From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment *

Karsten Müller† and Carlo Schwarz‡

October 31, 2019

Abstract

We study whether social media can activate hatred of minorities, with a focus on Donald Trump's political rise. We show that the increase in anti-Muslim sentiment in the US since the start of Trump's presidential campaign has been concentrated in counties with high Twitter usage. To establish causality, we develop an identification strategy based on Twitter's early adopters at the South by Southwest (SXSW) festival, which marked a turning point in the site's popularity. Instrumenting with the locations of SXSW followers in March 2007, while controlling for the locations of SXSW followers who joined in previous months, we find that a one standard deviation increase in Twitter usage is associated with a 38% larger increase in anti-Muslim hate crimes since Trump's campaign start. We also show that Trump's tweets about Islam-related topics are highly correlated with anti-Muslim hate crimes after the start of his presidential campaign, but not before. These correlations persist in an instrumental variable framework exploiting that Trump is more likely to tweet about Muslims on days when he plays golf. Trump's tweets also predict more anti-Muslim Twitter activity of his followers and higher cable news attention paid to Muslims, particularly on Fox News.

*A previous version of this paper was circulated under the title "Making America Hate Again? Twitter and Hate Crime Under Trump." We are grateful to Roland Bénabou, Dan Bernhardt, Leonardo Bursztyn, Rafael di Tella, Mirko Draca, Ruben Durante, Didi Egerton-Warburton, Ruben Enikolopov, James Fenske, Thomas Fujiwara, Scott Gehlbach, Matthew Gentzkow, Andy Guess, Atif Mian, Jonathan Nagler, Maria Petrova, Joshua Tucker, Alessandra Voena, Hans-Joachim Voth, Fabian Waldinger, and seminar participants at the University of Warwick, Princeton University, New York University, the RES Conference 2019, the Young Economist Symposium 2019, the EMCON Conference 2019, the Political Economy Workshop Barcelona 2019, the Galatina Summer Meetings 2019, and the EEA Conference 2019 for their helpful suggestions. Christian Kontz provided excellent research assistance. Schwarz was supported by a Doctoral Scholarship from the Leverhulme Trust as part of the *Bridges* program.

†Princeton University, Julis-Rabinowitz Center for Public Policy and Finance, karstenm@princeton.edu

‡Department of Economics, University of Warwick, www.carloschwarz.eu, c.r.schwarz@warwick.ac.uk

1 Introduction

In this paper, we study whether social media platforms can affect anti-minority sentiments online and offline. We investigate this question in the context of a particularly notable case study: the political rise of Donald Trump. Trump has been widely criticized for his inflammatory rhetoric on Twitter and is frequently cited as an example of how social media can increase anti-minority sentiments (New York Times, 2017). Minnesota congresswoman Ilhan Omar, for example, has linked tweets by Trump targeting her Muslim faith to "an increase in direct threats on my life - many directly referring or replying to the president's video" (BBC, 2019).

We interpret Trump's presidential campaign as a shock to the salience of anti-Muslim views, particularly for those exposed to his rhetoric on social media. This interpretation is in line with experimental evidence that Trump's popularity on the campaign trail and subsequent election win increased people's willingness to publicly express xenophobic views (Bursztyn et al., 2017). Building on this insight, we ask if social media may play a role in propagating of anti-Muslim sentiment and real-life violence.

We start by documenting that the frequency of anti-Muslim hate crimes has doubled since Donald Trump's presidential campaign compared to the presidencies of Barack Obama and George W. Bush. This is particularly striking because Bush's term included a temporary ten-fold increase in such crimes following the 9/11 terror attacks, the largest spike since the beginning of the FBI records in 1990 (Gould & Klor, 2016; Panagopoulos, 2006; Hanes & Machin, 2014). It is also consistent with evidence that the Muslim community has been particularly affected by Trump's political rise (e.g. Hobbs & Lajevardi, 2019).

We investigate the potential role of social media in enabling such hate crimes using a difference-in-differences approach. We find that the increase in hate crimes targeting Muslims predominantly originates in counties with high Twitter usage. We also observe disproportionate increases in tweets containing the hashtags #BanIslam and #StopIslam in these counties. These regressions, however, may not isolate a pure "social media effect" because counties with many Twitter users likely also differ in many unobservable dimensions. This may bias our estimates upwards or downwards, depending on how individuals select into social media usage. For example, areas where many people use relatively new technologies such as Twitter may react less because they are more liberal and tolerant, which could bias our estimates downwards. On the other hand, such areas may have a larger share of minority groups and thus more potential targets for perpetrators of hate crimes.

To overcome these concerns, we construct an instrument for county-level Twitter usage in the United States based on the home towns of the platform's early adopters at the South by Southwest (SXSW) Festival in March 2007.[1] SXSW is widely regarded as the tipping point for Twitter's popularity and an important early catalyst for the site's success. One indication of SXSW's importance in explaining Twitter's trajectory is that the number of daily tweets *tripled* during the festival. We also find that tweets about SXSW are a clear outlier in 2007 compared to those about other, considerably more popular festivals, such as Burning Man, Coachella or Lollapalooza. We show that activity on Twitter grew rapidly in the weeks following SXSW 2007, and disproportionately so in the home counties of SXSW followers who signed up in March 2007.

In line with the literature on path dependence in technology adoption (e.g. Arthur, 1989, 1994; Liebowitz & Margolis, 1999; Arrow, 2000), this early expansion left its imprint on the geographical distribution of social media usage in the United States. The locations of Twitter's early adopters at SXSW are a strong predictor of county-level Twitter usage today, even after controlling for the locations of SXSW followers that had already signed up prior to the festival. This result is also robust to using alternative control sets, e.g. using the locations of Twitter users mentioning other major festivals in 2007 or those tweeting about SXSW before the 2007 event. Similar to the strategy of Enikolopov et al. (2016), the identifying assumption is that differences in the locations of SXSW followers in March 2007 relative to earlier months are not related to unobserved county characteristics that explain the rise in anti-Muslim sentiment with the 2016 presidential campaign. Because Twitter was largely unknown before SXSW, and these counties do not systematically differ in many observable characteristics, we believe this assumption is credible.

Instrumenting for Twitter usage with SXSW followers in March 2007, we confirm that measures of anti-Muslim sentiments disproportionately increased in areas with higher social media usage. We find that a one standard deviation higher exposure to social media is associated with a 38% larger increase in hate crimes between 2010 and 2017. This increase in hate crimes against Muslims is entirely accounted for by assaults. Exploiting heterogeneity across counties, we further show that most of this effect is driven by areas with higher pre-existing anti-minority bias. These findings suggest that social media platforms may have played a role in the recent spread of anti-Muslim sentiment in the United States by reinforcing existing tensions.

We also find a similar but slightly weaker pattern for hate crimes targeting Hispanics, the second minority group often targeted by Trump. While data from the FBI suggest that

[1] SXSW is an annual event, held since 1987, that comprises a number of festivals, conferences, trade shows, and exhibitions. In 2019, more than 230,000 people attended the festivals, where almost 2,000 acts from all over the world performed. More than 70,000 people attended the SXSW conference, which featured almost 4,800 speakers. Around 30,000 people attended SXSW Interactive, which focuses on emerging technology. For simplicity, we refer to the event as "SXSW festival" or similar short forms throughout the paper.

the frequency of these incidents has been largely unchanged, our results point to a potential role of social media in contributing to a geographical reallocation of these crimes.

To determine if Trump's tweets contributed to the increase of anti-Muslim sentiment on Twitter, we analyze Trump's Twitter feed. We find a strong time series correlation between Trump's tweets on Islam-related topics and the number of anti-Muslim hate crimes after the start of his presidential campaign, even after controlling for general attention paid to topics associated with Muslims. There is no correlation between Trump's tweets and hate crimes with other motives (e.g. racial hate crime), which suggests that we are not merely capturing waves of general anti-minority sentiment. We also find no such link for the period before the time of Trump's presidential campaign.

To establish causality, we leverage Trump's well-documented golf habit. This analysis is motivated by the fact that many commentators have argued that golfing shifts Trumps state of mind. In 2017 alone, Trump played golf on more than 90 days. In the data, we find a clear pattern: Trump's golf days coincide strongly with changes in the content, but not the number of his tweets. In particular, Trump is more likely to send messages aimed at Muslims and the media on his golf days, and fewer about policy, a fact we exploit in an instrumental variable framework. One intuitive explanation of this finding is that day-to-day politics may be less salient to the President when outside of Washington, DC. Additionally, there is anecdotal evidence that Trump may be influenced by his social media director Dan Scavino – former manager of Trump National Golf Club Westchester and Trump's former caddie – who has been linked to particularly inflammatory tweets (New York Times, 2018).

Using golf days as an instrument, we find evidence consistent with the idea that Trump's tweets about Muslims "trigger" waves of anti-Muslim sentiment. In particular, we find that his instrumented tweets not only continue to predict the frequency of hate crimes, but also measures of media attention paid to Muslim-related topics. Using transcript data on the reporting of the major cable news networks Fox News, CNN, and MSNBC, we show a time series correlation between Trump's golf-induced tweets and mentions of Muslims. This link seems to be largely driven by Fox News, which tends to support rather than oppose Trump's rhetoric. Analyzing over 100 million tweets, we also find that Trump's anti-Muslim tweets are widely shared by his followers, who further produce their own anti-Muslim content.

Additionally, we investigate whether the transmission effects of Donald Trump's tweets are stronger in counties with more Twitter users in a panel regression setting. Interacting county-level Twitter usage and Trump's Twitter activity, we document that the spike in anti-Muslim hate crime in the days after Donald Trump's tweets is driven by counties with higher Twitter penetration. These findings also persist when we estimate regressions in reduced form and two-stage least squares using our SXSW instrumental variable strategy.

Taken together, our evidence is consistent with the interpretation that, with the start of Donald Trump's presidential campaign, social media may have come to play a role in the increase of anti-Muslim sentiments in the United States. The existing literature broadly suggests three possible mechanisms to explain our findings: coordination capabilities, persuasion, and changes in social norms. We discuss how our findings line up with these three mechanism at the end of the paper. While all are likely at play, some of our results suggest that social media may influence the perception of which beliefs about minorities are socially acceptable. In other words, social media could have enabled changes in social norms for people at the fringes of the political spectrum. Because Twitter users are predominately male and more ideologically extreme than the general population (Barberá & Rivero, 2015), this may explain how social media can contribute to an increase in hate crimes.[2]

Our paper contributes to the literature on the relationship between media consumption and violence. Yanagizawa-Drott (2014), Adena et al. (2015), and DellaVigna et al. (2014) find that traditional media can contribute to ethnic hatred and violence. Other research has linked media such as television (Card & Dahl, 2011) and movies (Dahl & DellaVigna, 2009) to short-lived spikes (or decreases) in violence. Bhuller et al. (2013) document increases in sex crime associated with the roll-out of broadband internet in Norway; Chan et al. (2016) find a correlation between broadband availability and hate crimes in the US. Our findings speak to the role of social media in the spread of violence against minority groups.

We most directly contribute to a growing literature on the influence of social media on real life outcomes. Enikolopov et al. (2016) show that social media can increase participation in protests in Russia by reducing coordination costs. Petrova et al. (2017) study whether adopting Twitter helps politicians attract donations. In previous work, we found evidence that social media affects the propagation of anti-refugee incidents in Germany, using Facebook and internet disruptions as a source of short-lived exogenous variation (Müller & Schwarz, 2018). Here, we study the medium-term effects of social media and highlight a potential social norms channel, based on the particularly salient case study of Trump's presidency.

A separate related literature studies political polarization. While there is evidence that polarization has increased over the past decades (Fiorina & Abrams, 2008; Gentzkow, 2016; Draca & Schwarz, 2018), existing studies have found no or even a negative correlation with social media use (Boxell et al., 2017; Barberá, 2014).[3] One interpretation of our findings is

[2]These findings are also consistent with studies on the demographics of social media consumption. Guess et al. (2018) and Guess (2018), for example, show that consumption of fake news articles and ideologically extreme content is driven by relatively few people, which might overlap with the few potential perpetrators of hate crimes.

[3]A separate literature has analyzed the effects of the media on elections and other political outcomes. See, among others, the work by Adena et al. (2015), DellaVigna et al. (2014), Stephens-Davidowitz (2014), Gavazza et al. (2015), Gentzkow (2016), and Martin & Yurukoglu (2017).

that social media may not necessarily affect *average* outcomes, but rather enable those with extreme viewpoints to find sources of social legitimacy. A widely shared discriminatory tweet by the President, for example, could signal to potential perpetrators of hate crimes that their actions are more widely accepted than they really are.

