitter accounts used for validation

Twitter account	**Description**
@ACLU	The American Civil Liberties Union
@ADL_National	Anti-Defamation League
@AmRenaissance	Race-realist, white advocacy magazine
@BarackObama	44th President of the United States
@BernieSanders	American politician, Senator
@Blklivesmatter	Black Lives Matter organization
@BuzzFeed	Internet media
@CCriadoPerez	Feminist activist
@CNET	American media website on technology
@Cernovich	Alt-right social media personality
@Eminem	Musician
@FCBarcelona	Spanish football club
@FeminismDaiIy	Pro-feminist Twitter account
@Gavin_McInnes	Far-right commentator & comedian
@HillaryClinton	Politician, Democratic party
@Impeach_D_Trump	Anti-Trump Twitter account
@JamilahLemieux	African-American columnist
@JuddLegum	Liberal journalist, editor of ThinkProgress
@LGBT_news	Pro-LGBT Twitter account
@Lauren_Southern	Canadian right-wing activist
@MGTOW	Men Going Their Own Way account
@NewAltRight	Alt-right blog
@PamelaGeller	Anti-Islamic political commentator
@PrisonPlanet	Paul Joseph Watson, editor-at-large at Infowars.com
@RichardBSpencer	American white nationalist
@TheMadDimension	Organizer of Unite the Right rally
@Yankees	American baseball team
@feminismvibes	Pro-feminist Twitter account
@jartaylor	Editor of American Renaissance pro-white magazine
@latimesmovies	LA Times Movie News
@orensegal	Director of Center on Extremism (ADL)
@ramzpaul	Alt-right YouTube personality
@realDonaldTrump	45th President of the United States
@splcenter	The Southern Poverty Law Center

Table A26: Optimal parameters of the model

Parameter	**Values**
Preprocessing options	
Stemming	No
Removing stop-words	Yes
Removing punctuation	Yes
Lowercasing	Yes
fastText paramters	
Epochs	15
Learning rate	0.01
Dimension	100
n-grams	2
Minimal count	5
Buckets	2000000

This table demonstrates the optimal values of hyperparameters of the model. The best model correctly predicts group for 45.8% of the comments in the test dataset.

Table A27: Subreddit-based Classification of popular Twitter accounts

Category	**Accounts**
Anti-Trump	ACLU, DC_Resister_Bee, JuddLegum, NancyPelosi, TheDemocrats, splcenter
Religion	ADL_National, orensegal
Altright	AmRenaissance, Mathiasian, NewAltRight, RichardBSpencer, jartaylor, ramzpaul
Black	BLMNYC, Blklivesmatter, Eminem
Liberal	BarackObama, BernieSanders, HillaryClinton
/r/The_Donald	BreitbartNews, Cernovich, FoxNews, GOP, Gavin_McInnes, JackPosobiec, LauraLoomer, Lauren_Southern, PrisonPlanet, SenJohnMcCain, TheMadDimension, lucianwintrich, realDonaldTrump
Misc	BuzzFeed, CNET, NYUDataScience, latimesmovies, nytimes
Sport	FCBarcelona, Yankees
Feminist	FeminismDaiIy,WeNeedFeminlsm, feminismvibes
LGBT	LGBT_news
Anti-feminist	MGTOW
/r/uncensorednews	PamelaGeller
Conservative	marcorubio

Table A28: Effect of Election on Alt-Right Reddit Similarity (Clinton Data)

	Model 1	Model 2
Baseline	0.02921019***	0.02985143***
	(0.00096843)	(0.00144286)
Pre-Election Trend	−0.00000520	−0.00001264
	(0.00000327)	(0.00001299)
Election Level Change	−0.00196347	−0.00446807$^{\cdot}$
	(0.00173207)	(0.00252757)
Election Slope Change	0.00000093	0.00004704
	(0.00001201)	(0.00004765)
Pre-Election Trend^2		0.00000001
		(0.00000002)
Election Slope Change^2		−0.00000026
		(0.00000020)
AIC	-5398.19670180	-5333.83733947
BIC	-5370.68795345	-5297.18113836
Log Likelihood	2705.09835090	2674.91866974
Num. obs.	728	728

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A29: Effect of Election on Alt-Right Reddit Similarity (Trump Data)

	Model 1	Model 2
Baseline	0.03083694^{***}	0.03012187^{***}
	(0.00083470)	(0.00125485)
Pre-Election Trend	-0.00001285^{***}	-0.00000468
	(0.00000281)	(0.00001127)
Election Level Change	0.00132778	0.00414511^{*}
	(0.00144906)	(0.00206040)
Election Slope Change	-0.00001395	$-0.00007085^{\cdot}$
	(0.00001025)	(0.00004073)
Pre-Election Trend^2		-0.00000002
		(0.00000002)
Election Slope Change^2		$0.00000031^{\cdot}$
		(0.00000017)
AIC	-6090.10560358	-6027.00732968
BIC	-6062.59685522	-5990.35112856
Log Likelihood	3051.05280179	3021.50366484
Num. obs.	728	728

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A30: Effect of Election on Alt-Right Reddit Similarity (Random Sample Data)

	Model 1	Model 2
Baseline	0.03492992^{***}	0.03355044^{***}
	(0.00039590)	(0.00054855)
Pre-Election Trend	-0.00001737^{***}	-0.00000116
	(0.00000133)	(0.00000493)
Election Level Change	0.00198372^{*}	0.00298350^{**}
	(0.00078714)	(0.00100440)
Election Slope Change	0.00000516	0.00002559
	(0.00001029)	(0.00003701)
Pre-Election Trend ^{2}		-0.00000003^{***}
		(0.00000001)
Election Slope Change ^{2}		0.00000002
		(0.00000028)
AIC	-6496.41447680	-6439.97483664
BIC	-6469.74015788	-6404.43451531
Log Likelihood	3254.20723840	3227.98741832
Num. obs.	634	634

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$