In Section 2, we introduce the data sources and present descriptive evidence on hate crimes since 1990. In Section 3, we discuss our empirical strategy and introduce our instrument for Twitter usage based on the SXSW festival. Section 4 presents the main empirical results. In Section 5 we discuss evidence for the link between Trump's tweets and anti-Muslim sentiment. In Section 6 we show that the relationship between Trump's tweets and anti-Muslim hate crime is driven by counties with high Twitter usage. Section 7 discusses plausible mechanisms behind our results and potential reporting biases. Section 8 concludes.

2 Data and Background

We create two datasets for our analysis. First, we build a county-level dataset for the US containing information on hate crimes, Twitter usage, and numerous other variables. Second, we construct a daily time series dataset that combines Trump's daily Twitter activity, the number of total hate crime incidents in the US, data on TV news coverage, and time series control variables. The key sources we draw on are (1) hate crime data reported by the FBI's Uniform Crime Reporting (UCR) program; (2) a county-level measure of Twitter usage based on 475 million tweets collected by Kinder-Kurlanda et al. (2017); (3) hand-collected county-level data on the locations of early adopters of Twitter in 2006 and 2007; and (4) information on Trump's golf activity from his inauguration in early 2017 until the end of that year. We describe these and all other data sources in more detail in the following subsections. Table A.10 and Table A.11 in the online appendix present the full descriptive statistics.

2.1 FBI Hate Crime Data

The data on hate crime in the US come from the FBI and are available for the years 1990 until 2017.[4] The data set contains all hate crimes in the US that are reported to the FBI as part of the Uniform Crime Reporting (UCR) program. The FBI defines a hate crime as:

> "[...] criminal offenses that are motivated, in whole or in part, by an offenders bias against a race, religion, disability, sexual orientation, ethnicity, gender, or gender identity." (FBI, 2015, p. 4)

[4]Note that data for the year 2018 will only become available in November 2019.

To classify hate crimes, the FBI uses a two-tier decision making process. First, the law enforcement officer recording an incident has to decide whether it might constitute a hate crime. Second, the potential hate crime cases are forwarded to and evaluated by officers with special training in hate crime matters. The FBI (2015) states (p. 35): “For an incident to be reported as a hate crime, sufficient objective facts must be present to lead a reasonable and prudent person to conclude that the offenders actions were motivated, in whole or in part, by bias.” For more information on the FBI classification procedure see appendix A.1.

Because considerable evidence needs to be available for an offense to be classified as a hate crime, the numbers reported by the FBI have been criticized as underestimates (ProPublica, 2017; NBC News, 2017).[5] Nonetheless, the FBI data constitute the most complete record of hate crimes committed in the United States for which incident details are available. Among others, they include information on the exact date of the crime, the type of crime (e.g. vandalism, theft, assault), the number of victims, and the number of perpetrators. The data further make it possible to assign hate crimes to counties using the county location of the more than 32,000 original reporting agencies based on their Originating Agency Identifier (ORI).[6] Figure 2a plots the geographic distribution of hate crimes across the mainland USA.[7] The counties in grey never report any hate crime to the FBI.

The FBI differentiates hate crimes by motivating bias (e.g. anti-Muslim). Overall, they report 34 bias motivations for the broad categories race, religion, sexual orientation, disability, and gender/gender identity. We report all codes for the motivating bias in Table A.4. We use this classification to identify hate crimes against Muslims. The other categories used in the paper are defined according to the codes listed in Table A.3.

Presidents and Trends in Hate Crimes To motivate our analysis, we begin by investigating how the number of hate crime incidents has evolved over time. In particular, we test for changes in anti-Muslim hate crimes since the commencement of Trump's presidential run. Panel A of Figure 1 plots the average number of weekly anti-Muslim hate crimes for

[5] Note that time-invariant reporting bias across counties is unlikely to drive our results. First, the US-wide trend of hate crimes reported to the FBI is likely to be highly correlated with the “true” hate crimes trend. Second, we accommodate potential geographical reporting differences in our cross-sectional tests by estimating our model in first-differences. In further robustness checks we restrict the sample to counties where at least one hate crime is reported. We discuss the extent to which changes in reporting over time may explain our results in the results section.

[6] In the rare cases where an agency is located in more than one county we assign the hate crime to all counties the agency is active in; this only applies to 0.08% of all incidents.

[7] The FBI hate crime data do not contain information on the US territories of Virgin Island, Puerto Rico, Northern Mariana Islands, American Samoa, and Guam.

each president since George H. W. Bush; we also plot the 95% confidence interval around the mean.[8]

We split the presidency of Barack Obama into two periods based on Trump's official campaign start. We use this time split because Trump's presidential run does not only mark a cesura for Trump's presence in the media, but is also an important breaking point in his Twitter reach. Figure 3a shows that the number of retweets Trump received grew considerably with each month of his presidential campaign.

[Figure 1 about here.]

Over the 27-year period for which the FBI publishes data, the number of hate crimes against Muslims in the United States has increased. Anti-Muslim hate crimes were somewhat less common under Obama than under George W. Bush. Most strikingly, the period after Trump's presidential campaign commenced is a clear outlier by historical standards: the average number of anti-Muslim hate crimes doubled compared to Obama's presidency before Trump's campaign. This increase still stands out in comparison to George W. Bush's presidency, which included the largest recorded spike in anti-Muslim hate crimes in the wake of the 9/11 terror attacks (Gould & Klor, 2016; Panagopoulos, 2006; Hanes & Machin, 2014).

We plot the number of total hate crimes, for which we do not observe a similar increase, in Panel B of Figure 1. While we still observe slightly higher numbers compared to Obama, the frequency of hate crimes is lower under Trump than under Clinton or George W. Bush. We show in Appendix A.2. that this finding also holds true when we split the total number of hate crimes into the underlying categories (e.g. hate crimes motivated by racial bias). We conclude that the beginning of Trump's presidential campaign appears to coincide with a rise in anti-Muslim sentiment in the United States.

2.2 Measuring County-Level Twitter Usage

Twitter does not publish statistics on the number of active users per US county. We create an approximate measure of Twitter usage in each US county using 475 million geo-located tweets collected by Kinder-Kurlanda et al. (2017) made available through the Gesis Datorium. The data were collected between June and November in 2014 and 2015 by repeatedly calling the Twitter streaming API, restricted to US tweets. The streaming API provides a 1% sub-sample of public tweets each time it is called. While the exact underlying sampling procedure is unknown, this process should result in a good approximation of overall Twitter activity.

[8] For Trump's presidency, we only have information until December 31, 2017, since the FBI only publishes hate crime data for the previous year in November. For the presidency of George H. W. Bush we only have data from 1991 onward.

These tweets were assigned to counties based on the geographic location of each tweet. Figure 2b visualizes the Twitter activity per capita. Unfortunately, the data do not contain information for Alaska and Hawaii; our analysis therefore focuses on the continental US.

[Figure 2 about here.]

2.3 Measuring Trump's Twitter Activity

To understand Trump's Twitter activity, we collect the universe of his tweets from the Trump Twitter Archive (Brown, 2018). Our version of this data set contains 35,137 tweets for the time period of April 2009 to November 2018. The data contain the date, time, and text of each tweet and the number of retweets a tweet received.

Identifying Trump's anti-Muslim Tweets We use the text of Trump's tweets to identify tweets about Muslims or Islam-related topics. We start by hand-coding a random subsample of 5000 tweets in which we tag anti-Muslim tweets. These 5000 tweets form the training sample for a machine learning classifier. In preparation for machine learning we remove stopwords from and reduce all words to their morphological routs, so called lemmas. We then extract all unigram, bigrams and trigrams which appear in at least 3 tweets. The extracted n-grams are reweighted using term frequencyinverse document frequency (tf-idf). In this step the the frequency of a n-gram v in document d is replaces by $tfidf(f_{d,v}) = (1 + \ln(f_{d,v}) \cdot (\ln(\frac{1+D}{1+d_v}) + 1)$, where d_v is the number of documents n-gram v appears in. Afterwards, we train a classifier based on a logistic regression model with L1 regularization. We decide the optimal regularization strength using 5-fold cross-validation. The final model achieves and out-of-sample F1 score of 0.97. In the total sample of Trump's tweets the classifier tags 266 anti-Muslim tweets.

As we use the words "muslim", "islam", "terror", "mosque", "refugee", and 'sharia" to collect data on Google searches and news reports on Muslims, we add any tweet containing these words to the set of potential anti-Muslim tweets. This process tags an additional 57 Tweets as anti-Muslim. To rule out that we are picking up unrelated topics by mistake and change the coding of tweets if necessary. In the in the online appendix, we list examples of anti-Muslim tweets (see Table A.5) and the 25 tweets we removed in the hand-coding step (see Table A.6).

To further understand the topics of Trump's tweets during his presidency, we use Amazon Mechanical Turk (mTurk) and let three individuals code Trump's tweets in 2017 into the following categories: Media, Islam and Terrorism, Party Politics, Immigration, Foreign Policy, Domestic Policy and Other. We also code the sentiment of each tweet. More specifically, the same three individuals code the sentiment of each tweet either as "very negative", "negative', "neutral", "positive" or "very positive". We recode these categories into a scale from -2 (very negative) to 2 (very positive). In our analysis we then use the modal topic and the average sentiment coded by the three individuals.

Understanding Trump's Twitter reach. Figure 3 shows that Trump has the Twitter reach to potentially influence a considerable fraction of Americans. Figure 3a plots the monthly number of retweets he received since joining Twitter. It is apparent that the number of retweets increased with Trump's presidential run (marked by the vertical line). This suggests that a large number of people read his tweets. In Figure A.2 in the online appendix we additionally show that Trump's tweets about Muslims are significantly more widely shared than his tweets about other topics.

In Figure 3b, we plot the number of tweets using the hashtags #StopIslam and #BanIslam, as well as the number of these tweets coming from Trump's Twitter followers (see section 2.6). To construct these counts, we obtained the Twitter user IDs of all people who follow Trump on Twitter. The figure shows that the majority of the tweets using these hashtags come from Trump's followers. This lends credence to the idea that many people who harbor anti-Muslim sentiments self-select into following Donald Trump on Twitter, which exposes them to his tweets.

To provide direct evidence for the spillovers of Trump's anti-Muslim tweets on his followers, we collect the tweets for a random 1% sample of Trump's followers. These over 115 million tweets allow us to investigate if Trump's followers react to his content about Muslims.

[Figure 3 about here.]

2.4 Twitter Data for South by Southwest and Other Festivals

To construct our instrument we collect data using the Twitter application programming interface (API). In particular, we collect the universe of people following the Twitter account of SXSW Conference & Festivals (SXSW). This yields 658,240 unique user IDs. For each of these users, we collect information on their location and the date the account was created. In line with the findings of Takhteyev et al. (2012), around 75% of Twitter users in the sample report their geographical location. Previous research suggests that these user locations yield valid proxies for Twitter usage (e.g. Takhteyev et al., 2012; Haustein & Costas, 2014). As an alternative measure, we also search for tweets containing the term "SXSW" in the year 2007. We do not search for hashtags, since Twitter only formally adopted these in July 2009. In total, we find 5,933 tweets mentioning the SXSW festival.

To compare Twitter activity at the 2007 SXSW festival to other festivals in the same year, we additionally collect the tweets and user data for the Austin City Limited Festival, Burning Man, Coachella, Electric Daisy Festival, New Orleans Jazz and Heritage Festival, Lollapalooza, Pitchfork Music Festival and the West by Southwest Festival. The full list of search terms for these festivals can be found in Table A.7.

Since we are also interested in the impact of the SXSW festival on overall Twitter activity, we create a proxy for the total number of tweets using the 100 most common English words for January through March 2007 (the full list of words is reported in Table A.8). While this approach does not give us the universe of tweets in this time window, it should serve as a valid proxy for how many people are using Twitter over time.

2.5 Information on Trump's Golf Trips

Information on Trump's golf outings was collected by the New York Times (NYT, 2019). The information covers Trump's travels and identifies sources indicating that he was in fact golfing on any given trip. We cross-check these data using information from *trumpgolfcount.com* and the official Presidential schedule from the White House. In this process we add a few additional days of golf. Table A.9 in the online appendix describes these sources in more detail; Figure A.11 graphs the days in 2017 Trump spent golfing, where the darker shade of orange indicates golf outings longer than three days. More than two thirds of golf days are on the weekend, although he has also golfed multiple times on all days of the week (also see Table A.24 in the online appendix).

2.6 Additional Data Sources

We construct a large number of additional variables, which mostly serve as controls. A more detailed variable description and the relevant data sources can be found in Table A.1.

County-level variables We collect demographic control variables at the county level from the United States Census and the American Community Survey. In particular, we use information on the yearly population, the share of the population by age group, the ethnic composition of the population, the poverty rate and education levels. Information on a county's unemployment rate and industry level employment shares were obtained from the Bureau of Labor Statistics. County-level election results are available from the webpage of the MIT Election lab. The number of Muslims in each US county is derived from the 2010 US Religious Census. Additionally, we make use of county-level crime statistics based on the FBI's UCR data. Information on TV viewership patterns was collected from Simply Analytics.

We create proxies for anti-Muslim Twitter content by collecting tweets containing the hashtags "#BanIslam" or "#StopIslam" from 2010 to 2017. We selected these hashtags because they are both clearly anti-Muslim and commonly used on Twitter (Miller & Smith, 2017). Following the same procedure as for the SXSW tweets, we assign these tweets to counties based on the location of the users.

Lastly, we study potential preexisting prejudices and xenophobic sentiments at the county level based on data on hate groups from the webpage of the Southern Poverty Law Center (SPLC). The data contain information on the name of the state and city a hate group is active in. We use this information to assign the hate groups to counties. While the classification of hate groups is subjective and subject to controversy, the information gathered by the SPLC is widely used as a proxy for where hate groups are located.[9]

Time series variables To study the content of cable news, we collect TV news mentions of Muslims from the TV News Archive of the Internet Archive. We scrape news mentions for Fox News, CNN and MSNBC based on the same search terms we used for the initial classification of Trump's tweets ("sharia", "refugee","mosque", "muslim", "islam"). In total we collect 82,520 news mentions from the start of Trump's presidential campaign to the end of 2017.

We are also interested in the overall salience of Islam-related topics on the internet. We use Google Trends to obtain daily trends for the above search terms for the US. Unfortunately, Google trends only allows us to collect the daily search interest for a 90 day period. We therefore separately collect the Google trends in 90 day intervals for the period since Trump's presidential campaign commenced. Since Google normalizes the search interest between 0-100 for each 90 day period, we use the weekly search interest, which is available for the period as a whole to bring the daily search to the same scale. We describe this process in more detail in Appendix A.1.4.

Lastly, we compile information on terror attacks by Islamist from the Global Terrorism Database. In particular, we calculate the daily number of Islamist terror attacks. We split terror attacks by their location and consider terror attacks that occur in the US, Europe, or other locations separately. For the years 2015-2017 our data contain 182 terror attacks.

[9]Note that, as long as the geography of potential misclassification of hate groups by SPLC is random, this will bias our estimates towards zero.

3 Social Media and Anti-Muslim Sentiment

3.1 Introductory Correlations

Could social media play a role in the spread of anti-Muslim sentiments starting around the time of the 2016 presidential campaign? If that were the case, we would expect the increase in hate crimes documented in Figure 1 to be concentrated in areas where many people use Twitter. To get a first pass at this question, we estimate panel regressions in the following form:

$$\begin{aligned} Hate\ Crimes_{cw} = &\sum_{y=2010}^{2017} \beta_{\tau=y} \cdot Twitter\ Usage_c + \mathbf{X'_{cw}}\gamma \\ &+ County\ FE + Week\ FE + \epsilon_{cw} \end{aligned} \tag{1}$$

where the outcome variable is the natural logarithm of anti-Muslim hate crimes in county c and week w (with one added inside). $Twitter\ Usage$ is the natural logarithm of the total number of tweets in a county (also with one added inside). To simplify the interpretation of the coefficients we standardized the variables to have a mean of zero and standard deviation of one. The county fixed effects in the regression control for underlying differences in the number of hate crimes per county, while week fixed effects absorb changes in such crimes that affect all counties to the same extent. The main regressors of interest are β_τ, which measure the differential change in anti-Muslim hate crimes in counties with higher Twitter usage in year τ.

[Figure 4 about here.]

Figure 4a plots the estimated coefficients of Equation (1). The figure reveals that the increase in anti-Muslim hate crimes starting in 2015 appears to be concentrated in areas with high Twitter usage. The coefficients for previous years are close to zero and not significant, which suggests the counties followed similar trends in the pre-period. Given that all coefficients have been standardized the magnitude of the coefficients indicate that a one standard deviation increase in Twitter usage is associated with an 0.1 standard deviation increase in anti-Muslim hate crime.

As corroborating evidence for the spread of anti-Muslim sentiment via Twitter, we repeat the event study regressions for the hashtags #StopIslam and #BansIslam. Figures 4b and 4c plots the estimates for these outcome variables. The figures suggest that not only offline but also online sentiments about Muslims grew disproportionately more negative in counties with higher social media penetration.

The evidence here suggests a potential connection between anti-Muslim sentiment and Twitter usage. However, our proxy for Twitter usage is likely correlated with a host of observable and unobservable factors that might also affect hate crimes. To overcome this challenge, in the next section we develop an identification strategy to isolate the effect of social media.

3.2 Identification Strategy

The evidence in the previous sections suggests that the increase in anti-Muslim hate crimes around Trump's presidential run has been concentrated in areas with high social media usage. In this section, we address the concern that social media usage may be correlated with other factors by developing an instrumental variable strategy based on the early diffusion of Twitter.

The starting point is a county-level first-difference model relating the shift in anti-Muslim hate crimes in mid-2015 to a measure of social media usage:

$$\Delta Hate\ Crimes_c = \alpha + \beta \cdot Twitter\ Usage_c + \mathbf{X_c'}\gamma + State\ FE + \epsilon_c. \tag{2}$$

As a baseline, $\Delta Hate\ Crimes$ will refer to the log-change of hate crime incidents aimed at Muslims or other groups (with one added inside) with Trump's presidential run. The pre-period is defined as the years from 2010 onward.[10] $Twitter\ Usage$ is the natural logarithm of tweets sent from a given county, our measure of social media use. All regressions will control for state fixed effects and dummies for each decile of the population distribution.

$\mathbf{X_c}$ is a vector of control variables that further includes demographic controls for population growth and the share of the population in five-year age buckets; the linear distance from each county centroid from Austin Texas, the location of the SXSW festival we will describe in more detail below; controls for ethnic composition and the share of Muslims; socioeconomic controls including the share of high school graduates or people with a graduate degree, the poverty rate, the unemployment rate, local GINI index, the share of uninsured individuals, the log median household income, the employment shares in eight sectors; media controls for the viewership share of Fox News, the cable TV spending to population ratio, and the prime time TV viewership to population ratio; and the county-level vote share of the Republican party in 2012. Standard errors in all specifications are clustered at the state level.[11]

[10] In further robustness checks we show that our results neither depend on the pre-period we use in the first-difference nor on the specific functional form. The results also hold for the *level* of hate crimes after Trump's presidential run.

[11] In Table A.20 in the online appendix, we show that our results also hold using alternative ways to construct standard errors.

When estimating equation (2) using OLS, the point estimates for β in Equation (2) are likely biased because Twitter usage is not exogenous. In particular, one may be concerned that the factors driving people to commit hate crimes are correlated with the decision to adopt social media. This could give rise to alternative interpretations of the graph in Figure 4a and the β estimate in Equation (2). To give one example, perhaps the potential perpetrators of hate crimes live predominantly in areas with a sizable presence of minority groups, and those areas are also more likely to use Twitter. In that case, the period around Trump's campaign start could still be interpreted as a trigger point for anti-Muslim sentiments, but it is not clear whether or to what extent social media plays a role.

To circumvent this issue, we exploit plausibly exogenous variation in the early adoption of Twitter in the United States. More precisely, we make use of the fact that Twitter's popularity reached a tipping point at the SXSW conference and festival in 2007. During the event, the daily volume of tweets increased from around 20,000 to 60,000 (Gawker, 2007). Figure 5a gives a first indication that SXSW may have led to a trend break in the success of Twitter: we see a clear spike of tweets about the event during the SXSW conference in mid-March 2007, followed by an upward shift in the growth of the total number of tweets. While total tweets grew by 60% from February to March, this growth accelerated to over 240% from March to April. March 2007 is also a clear outlier in the number of SXSW followers that signed up to Twitter (see A.9 in the online appendix).

[Figure 5 about here.]

A number of facts suggest that the early adopters at SXSW were key to Twitter's rise. As a first indication, in 2007 there were more tweets about SXSW than about other major festivals (see Figure 5b).[12] This is noteworthy because of the lower attendance at SXSW Interactive. We can also see that the spread of Twitter across counties followed the early adopters. To show this, we run event study panel regressions to compare Twitter activity in counties with and without new SXSW followers in March 2007. Figure 6 plots the results. Areas with early adopters at SXSW did not exhibit a higher growth rate of Twitter activity prior to SXSW Interactive 2007 but the growth rate increased in its aftermath. Quantitatively, counties with a one standard deviation higher number of SXSW followers in March (1.91) increased their local twitter activity by 10% of a standard deviation in April compared to February 2007.

[Figure 6 about here.]

[12]This pattern also holds when we consider tweets about the festivals for the whole of 2007 (see Figure A.8).

We exploit that this pattern of technology adoption persists until today. As we will show below, the number of SXSW followers in a county who registered during the festival period are predictive of Twitter penetration across US counties. This is in line with the literature on the path dependence of technology adoption (e.g. Arthur, 1989, 1994; Liebowitz & Margolis, 1999; Arrow, 2000). Crucially, this is still true after controlling for the number of SXSW followers in a county *prior* to the tipping point in March 2007, or alternatively for users tweeting about the much more popular festivals Coachella, Burning Man, and Lollapalooza in the same year.

The historical diffusion of Twitter gives rise to a difference-in-difference instrumental variable framework. We collapse the time dimension into an IV setting where the first stage equation is given by:

$$\begin{aligned} Twitter\ Usage_c = \alpha &+ \delta_1 \cdot SXSW\ followers,\ March\ 2007_c \\ &+ \delta_2 \cdot SXSW\ followers,\ Pre_c \\ &+ \mathbf{X_c'}\psi + State\ FE + \xi_c, \end{aligned} \tag{3}$$

where $SXSW\ followers,\ March\ 2007$ is the number of SXSW followers in county c that joined Twitter in March 2007, which serves as the excluded instrument. $SXSW\ followers,\ Pre$ are followers that joined before the festival at any point in 2006. This controls allows us to address the concern of inherent differences of counties with SXSW followers.[13]

Similar to Enikolopov et al. (2016), the identifying assumption underlying our empirical strategy is that, conditional on a large number of county characteristics, the decision to start following SXSW in March 2007 rather than in the months before drives increases in anti-Muslim sentiments with the 2016 presidential campaign only through the diffusion of Twitter usage.[14] Three pieces of evidence suggest that this assumption is reasonable. First, as shown above, counties with Twitter adopters around SXSW did not differ in Twitter adoption prior to the festival. This suggests that these counties are not inherently different. Second, a comparison of the Twitter profiles of users signing up for Twitter around SXSW with those who signed up before suggests that they are highly similar. Table A.13 shows that users' first names and the terms they use to describe themselves are almost indistinguishable between these two groups. The correlation of words mentioned in the "bio" of these groups is 0.92. Third, the home counties of SXSW followers who signed up during the 2007 event do not

[13] In the robustness section below, we consider a large range of alternative control sets based on different time periods to hold selection into social media usage constant.

[14] With the alternative festival controls, the assumption is similar in that attending SXSW rather than other festivals in 2007 should only affect outcomes through this social media adoption channel.

systematically differ in observable characteristics from those of users who signed up before (see Table A.12).

Figure A.1 in the online appendix plots the distribution of our proxy of new SXSW followers in March 2007 across US counties. People from 155 counties were early adopters of Twitter at or around the time of SXSW. Table A.14, also in the online appendix, plots the correlation coefficients between the county-level SXSW measures and those for the other festivals. Although these variables are strongly correlated, as one would expect, there is enough variation in the locations of SXSW followers we can exploit in our empirical strategy. In robustness exercises, we consider a large range of alternative SXSW metrics, some of which show a considerably lower correlation between "treatment" and "control" group.

Since our baseline outcome variable is differenced over time, we also require that the parallel trends assumption holds. We already saw in Figure 4a above that hate crimes against Muslims disproportionately increased in areas with high Twitter usage only *after* Trump's presidential campaign started. In the online appendix in Figure A.4 and Figure A.7, we provide additional reduced form evidence in support of parallel trends when comparing areas with and without users that likely attended SXSW in March 2007.

3.3 South by Southwest and Twitter Adoption: First Stage Results

To assess whether the initial diffusion of Twitter at SXSW still matters for social media use today, we report the results of estimating the first stage Equation (3) in Table 1. We can see that across the board the number of new Twitter users in March 2007 who followed SXSW is highly predictive of Twitter usage today. The point estimates are always statistically significant at the 1% level. The coefficient for SXSW followers in the months prior to the 2007 event is not statistically significant as soon as we control for observable county characteristics. Indeed, an F-test for the equality of coefficients suggests that the March 2007 and pre-period estimates are also statistically different from each other. Importantly, the coefficient estimates for March are highly stable and do not depend on the included covariates. Quantitatively, the estimate of 0.362 in column 8 implies that a one standard deviation increase in the log number of new SXSW followers in March (0.32) is associated with 12% higher Twitter usage today. The estimated effect based on the pre-period estimate implies 1% more users, which is not distinguishable from zero.

[Table 1 about here.]

Based on these estimates and the event study plot in 6, we conclude that county-level differences in the early adoption of Twitter spread through the 2007 SXSW conference and festival are a reliable predictor of Twitter usage in the United States today. Because the location of early adopters in the period before the festival does not predict Twitter usage, it is unlikely that this result is driven by selection into following the SXSW festival's Twitter page. In the next sections, we will conduct more robustness checks to test the validity of this insight and will employ the strong first stage result to estimate the effect of social media propagation on the recent rise in anti-minority sentiments, particularly those aimed at Muslims.

4 Main Results

4.1 Reduced Form Estimates

We next turn to the reduced form estimation results for the change in hate crimes against Muslims around Trump's presidential campaign start. Table 2 presents these results. Across a large number of different specifications, we find that the early adoption of Twitter – measured by the number of SXSW followers who joined Twitter in March 2007 – is associated with an increase in hate crimes against Muslims. The estimates for the March coefficient are strikingly similar irrespective of the included control variables. The estimates on new SXSW followers in previous months are not statistically significant and considerably smaller.[15]

Figure A.4 in the online appendix plots the reduced form estimates from difference-in-difference panel regression of the type in Equation (1). Note that this regression also controls for the locations of SXSW followers in previous months interacted with year fixed effects. As above, we find that hate crimes against Muslims did not disproportionately increase in areas with new SXSW followers in March 2007 prior to the period of Trump's presidential campaign. Afterwards, however, these counties experienced a large upward shift in such incidents.

Taken together, we interpret these results as first evidence that social media may play a role in the propagation of hate crimes as a result of Donald Trump's campaign. Because we control for the number of SXSW followers in the months before SXSW 2007, these results are unlikely to be driven by a selection of individuals from areas prone to hate crimes into participation in that particular festival. In the next sections, we provide the formal two stage least squares estimates and conduct further robustness checks in support of this interpretation.

[Table 2 about here.]

[15] Note that the standard deviation of these pre-SXSW users is around half that of the March 2007 variable.

4.2 IV Estimates

The results in the previous section can be interpreted as evidence that social media plays a role in the recent increase in hate crimes in the United States. In this section, we use the proxy for new SXSW followers in March 2007 as an instrument for Twitter usage across the US today, while holding interest in SXSW prior to the key event constant to alleviate selection concerns.

Table 3 provides two sets of results. In panel A, we plot the OLS results from regressions of the change in hate crimes against Muslims on our measure of Twitter usage. In panel B, we report the 2SLS results and a number of diagnostic tests. The results suggest that social media penetration, measured by Twitter usage, is positively associated with the increase in hate crimes against Muslims. The 2SLS estimates in column 8 imply that a one standard deviation increase in Twitter usage (1.91) is associated with a 38% larger increase in hate crimes after Trump's presidential campaign launched.

A well-known concern with IV estimation is the weak instruments problem, which can lead to biased point estimates. We believe that our estimation does not suffer from a weak first stage for three reasons. First, the robust F-statistic for the excluded regressor ranges between 41 and 68 in columns 1 through 8.[16] Second, the values of the F-statistic are above the critical values to reject the null hypothesis of a 5% potential bias with 5% statistical significance derived in Olea & Pflueger (2013), which is 37.42. These authors extend the well-known thresholds of Stock & Yogo (2005) to the case of heteroskedasticity-robust and, relevant in our case, clustered standard errors.

[Table 3 about here.]

We also assess the significance of our main estimates using confidence sets based on test inversion that are valid whether or not instruments are weak. For the case of a single instrument we study here, Andrews et al. (2019) recommend reporting Anderson-Rubin (AR) confidence sets that are efficient and robust to weak identification (Anderson et al., 1949). Andrews (2018) develops a two-step approach to construct these confidence sets that is implemented in Stata by Sun (2018). Basing inference on this two-step approach sidesteps the issue that the usually reported (Wald) confidence intervals for 2SLS estimates can exhibit large coverage distortions. This is because AR confidence sets allow for inference without assessing the strength of first-stage results in a separate initial step. As such, we can determine whether our second stage coefficients are likely to be non-zero even if our instrument was

[16] Note that because the model is just-identified, the robust F-statistic (sometimes also called Kleibergen-Paap) is equivalent to the effective F-statistic derived in Olea & Pflueger (2013).

indeed weak. Reassuringly, the AR confidence sets reported below the (instrumented) Twitter usage in panel B always exclude zero.

Because our estimations do not appear to suffer from a weak instrument problem, we can use the comparison of the OLS and 2SLS estimates to assess whether the selection of individuals into social media adoption is positively or negatively correlated with the incidence of hate crimes. In other words, we can test whether the OLS estimates are upward or downward biased. Across all specifications in Table 3, the OLS estimates are highly statistically significant, but considerably smaller than those obtained using 2SLS. This difference suggests negative selection into social media usage. To give one example, if people in particularly xenophobic areas commit more hate crimes but are less likely to use Twitter, the OLS estimate would be downward biased. This selection effect is also consistent with Enikolopov et al. (2016): for the case of social media and protest participation in Russia, they find much larger IV estimates compared to OLS.[17]

In Table A.19 in the online appendix, we investigate which types of hate crimes increased particularly in areas with higher social media usage. It turns out that our results seem to be almost entirely driven by a rise in assaults. This makes it unlikely that we are capturing changes in *reporting* rather than the actual incidence of hate crimes, since we have no reason to expect reporting changes to be limited to particularly severe cases. We relegate a more extensive discussion of reporting changes to Section 7

A conceptual question raised by these estimates is the extent to which any potential causal effect of social media can be directly attributed to Twitter, rather than other platforms. While the initial diffusion through SXSW in 2007 was probably specific to Twitter, there were likely significant spillovers in the adoption of other social media platforms. Since we only observe the equilibrium outcome of these spillovers today, our estimates might not identify a pure "Twitter effect". What matters for the interpretation of our estimates is that this diffusion is limited to social media, which we believe is plausible.

4.3 Robustness

We consider a range of sensitivity checks to validate the robustness of our main findings. We begin by reporting an additional set of results that test alternative ways to account for the selection of users into events such as SXSW. In particular, we replace the control variables for new followers of SXSW at any point in 2006 with users tweeting about *other* festivals in 2007 that are, in many respects, very similar to SXSW. We consider tweets about three of the most popular festivals in the United States: Coachella, Burning Man, and Lollapalooza.

[17] Another interpretation of the 2SLS estimate is that counties with more SXSW followers that signed up in March 2007 have a higher local average treatment effect (LATE).

More precisely, we define control variables that capture the log number of users from each county that tweeted about these festivals in the month of 2007 in which they were held.

Table A.16 in the online appendix reports the results for the reduced form and 2SLS estimations with these alternative controls in panel B and C, respectively. To aid comparison, we again plot the OLS results in panel A. As before, we find that the impact of Twitter usage on changes in anti-Muslim hate crimes is highly statistically significant throughout. Crucially, the log number of users tweeting about the other festivals is statistically insignificant, which is another indication that we are not merely capturing a selection of particular people into areas with hate crimes and high Twitter usage. The estimates and F-statistics for the 2SLS results are somewhat larger than the baseline findings in Table 3.

We also consider alternative transformations of the SXSW variables in Table A.18 in the online appendix. In column 1, we begin by showing that the results also hold when dropping the SXSW control, which makes the results somewhat stronger. In columns 3 through 6, we consider alternative time periods for the pre-period variable or alternatively control for the individual months. Columns 7 through 11 replace the SXSW follower variables with dummies for counties in which we can locate any tweet about SXSW in March 2007 or previous periods. Importantly, these specifications vary widely in the number of "treatment" and "control" counties, as well as the correlation between the treatment and control SXSW variables. Our results are robust throughout, which suggests our findings are not driven by any particular specification.

We also use alternative metrics of Twitter usage in Table A.17 in the online appendix. We consider two survey measures of Twitter usage provided by GfK Mediamark Research & Intelligence (via SimplyAnalytics), as well as two alternative transformations of the GESIS Twitter data (only tweets before June 2015 or the number of Twitter *users*, rather than the number of tweets). All of these measures yield similar estimates.

In Table 4, we present additional robustness checks. In column 1, we drop state fixed effects, which makes little difference to the point estimates. In column 2, we consider the change in anti-Muslim hate crimes since 1990 (rather than 2010); this yields larger estimates throughout. In column 3, we replace the change in hate crimes with the log number of hate crimes after Trump's presidential run as dependent variable. This also yields significant estimates.

In columns 4 through 6 of Table 4, we address the concern that anti-Muslim hate crimes reported by the FBI mainly occur in a relatively small fraction of all counties. In column 4, we begin by dropping all counties that report a zero change in anti-Muslim hate crimes between 2010 and 2017. Because this applies to the majority of counties, the sample size shrinks considerably. One way to think about this estimation is that it captures the intensive margin of hate crimes. Despite the drop in observations, our estimates remain statistically significant. In column 5, we next drop counties for which the FBI always reports zero hate crimes. Reporting may be less reliable for these counties. As it turns out, this exclusion makes little difference for our estimates. As a last exercise, we drop all counties for which the (rounded) estimated share of Muslims in the total population is zero from the sample in column 6.[18] Again, the results we obtain in this sample are very similar to those in the main sample.

In column 7, we weight all estimates by population, which makes little difference to the results. In column 8, we restrict the sample to neighbouring counties where one has no new SXSW followers in March 2007 and the other one has at least one. This is to purge the estimates of potential unobserved local confounders. This yields similar estimates. At last, in column 9, we transform the dependent variable into an index equal to 1 for increases in anti-Muslim hate crimes, 0 for no change, and −1 for decreases; again, our findings remain similar.

[Table 4 about here.]

4.4 Social Media and Changes in Other Hate Crimes

Up to this point, we have focused on changes in anti-Muslim hate crimes, motivated by the fact we found little change in the frequency of other types of hate crimes around the start of Trump's presidential campaign in the FBI data. However, one might expect Trump's presidential run to also affect other categories of hate crimes, in particular anti-Hispanic incidents.[19] If social media plays a role, such incidents may have become more common in areas with high Twitter usage even if their total number remained unchanged.

In Table 5, we consider this possibility empirically by replacing the dependent variable with the log change in hate crimes targeting on Hispanic ethnicity, other ethnicities, race, sexual orientation or religion (excluding anti-Muslim bias). We also consider hate crime data from the Anti-Defamation League (ADL) as an alternative data source in column 7. The ADL only appear to report a large number of hate crimes from 2016 on, so we focus on the *level* rather than the change in hate crimes.[20]

[18] Although the Religious Census reports no Muslims living in these counties, this might be the artifact of a very small number, rather than an actual zero.

[19] In his presidential campaign announcement speech, Trump infamously singled out Hispanics and Arab Muslims: "When Mexico sends its people, theyre not sending their best. ... Theyre bringing drugs. Theyre bringing crime. Theyre rapists. And some, I assume, are good people. ... Theyre sending us not the right people. Its coming from more than Mexico. Its coming from all over South and Latin America, and its coming probably – probably – from the Middle East."

[20] In unreported results, we find similar results using a measure of the change in local hate crimes as reported by ADL.

Overall, we also find a role for social media in explaining increases in the total number of hate crimes and those targeting Hispanics, the other minority group frequently singled out by Donald Trump. However, only anti-Muslim hate crimes show a consistent pattern across the OLS and 2SLS estimates. There is little evidence for a reallocation of other hate crimes towards areas with higher Twitter usage. In the 2SLS estimation, a one standard deviation increase in Twitter usage is associated with a 35% larger increase in total hate crimes, and a 33% larger increase for incidents targeting Hispanics.[21] The difference of these estimates compared to the OLS results likely arises because of selection: social media, and Twitter in particular, is likely adopted more by areas with more technologically-savvy people who are probably less likely to commit hate crimes. This creates a downward bias for the OLS estimates.

[Table 5 about here.]

4.5 Heterogeneous Effects: Social Media and Pre-Existing Bias

The results in the previous sections raise the question whether exposure to social media is changing people's beliefs about Muslims or if social media rather reinforces existing prejudices. To address this question, we investigate whether the effect of Twitter usage is driven by counties that are more likely to be susceptible to anti-Muslim messaging.

In particular, we repeat the event study regressions from Section 3.1 and split counties by whether the Southern Poverty Law Center (SPLC) identifies at least one hate group. Note that these sample splits do not estimate whether anti-Muslim hate crimes increased in counties with hate groups but rather whether Twitter usage has a different impact in these counties.

Figure 7 plots the estimated coefficients from this exercise.[22] We find that higher Twitter usage is only associated with more anti-Muslim hate crime in counties with hate groups. In contrast, counties with high Twitter usage but no hate group continue to follow the same trajectory as low Twitter usage counties. Quantitatively, among the counties with at least one hate group a one standard deviation increase in Twitter usage is associated with a 0.6 standard deviation increase in anti-Muslim hate crime. In Panel (b), we provide similar evidence for counties that are above the 90th percentile of hate crime per capita (all motivating biases) in the pre-period. We again observe that the increase in anti-Muslim hate crimes is driven by counties with high Twitter usage and pre-existing biases.

[Figure 7 about here.]

[21] Figure A.5 and Figure A.6 in the online appendix plot the OLS and reduced form event study graphs.

[22] To reduce clutter, the figures report the estimated coefficients without confidence bands. We report the full regression results with standard errors in Table A.21 in the online appendix.

Taken together, the findings are at least some evidence that social media did not necessarily change people's beliefs, but rather triggered existing negative attitudes towards Muslims around the time Trump started his presidential run. This is consistent with the view that people infer information about the social acceptability of viewpoints and actions based on what they see online. As such, it appears possible that after observing increased anti-Muslim rhetoric on Twitter (as documented above), already radicalized individuals might have become more willing to commit violent acts against Muslims in real life. If this is the case, spikes in anti-Muslim sentiment on social media might work as "triggers", a possibility we investigate in the next section.

It is also worth noting that the sample splits are another indication that we are unlikely to capture changes in the propensity to report hate crimes rather than an actual increase in incidents. We discuss this issue in more detail in Section 7.

5 Trump's Tweets and Anti-Muslim Sentiment

The previous section suggests that social media may have played a role in the spread of anti-Muslim sentiment associated with the start of the Trump campaign. An often-voiced hypothesis is that Trump actively contributes to anti-Muslim sentiment through his incendiary comments on Twitter. Indeed, there is some existing evidence that influential individuals can have a disproportionate effect on public opinion (e.g. Beaman et al., 2009; Bursztyn et al., 2017; Alatas et al., 2019).

We attempt to shed some light on this mechanism by analyzing the time series relationship between Trump's tweets about Muslims, anti-Muslim hate crimes, and media attention. We attempt to get at the issue of causality by again leveraging an instrumental variable. The main purpose is to provide evidence for a channel through which social media could contribute to a climate that enables hate crimes and investigate the importance of prominent only figures. Table A.23 and Table A.29 plot the summary statistics.

5.1 Trump Tweets and Hate Crimes

If there is a relationship between Trump's Twitter activity and physical hate crimes, the timing of both should coincide. We thus begin by plotting the number of Trump's tweets about Islam-related topics and anti-Muslim incidents over time in Figure 8. We define these tweets based on a careful reading of Trump's Twitter feed, combined with a machine learning algorithm; see the data section and online appendix Table A.8 for more details. Since the daily number of tweets is highly volatile, we plot the 14-day moving average of the series; collapsing the data on the weekly level looks very similar (unreported).

It is immediately apparent that Trump's tweets about Muslims and anti-Muslim hate crimes are highly correlated. This correlation could reflect that Trump reacts to US-wide anti-Muslim sentiments driven by observable and unobservable factors, e.g. terrorist attacks. It could also be that Trump's tweets themselves contribute to a climate that enables hate crimes. Clearly, we cannot disentangle these possibilities using the graphical evidence from the data nor using a simple OLS regression of hate crimes on tweets.

[Figure 8 about here.]

We propose an instrumental variable strategy to get around the most obvious reverse causality concerns. In particular, we leverage Trump's passion for golf: in 2017 alone, Trump likely golfed on 92 days. As it turns out, the data suggest a strong link between Trump's golf outings and his Twitter feed: Figure 9 shows that while the total number of tweets he sends are unchanged on golf days, the *content* of his tweets sharply tilts towards negative, Muslim-related rhetoric. In 2017, 15 out of the 34 tweets we classify as negatively mentioning Muslims were sent on golf days. In Figure A.13 in the online appendix, we show that the topic shift is explained by a drop in policy-related tweets and more frequent mentions of Muslims and the media. Figure A.14 shows that his tweets also become more negative in sentiment. One intuitive explanation for this pattern is that once Trump is away from the White House, his attention shifts away from policy issues. Another influence on Trump's social media activity that is likely stronger on golf days is his social media manager Dan Scavino, who is know to have suggested tweets and topics to Trump (Edwards, 2018). Figure A.15 in the online appendix provides additional evidence that Trump's daily schedule influences the content of his tweets. In particular, we show that Trump is more likely to tweet about foreign politics when he is abroad and more likely to tweet about domestic and party politics on days he receives a policy briefing.

[Figure 9 about here.]

Because the President's schedule is to some extent predetermined to accommodate security concerns and meetings, it is plausibly exogenous with respect to hate crimes against Muslims. What matters for our identification strategy is that Trump's golf outings are not systematically correlated with unobservable anti-Muslim sentiment. One disadvantage of this strategy is that we can only analyze 2017, for which we have both details about Trump's schedule and data on hate crimes. We also present OLS regressions for the IV sample and using the full time period since Trump joined Twitter in 2009 below.

More formally, we run time series regressions using the following framework:

$$Hate\ Crimes_{t+h} = \alpha + \beta \cdot Muslim\ Trump\ Tweets_t + \mathbf{X_t'}\gamma + \epsilon_t \tag{4}$$

$$Muslim\ Trump\ Tweets_t = \alpha + \delta \cdot I[Trump\ golfs]_t + \mathbf{X_t'}\psi + \xi_t \tag{5}$$

The dependent variable in equation (4) is the natural logarithm of US-wide hate crimes against Muslims at day $t + h$ (with one added inside). The main regressor of interest is the natural logarithm of the number of Donald Trump's Muslim tweets (again with one added inside). In the baseline specification, the vector X_t includes time trends and a full set of day-of-week as well as year-month fixed effects.

Naively estimating equation (4) would not be informative about whether Trump's Twitter activity might contribute to driving sentiments because his tweets cannot be regarded as random. We will thus instrument for tweets about Muslims in equation (5) using $I[Trump\ golfs]_t$, an indicator variable that is 1 for days on which Trump plays golf (see Section 2 for more details). We base inference on Newey-West standard errors that allow for heteroscedasticity and autocorrelation.

The appropriate choice of the prediction horizon h depends on the lead-lag relationship between Trump's tweets and real-life hate crimes. We plot the result from estimating equation (4) with OLS using values for h from -4 to 4 in panel (a) of Figure 10. As we can see, the log number of anti-Muslim hate crimes is essentially flat prior to Trump's tweets and subsequently rises to its peak in T+2. In our baseline regressions, we will thus set h to 2. We repeat the baseline estimations for different time windows in Table A.27 in the online appendix. Panel (b) also plots the dynamic relationship between Trump's golf outings and tweets about Muslims. We can see that his tweets only increase on the days he golfs, with no similar spikes before and after.

[Figure 10 about here.]

Table 6 presents the regression results of this exercise. We plot the OLS coefficients in panel A, first stage coefficients in panel B, reduced form coefficients in panel C, and the 2SLS estimation in panel D. Across the different specifications, the estimations suggest a clear link between Trump's tweets about Muslims and subsequent real-life hate crimes. Notably, the reduced form and 2SLS coefficients are almost fully unchanged when we include controls for measures of the salience of Muslim-related topics based on Google searches and the number of mentions on the big three TV networks (Fox News, CNN, and MSNBC). Taken at face value, this indicates that his golf outings are indeed not timed to coincide with periods of high Muslim salience.

[Table 6 about here.]

As mentioned above, a concern with instrumental variable estimation is the weak instruments problem. Because we only have one year of data to work with, this is a particular challenge in our setting. However, two pieces of information suggest that our estimates are meaningful. First, the robust F-statistics we find are consistently above the widely used linear IV rule of thumb of 10. Most of them are above the critical value for a worst case bias of 30% (at 5% statistical significance) using the cutoffs from Olea & Pflueger (2013). Second, the Anderson-Rubin confidence sets constructed using the two-step approach proposed in Andrews (2018) always exclude a zero estimate even if we assume that the instrument is weak. The reduced form and 2SLS results thus suggest that Trump's tweets could indeed be a contributing factor triggering potential perpetrators to commit real-life hate crimes.

To get a sense of the implied magnitudes, consider the estimate in column 7 of panel D Table 6. The coefficient of 1.659 implies that a one standard deviation increase in the log number of tweets about Muslims (0.25) is associated with a 41 log-point increase in hate crimes. This effect is large and, importantly, much larger than the OLS estimate of 0.116. An obvious explanation for this difference would be the presence of a weak instrument. However, given that the diagnostic tests discussed above are relatively encouraging, another possibility is that unobserved third factors lead to a downward bias of the OLS estimates. For example, Trump's tweets about Muslims might coincide with periods of *low* pre-existing anti-Muslim sentiment. In that case, the relationship between his tweets and hate crimes estimated via OLS would be downward biased because it conflates the true Trump effect with low general anti-Muslim sentiment. This explanation is also consistent with the finding that controlling for general attention paid to Muslims or terror attacks in columns 4 through 6 *increases* the point estimates relative to the baseline specification.

A limitation of these findings is that they are limited to the year 2017. In Table A.30 in the online appendix, we re-run the OLS estimation for the entire period since Trump's first tweet in 2009 and split the sample into the period before and after the launch of his presidential run on June 16, 2015. We find very similar OLS estimates on his tweets about Muslims, but only after the start of his presidential campaign. For the much longer period from 2009 to mid-2015, his tweets seem to be uncorrelated with anti-Muslim hate crimes. While many factors may explain this finding, it is at least some indication that we are not capturing a phenomenon that is limited to a single year.

In Table A.25 in the online appendix, we report more robustness results. Our results remain largely unchanged when we control for more lags of the dependent variable to capture stronger serial correlation in hate crimes. We further experiment with additional controls for the total length of Trump's golf outings in column 3, a control if Trump golfed in the previous week (column 4), or alternative definitions of the golf dummy in columns 6 and 7. Our results are also robust to using a dummy for days with *any* Islam-related tweet from Trump (column 5).

Given the relatively short sample period, how likely would it be to find an effect if we picked golf days at random? Figure A.12 reports the results of a randomization test for the first stage regression of Trump's tweets about Muslims on a golf dummy, where we randomly pick 92 golf days in 2017 (except the ones used in the actual variable). The distribution of the resulting t-statistics of the golf day dummy suggests that none of the placebo coefficients are close to our estimate.

We further investigate which type of anti-Muslim hate crimes drive our results. Based on the most common criteria in the FBI data, we divide anti-Muslim incidents into vandalism, theft, burglary, robbery, and assault. The results of this exercise are presented in Table A.26 in the online appendix. Our high-frequency results appear to be mainly driven by cases of vandalism.[23]

As a simple validation exercise, we also investigate whether Trump's messages about Muslims are also correlated with hate crimes against other minorities. In particular, we consider incidents motivated by ethnicity, race, sexual orientation, or religions other than Islam. Table A.31 plots the predictive ability of Trump's tweets about Islam-related topics for these different types of hate crimes. We only find clear-cut correlations with crimes against Muslims, not other hate crimes. This suggests that we are not merely capturing anti-minority sentiment, but rather something Muslim-specific. We also replicate this finding using simple OLS regressions for the full sample in Table A.32. Again, we find that only hate crimes targeting Muslims are correlated with Trump's anti-Muslim tweets; the correlation with other types of hate crimes is close to zero, both before and after the start of his presidential run.

[Table 7 about here.]

5.2 Trump Tweets and Twitter Spillovers

We next provide evidence for the fact that Trump's negative tweets about Muslims have a direct effect on his followers. In particular, we analyze if Trump's followers become more willing to express anti-Muslim content. For this analysis we use more than 115 million tweets drawn from a random 1% sample of Trump's followers (around 630,000 users). In this dataset, we identify tweets that are retweets of Trump's negative content about Muslims, tweets that

[23] Note that this does not stand in contradiction to our cross-sectional results, for which we find the largest role for assault. The daily variation we exploit here likely picks up more spontaneous anti-Muslim incidents relative to the medium-term effects in the cross-section.

refer to Muslim-related topics but are not retweets of Trump, and tweets that contain the hashtag #BanIslam.

We continue to run time series regressions of the type in equation (4). To start, we plot dynamic correlations in Figure 11, where the dependent variables are different measures of tweets (in natural logarithm). The results show a clear pattern. Trump's negative tweets about Muslims are not only widely shared by his followers over the next days but also systematically followed by a spike in new content about Muslims. The time series pattern suggests no increase of anti-Muslim sentiment before Trump's tweets.

Columns 1 through 3 in Table 7 provide evidence that these patterns also hold when we instrument for the tweets using golf days. We focus on contemporaneous correlations, as suggested by the pattern in Figure 11. The reduced form and 2SLS specifications are highly statistically significant, and the weak IV confidence sets always clearly exclude zero. The 2SLS estimates suggest that a one standard deviation increase in Trump's Muslim tweets (0.25) is followed by a doubling of retweets and an almost 30% increase in new messages about Muslims that do not mention Trump. They are also followed by a 58% increase in the use of the hashtag #BanIslam by Trump followers.

These results lend credence to the idea that Trump's tweets are trigger points for anti-Muslim sentiment among his followers. The willingness of Trump's followers to produce their own anti-Muslim messages speaks to changes in the perceived acceptability of such content after a tweet by the president.

[Figure 11 about here.]

5.3 Trump Tweets and the News Cycle

As a last time series exercise, we ask whether Trump's tweets about Muslims may have the ability to affect the news cycle. This is important to understand because, unlike for the social media channel we study here, there is ample evidence that other types of media can persuade people to participate in spontaneous, potentially violent outbursts (see e.g. DellaVigna & Gentzkow, 2010; Yanagizawa-Drott, 2014). As such, one obvious channel through which social media may affect offline outcomes is through influencing what other media report on. Indeed, it has been widely recognized that Twitter has become an important dissemination channel for journalists (Willnat et al., 2019); some estimates suggest that up to a quarter of Twitter users may be working for media outlets (Haje Jan Kamps, 2015).

We investigate the effect of Trump's tweets on media coverage using transcript data from the *TV News Archive*. In particular, we replace the dependent variable in equation (4) with the log number of mentions of Muslim-related topics on a given day by the three major cable news stations Fox News, CNN, and MSNBC. Columns 4 through 7 in Table 7 present the results of this exercise. Because we find a more immediate correlation between Trump's Twitter activity and news coverage, we report specifications with $h = 0$ as the prediction horizon.

Trump's tweets about Muslims are highly correlated with TV mentions in the OLS, reduced form, and 2SLS regressions. While the 2SLS estimates are still considerably larger than those obtained from OLS, they are less so than for the hate crime estimates. For overall news coverage in column 2, for example, we find that a one standard deviation increase in Muslim Trump tweets (0.25) is associated with a 74% increase in news coverage.

However, we urge caution in interpreting these results due to the short sample period. Nevertheless, the F-statistics are almost uniformly above the rule-of-thumb of 10, and mostly above the 12.04 threshold for a maximum 30% coefficient bias with 5% statistical significance derived in Olea & Pflueger (2013). Perhaps more importantly, the Anderson-Rubin confidence sets always clearly exclude zero.

We also consider heterogeneity across news stations. The correlation of instrumented Trump tweets with TV mentions appears to be strongest for Fox News (see column 5). Indeed, for CNN and MSNBC (columns 6 and 7), a zero effect is well within the AR confidence sets. This is interesting because Fox News is well-known to be supportive of Trump, following a longer term move towards more Republican-friendly reporting (Martin & Yurukoglu, 2017). This might allow Trump's comments to be broadcast uncritically and even more widely through the channel's considerable reach. Taken together, this suggests that social media may affect the news cycle, which could be one potential trigger point for potential perpetrators of hate crimes.

6 Panel Evidence: Trump's Tweets and Twitter Usage

As the last part of our analysis, we combine the cross sectional and time series evidence. If Trump's anti-Muslim rhetoric spreads through Twitter, we should observe large increases in anti-Muslim hate crime in counties with higher Twitter usage. We investigate this hypothesis with the following regression specification:

$$\begin{aligned} Hate\ Crimes_{cd} = \beta \cdot Twitter\ Usage_c \times Muslim\ Trump\ Tweets_d \\ + \mathbf{X'_{cd}}\gamma + County\ FE + Day\ FE + \epsilon_{cd} \end{aligned} \tag{6}$$

where $Hate\ Crimes_{cd}$ is the natural logarithm of one plus the number of hate crimes in county c on day d. The main coefficient of interest β is the interaction of county-level Twitter usage with Trump's tweets about Muslims. The coefficient measures if there are disproportionate changes in anti-Muslim hate crimes in counties with high Twitter usage on days Trump tweets about Muslims. To simplify the interpretation of the coefficients, we standardize all variables to have a mean of zero and standard deviation of one. The specification additionally controls for a vector of control variables $\mathbf{X_{cd}}$ and includes a full set of county and day fixed effects. We also allow for models that include lags of the dependent variable.[24] We cluster standard errors at the state level.

The setup in equation 6 is akin in spirit to a shift-share design, where $Twitter\ Usage$ measures the local exposure to aggregate shocks $Muslim\ Trump\ Tweets$. Because we are interested in estimating the effect of social media, the main concern with this empirical strategy is that the local exposure measure is co-determined with latent factors that may also lead to changes in outcomes when Trump tweets (Goldsmith-Pinkham et al., 2017). Apart from estimating equation 6 using OLS, we thus also present results based on 2SLS, where we again instrument for local Twitter usage using temporal fluctuations in when users started following SXSW around the 2007 festival. The exclusion restriction in this setting is that Trump's tweets about Muslims only affect areas with SXSW followers who joined in March 2007, compared to those who joined before, through its impact on Twitter usage. In support of this, we find that the interaction of Trump's tweets with SXSW followers who joined prior to March does not predict hate crimes.[25]

We first investigate the timing of Trump's tweets with real outcomes in this panel setting. To do so, we include interactions of local Twitter usage with leads and lags of Trump's tweets about Muslims. Figure 12 presents the estimates of this exercise. The graph indicates that we observe differential increases in anti-Muslim hate crime in counties with high Twitter usage one day after Donald Trump's tweets. This is similar to the one we observe in the time series regression. In the online appendix in Table A.33 we report the full set of estimated coefficients from this regressions in OLS and in reduced form.

[Figure 12 about here.]

[24]Estimates of dynamic panel models with fixed effects have an asymptotic bias of order $1/T$ (Nickell, 1981). Because we have a large T (930 days), this bias is likely negligible. Estimating the model with the GMM estimator of Arellano & Bond (1991) is not feasible because the number of moment conditions is of order T^2.

[25]Note that these regressions are highly demanding because hate crimes are relatively rare. In these specifications, less than 1,000 of the close to three million observations are non-zero. The results should thus be interpreted with caution. Nevertheless, we believe they are insightful because they provide an additional plausibility check for the evidence presented above.

Next, we test whether this finding is robust to the inclusion of additional fixed effects and compare the importance of Twitter usage relative to other cross-sectional predictors. In particular we analyze if exposure to Fox News or ideological alignment with Trump (measured by a high Republican vote share) are additional mediating factors.[26]

The results of these exercises can be found in Table 8. Overall the findings are remarkable robust to including interactions with these other cross-sectional exposure variables. The magnitude of the coefficients remains quantitatively unchanged, even when we include state $\times$ day, county $\times$ day of week and county $\times$ day of month fixed effects in columns 1-3. In the following two columns we show that the inclusion of Fox News exposure and the Republican vote share – both of which we interact with Trump's tweets – have less robust and quantitatively smaller predictive power for increases in anti-Muslim hate crime.

[Table 8 about here.]

Overall the findings in this section are again in line with the hypothesis that, when triggered by a shock such as Trump's tweets about Muslims, social media may contribute to anti-Muslim incidents in real-life.

7 Discussion

7.1 Potential Mechanisms

The evidence provided in the previous sections all support the hypothesis that social media began to play a role in the of the expression of anti-Muslim sentiment and the spread of anti-Muslim hate crimes with the 2016 presidential campaign. The existing literature suggests that our findings could be driven by coordination, persuasion or social norms. While all mechanism are likely at play to some extent in our setting, some findings are more consistent with a role for social norms.

To begin, our findings are unlikely to be driven by lower coordination costs due to social media. The main reason is that neither the 2016 presidential campaign period nor Trump's tweets sharply improved the coordination capabilities of perpetrators of anti-Muslim hate crimes. Further, because most content on Twitter is entirely public, one would not expect it to be the most likely place for plotting anti-Muslim attacks but rather a place to spread ideas.

Another hypothesis is that our findings are driven by the persuasiveness of Twitter content, and Trump's tweets in particular (see DellaVigna & Gentzkow, 2010, for a review

[26] Note that we focus on additional cross-sectional exposure variables because we are interested in the effect of social media per se. As we show above, measures of anti-Muslim sentiment (e.g. Fox News reports) are at least partially *outcomes* of Trump's tweets.

of the literature on persuasion). The short-lived spikes in anti-Muslim hate crime we are observing in the time series are perhaps most in line with a persuasion story. But while persuasion can explain some of our findings, there are some pieces of evidence that are not easily rationalized in a belief-based persuasion model. First, in most persuasion models, the updating of beliefs depends on the credibility of the receiver as well as the information provided (Kamenica & Gentzkow, 2011). However, Trump's tweets for the most part do not contain hard information. This makes it less likely that people are persuaded to commit hate crimes against Muslims compared to the possibility that Trump's tweets trigger people with existing anti-Muslim biases. Second, belief-based models of persuasion would suggest that people with weaker priors adjust their attitudes more strongly. In contrast, we find that the effects of Twitter usage are driven by areas with *higher* pre-existing prejudice. This is also in line with existing evidence of media persuasion: in the case of Nazi radio propaganda, Adena et al. (2015) show that it predominantly activated existing sentiments (also see Voigtlander & Voth, 2012). Third, most persuasion models would predict increases in *average* anti-Muslim hostility. Panel survey evidence in Hopkins & Washington (2019), however, suggests that white Americans' anti-minority prejudice, if anything, declined after Trump's political rise.

We also provide some additional evidence that is difficult to square with the idea that social media affects violence by making people more xenophobic, at least in our setting. Table A.22 reports the results from regressions of the type in 2, where the dependent variable is now the change in a measure of implicit bias against Muslims around Trump's presidential campaign start. This measure is based on mean scores on implicit association tests (IAT) from Project Implicit, which are based on the difference in an individual's ability to assign positive or negative words to Muslims or other people.[27]

We consider a range of specifications and sub-samples, including test scores restricted to whites or conservative, and find no evidence of an increase in implicit bias. In fact, both the time series mean and the estimates based on SXSW suggest that, if anything, people became *less* biased towards Muslims between 2000 and 2017. The estimates suggest that we can reject even small increases in implicit bias due to social media. The weak IV confidence set for the baseline estimate in column 1 is bounded at 0.03, which suggests we can likely rule out that a one standard deviation increase in Twitter usage increases implicit bias by more than 17% of a standard deviation.[28] This conclusion is also supported by the pattern of the

[27]We follow Chetty et al. (2018) and calculate mean IAT scores on the county-level. Participation in the IAT is online and largely voluntary, which may give rise to selection biases. While we cannot fully rule out such biases, we also consider a measure of implicit bias based on individuals who were obligated to take these tests, e.g. as part of a work program, and find similar results.

[28]To see this, consider that the standard deviation of *Log(Twitter usage)* in this sample is around 1.80. The standard deviation of the change in IAT scores is 0.313. That means the largest effect of a one standard deviation increase in social media usage in the confidence set is $(0.03 \times 1.80)/0.313 \approx 0.17$. In other words,

event study in Figure A.10.

A perceived shift in social norms among people who already harbor extreme viewpoints may be an alternative mechanism to explain why we observe an effect of social media on hate crime and expressed xenophobia, but no effect on implicit biases. The channel we have in mind is the following. A key feature of social norms is that they are based on people's *perceptions* of everyone else's beliefs. These perceptions, in turn, are shaped by the "sample" of beliefs that are most salient to an individual (e.g. Bursztyn & Jensen, 2015; Perez-Truglia & Cruces, 2017; Enikolopov et al., 2017). But the people are systematically wrong in their perception of what others believe, particularly when it comes to political topics (e.g. Westfall et al., 2015; Bordalo et al., 2016).[29]

By enabling relatively few but particularly visible individuals to affect the aggregate discourse, social media could shift beliefs about what is socially acceptable and make people more susceptible to extreme viewpoints. Such effects could be re-enforced by what has often been called "echo chambers" (e.g. Bessi et al., 2015; Del Vicario et al., 2016; Schmidt et al., 2017; Sunstein, 2017). This, in turn, could affect the willingness of a small set of potential perpetrators to take hateful actions online or offline.[30]

This interpretation is in line with the findings of Bursztyn et al. (2017), who show in a range of experiments that Donald Trump's 2016 election victory increased people's willingness to publicly express xenophobic views, as well as the tolerance towards such views. While our setting does not allow for a controlled experiment, our findings suggest that social media could contribute to such an unraveling of social norms.[31]

1% higher social media usage is unlikely to increase implicit bias against Muslims by more than 0.17%.

[29]See Bénabou (2008) for a model of how belief distortions can give rise to a persistence of ideologies in equilibrium; Bénabou (2013) studies "groupthink" more broadly. False beliefs can also result in an aggregate misperception, termed "pluralistic ignorance" (see Miller & Prentice, 1994; Kuran, 1995). In Saudi Arabia, for example, most men privately approve of women in the labor force but drastically underestimate approval among their peers (Bursztyn et al., 2018).

[30]This is related to Ali & Bénabou (2016), where the visibility of individuals makes aggregate behavior (*descriptive* norms) less informative about societal preferences (*prescriptive* norms). It is also related to Mukand & Rodrik (2018), where "political entrepreneurs" can change individuals' perception of whom they are, by increasing the salience of particular parts of their identity (e.g. a "true American"). Matz et al. (2017) provide evidence for the effectiveness of social media targeting based on psychological traits.

[31]For theoretical models of social norms see, for example, Bénabou & Tirole (2006), Bénabou & Tirole (2011), Ali & Lin (2013), and Ali & Bénabou (2016). Daughety & Reinganum (2010) study how agents adjust their actions if they are observable by others, which creates a costly social distortion. For empirical evidence on persuasion and social norms, see e.g. Cialdini et al. (2006), Gerber et al. (2008), DellaVigna & Gentzkow (2010), and Dellavigna et al. (2016).

7.2 Reporting Changes in Hate Crimes

A potential concern for interpreting our findings with regard to hate crimes could be reporting bias in the FBI data. We believe it is highly unlikely that our findings are solely driven by changes in the reporting rather than actual incidents of hate crimes.

First, our cross-sectional empirical strategy makes the most obvious types of reporting changes unlikely. We focus on within-county changes of hate crime after taking out state-level averages. This rules out any persistent differences in the propensity to report hate crimes, as well as dynamic changes across states. In our instrumental variable estimation, we exploit variation in the locations of SXSW followers who joined in March 2007, compared to those of SXSW followers from previous months. It is not clear why changes in reporting, without changes in actual hate crime incidents, would exhibit this particular correlation with early Twitter adoption. To the best of our knowledge, social media activity is not a major input in the two-tier process for the identification of hate crimes by the FBI.

Second, the heterogeneous patterns we find in the data are inconsistent with those one would expect for changes in hate crime reporting. The cross-sectional results are entirely driven by one crime category, assault. If social media only increased reporting, we would expect to see more reports on hate crimes of lower significance, such as minor cases of vandalism, which is not the case in the data. Reporting also does not explain why there should be larger effects in counties with pre-existing hate groups. If anything, one would expect reporting changes with the start of Trump's presidential run to be concentrated in more liberal counties. Further, Hobbs & Lajevardi (2019) find that the 2016 presidential election was associated with a partial withdrawal of Muslims from public life. In that case, changes in reporting would further bias our estimates downwards.

Third, the precise timing in our time series results speaks against reporting changes. While people might report more hate crimes after Trump's negative tweets about Muslims, they should also become more likely to report *past* hate crimes. This would lead to a very different time series pattern: increases in reporting should should translate into a larger number of hate crimes not only after but also *before* Trump's tweets. However, the data only shows a spike *after* the tweets. It also seems unlikely that the time series findings are driven by changes in the way the FBI classifies hate crimes, because the incident date rarely corresponds to the date a hate crime is reviewed by the FBI as part of the two-tier process. If Trump's tweets change the behavior of FBI analysts, this would again lead to increases in hate crimes before Trump's tweets, which we do not observe in the data.

Taken together, we believe our evidence to be more in line with changes in the actual number of hate crimes. This is also consistent with evidence using the alternative data from the Anti-Defamation League we use in robustness exercises.

8 Conclusion

Social media has recently come under scrutiny for its oft-alleged potential to increase citizen polarization by creating informational "echo chambers" (Sunstein, 2009, 2017). Yet, the empirical evidence on this question is limited and has led to widely differing conclusions (Boxell et al., 2017). Consistent with evidence that social media can motivate real-life action (Enikolopov et al., 2016; Müller & Schwarz, 2018), we find a tight link between Twitter usage, Donald Trump's tweets about Muslims, and different measures of anti-minority sentiment.

Using an instrumental variable strategy, we attempt to identify the causal effect of social media on anti-Muslim sentiment around the time that then-candidate Trump launched his campaign. We exploit the unique history of the diffusion of Twitter prompted by the service's surge in popularity at the SXSW conference in March 2007. This fact allows us to instrument for social media usage today using the locations of Twitter's early adopters while holding constant the locations of people following SXSW prior to the 2007 event or other events similar to SXSW. By identifying the effect through the time dimension, this approach allows us to abstract from endogenous selection into Twitter penetration under relatively mild identifying assumptions.

Our findings are consistent with a role for social media in the normalization of anti-minority sentiments. In line with this hypothesis, we find that Trump's tweets about Muslims are highly correlated with the number of anti-Muslim hate crimes, but only for the time period after the start of his presidential campaign. This correlation also persists using an instrumental variable strategy that leverages the fact that Trump tweets more about Muslims on days when he golfs. This is at least suggestive of the idea that social media, and Trump's tweets in particular, may contribute to a climate that reduces the social sanctions against and increases the incidence of hate crimes.

While this paper focused on particularly negative outcomes – hate crimes targeting minorities and other measures that broadly reflect xenophobia – social media may well have a positive impact in other areas. We would also like to caution against using our findings as a basis for policies directed at restricting online communication. History is ripe with cautionary tales of how excessive state power over the media can abet or enable authoritarian rule. The complex trade-offs that policy makers face in this regard thus require nuanced discussion and, above all, more evidence. Notwithstanding, our results suggest that social media can affect offline actions that might endanger minority communities, and should be taken seriously.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of The Nazis in Prewar Germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Alatas, V., Chandrasekhar, A. G., Mobius, M., Olken, B. A., & Paladines, C. (2019). When Celebrities Speak: A Nationwide Twitter Experiment Promoting Vaccination In Indonesia. Working Paper 25589, National Bureau of Economic Research.

Ali, S. N. & Bénabou, R. (2016). Image versus information: Changing societal norms and optimal privacy. Working Paper 22203, National Bureau of Economic Research.

Ali, S. N. & Lin, C. (2013). Why People Vote: Ethical Motives and Social Incentives. *American Economic Journal: Microeconomics*, *5*(2), 73–98.

Anderson, T. W., Rubin, H., et al. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Andrews, I. (2018). Valid Two-Step Identification-Robust Confidence Sets for GMM. *Review of Economics and Statistics*, *100*(2), 337–348.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak Instruments in IV Regression: Theory and Practice. *Annual Review of Economics*.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*(2), 277–297.

Arrow, K. J. (2000). Increasing Returns: Historiographic Issues and Path Dependence. *The European Journal of the History of Economic Thought*, *7*(2), 171–180.

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-in by Historical Events. *The economic journal*, *99*(394), 116–131.

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press.

Barberá, P. (2014). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US.

Barberá, P. & Rivero, G. (2015). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*, *33*(6), 712–729.

BBC (2019). Ilhan Omar: Muslim Lawmaker Sees Rise in Death Threats After Trump Tweet.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics*, *124*(4), 1497–1540.

Bénabou, R. (2008). Ideology. *Journal of the European Economic Association*, *6*(2-3), 321–352.

Bénabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, *80*(2 (283)), 429–462.

Bénabou, R. & Tirole, J. (2006). Incentives and Prosocial Behavior. *American economic review*, *96*(5), 1652–1678.

Bénabou, R. & Tirole, J. (2011). Laws and Norms. Technical report, National Bureau of Economic Research.

Bessi, A., Zollo, F., Vicario, M. D., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Trend of Narratives in the Age of Misinformation. *PLOS ONE*, *10*(8), 1–16.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband Internet: An Information Superhighway to Sex Crime? *Review of Economic Studies*, *80*(4), 1237–1266.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups. *Proceedings of the National Academy of Sciences of the United States of America*, 201706588.

Brown, B. (2018). The Trump Twitter Archive. http://www.trumptwitterarchive.com/ (accessed November 2nd, 2018).

Bursztyn, L., Egorov, G., & Fiorin, S. (2017). From Extreme to Mainstream: How Social Norms Unravel. Working Paper 23415, National Bureau of Economic Research.

Bursztyn, L., Gonzlez, A. L., & Yanagizawa-Drott, D. (2018). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. NBER Working Papers 24736, National Bureau of Economic Research, Inc.

Bursztyn, L. & Jensen, R. (2015). How does peer pressure affect educational investments? *The Quarterly Journal of Economics*, *130*(3), 1329–1367.

Card, D. & Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics*, *126*(1), 103–143.

Chan, J., Ghose, A., & Seamans, R. (2016). The Internet and Racial Hate Crime: Offline Spillovers from Online Access. *MIS Quarterly*, *40*(2), 381–403.

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2018). Race and economic opportunity in the united states: An intergenerational perspective. Working Paper 24441, National Bureau of Economic Research.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing Social Norms for Persuasive Impact. *Social Influence*, *1*(1), 3–15.

Colella, F., Lalive, R., Sakalli, S. O., & Thoenig, M. (2019). Inference with Arbitrary Clustering. IZA Discussion Papers 12584, Institute of Labor Economics (IZA).

Dahl, G. & DellaVigna, S. (2009). Does Movie Violence Increase Violent Crime? *The Quarterly Journal of Economics*, 677–734.

Daughety, A. F. & Reinganum, J. F. (2010). Public Goods, Social Pressure, and the Choice between Privacy and Publicity. *American Economic Journal: Microeconomics*, *2*(2), 191–221.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The Spreading of Misinformation Online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., & Zhuravskaya, E. (2014). Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia. *American Economic Journal: Applied Economics*, *6*(3), 103–32.

DellaVigna, S. & Gentzkow, M. (2010). Persuasion: Empirical Evidence. *Annual Review of Economics*, *2*(1), 643–669.

Dellavigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). Voting to Tell Others. *The Review of Economic Studies*, *84*(1), 143–181.

Draca, M. & Schwarz, C. (2018). How Polarized Are Citizens? Measuring Ideology from the Ground-up.

Edwards, B. T. (2018). Trump from Reality TV to Twitter, or the Selfie-Determination of Nations. *Arizona Quarterly: A Journal of American Literature, Culture, and Theory*, *74*(3), 25–45.

Enikolopov, R., Makarin, A., & Petrova, M. (2016). Social Media and Protest Participation: Evidence from Russia.

Enikolopov, R., Makarin, A., Petrova, M., & Polishchuk, L. (2017). Social Image, Networks, and Protest Participation. *Universitat Pompeu Fabra*.

FBI (2015). Hate Crime Data Collection Guidelines And Training Manual. *Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program*.

Fiorina, M. P. & Abrams, S. J. (2008). Political Polarization in the American Public. *Annual Review of Political Science*, *11*, 563–588.

Gavazza, A., Nardotto, M., & Valletti, T. M. (2015). Internet and Politics: Evidence from UK Local Elections and Local Government Policies.

Gawker (2007). Twitter Blows Up at SXSW Conference. https://gawker.com/243634/twitter-blows-up-at-sxsw-conference (accessed March 3rd, 2018).

Gentzkow, M. (2016). Polarization in 2016. *Toulouse Network of Information Technology white paper*.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment. *American political Science review*, *102*(1), 33–48.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2017). Bartik Instruments: What, When, Why, and How. *Working Paper*.

Gould, E. D. & Klor, E. F. (2016). The Long-run Effect of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West. *The Economic Journal*, *126*(597), 2064–2114.

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 U.S. Presidential Campaign. *Working Paper*.

Guess, A. M. (2018). (Almost) Everything in Moderation: New Evidence on Americans Online Media Diets. *Working Paper*.

Haje Jan Kamps (2015). Who Are Twitters Verified Users?

Hanes, E. & Machin, S. (2014). Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice*, *30*(3), 247–267.

Haustein, S. & Costas, R. (2014). Determining Twitter Audiences: Geolocation and Number of Followers. *ALM*, *4*, 6.

Hobbs, W. & Lajevardi, N. (2019). Effects of Divisive Political Campaigns on the Day-to-Day Segregation of Arab and Muslim Americans. *American Political Science Review*, *113*(1), 270–276.

Hopkins, D. J. & Washington, S. (2019). The Rise of Trump, the Fall of Prejudice? Tracking White Americans' Racial Attitudes 2008-2018 via a Panel Survey. *Working Paper*.

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research. *Big Data & Society*, *4*(2), 2053951717736336.

Kuran, T. (1995). *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press.

Liebowitz, S. J. & Margolis, S. E. (1999). Path Dependence. *Encyclopedia of law and economics*.

Martin, G. J. & Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, *107*(9), 2565–2599.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological Targeting as an Effective Approach to Digital Mass Persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719.

Miller, C. & Smith, J. (2017). Anti-Islamic Content on Twitter. *Centre for the Analysis of Social Media at Demos*.

Miller, D. T. & Prentice, D. A. (1994). Collective Errors and Errors about the Collective. *Personality and Social Psychology Bulletin*, *20*(5), 541–550.

Mukand, S. & Rodrik, D. (2018). The Political Economy of Ideas: On Ideas Versus Interests in Policymaking. Working Paper 24467, National Bureau of Economic Research.

Müller, K. & Schwarz, C. (2018). Fanning the Flames of Hate: Social Media and Hate Crime. *Working Paper*.

NBC News (2017). Advocates Warn of Possible Underreporting in FBI Hate Crime Data, by Chris Fuchs. https://www.nbcnews.com/news/asian-america/advocates-warn-possible-underreporting-fbi-hate-crime-data-n830711 (accessed March 3rd, 2018).

New York Times (2017). Trump Shares Inflammatory Anti-Muslim Videos, and Britains Leader Condemns Them, By Peter Baker and Eeileen Sullivan.

New York Times (2018). The Man Behind the Presidents Tweets.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, *49*(6), 1417–1426.

NYT (2019). Tracking Trump's Visits to His Branded Properties.

Olea, J. L. M. & Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Panagopoulos, C. (2006). The Polls-Trends: Arab and Muslim Americans and Islam in the aftermath of 9/11. *International Journal of Public Opinion Quarterly*, *70*(4), 608–624.

Perez-Truglia, R. & Cruces, G. (2017). Partisan Interactions: Evidence from a Field Experiment in the United States. *Journal of Political Economy*, *125*(4), 1208–1243.

Petrova, M., Sen, A., & Yildirim, P. (2017). Social Media and Political Donations: New Technology and Incumbency Advantage in the United States. *Working Paper*.

ProPublica (2017). Why America Fails at Gathering Hate Crime Statistics, by Ken Schwencke. https://www.propublica.org/article/why-america-fails-at-gathering-hate-crime-statistics (accessed March 3rd, 2018).

Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Anatomy of News Consumption on Facebook. *Proceedings of the National Academy of Sciences*, *114*(12), 3035–3039.

Stephens-Davidowitz, S. (2014). The Cost of Racial Animus on a Black Candidate: Evidence using Google Search Data. *Journal of Public Economics*, *118*, 26–40.

Stock, J. & Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, (pp. 80–108). New York: Cambridge University Press.

Sun, L. (2018). Implementing Valid Two-Step Identification-Robust Confidence Sets For Linear Instrumental-Variables Models. *The Stata Journal*, *18*(4), 803–825.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.

Sunstein, C. R. (2017). *# Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter Networks. *Social networks*, *34*(1), 73–81.

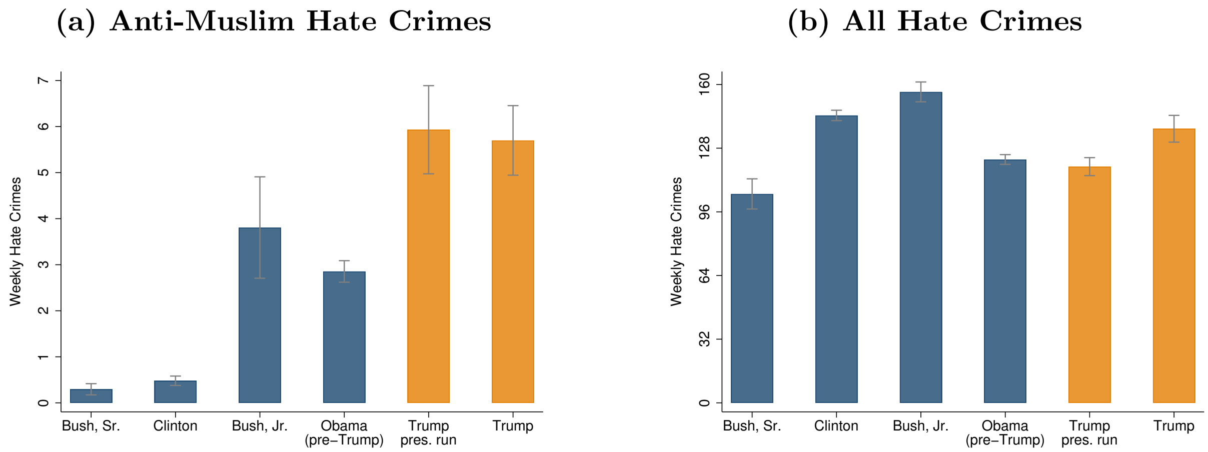
Voigtlander, N. & Voth, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-semitic violence in nazi germany. *The Quarterly Journal of Economics*, *127*(3), 1339–1392.

Westfall, J., Boven, L. V., Chambers, J. R., & Judd, C. M. (2015). Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide. *Perspectives on Psychological Science*, *10*(2), 145–158. PMID: 25910386.

Willnat, L., Weaver, D. H., & Wilhoit, G. C. (2019). The American Journalist in the Digital Age. *Journalism Studies*, *20*(3), 423–441.

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Evidence from the Rwandan Genocide. *The Quarterly Journal of Economics*, *129*(4), 1947–1994.

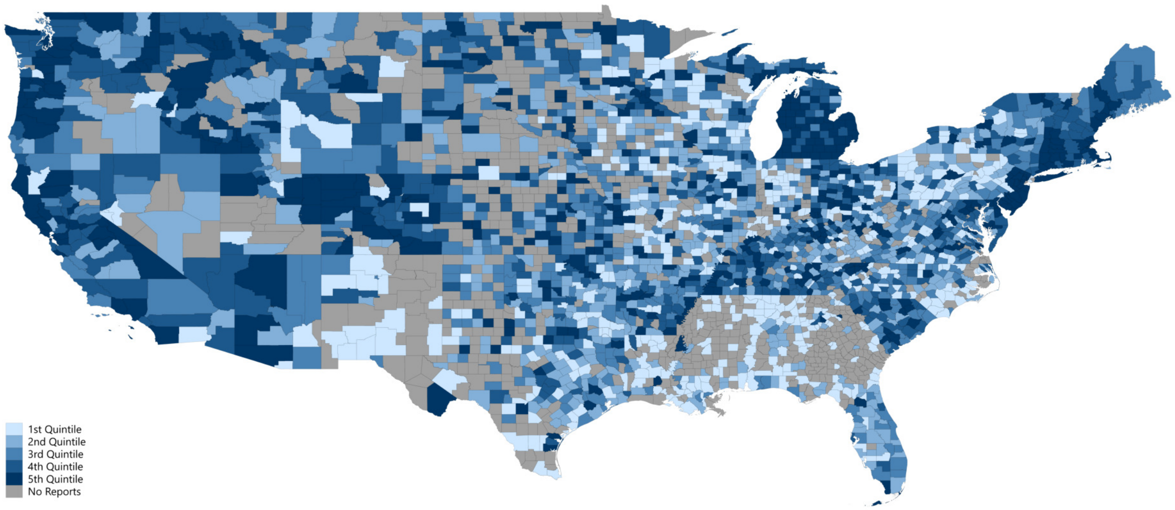
Figure 1: Average Weekly Anti-Muslim Hate Crimes Since 1990, by President



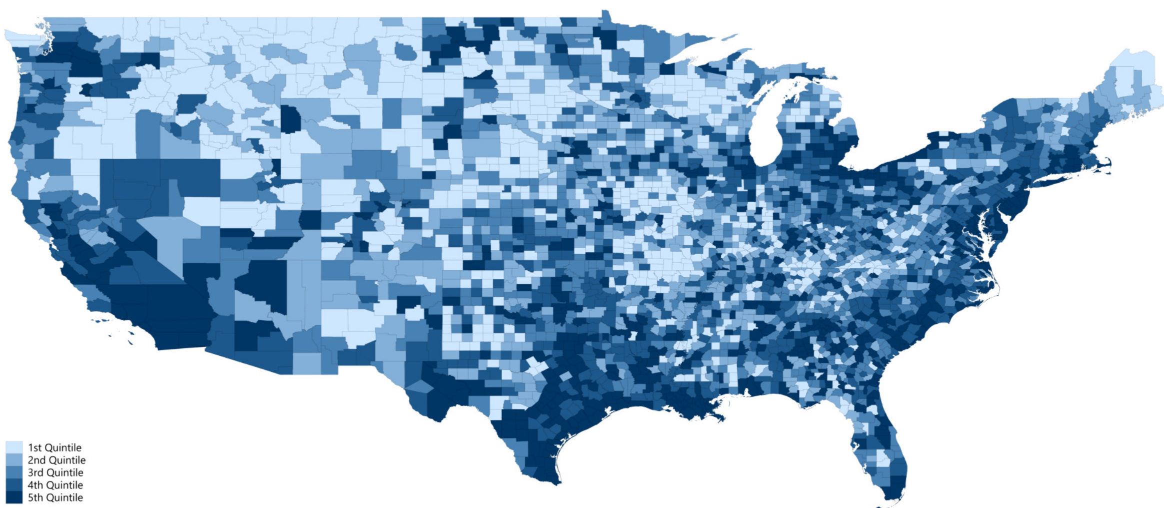
Notes: This figure plots the average weekly number of hate crimes reported by the FBI, by president. We divide Barack Obama's presidency into the period before and after Donald Trump's campaign start ("Obama (pre-Trump)" and "Trump pres. run", respectively). Panel (a) shows the number of anti-Muslim hate crimes. Panel (b) shows the total number of hate crimes. We also plot 95% confidence intervals.

Figure 2: Hate Crimes and Twitter Usage by US County

(a) Hate Crimes per Capita



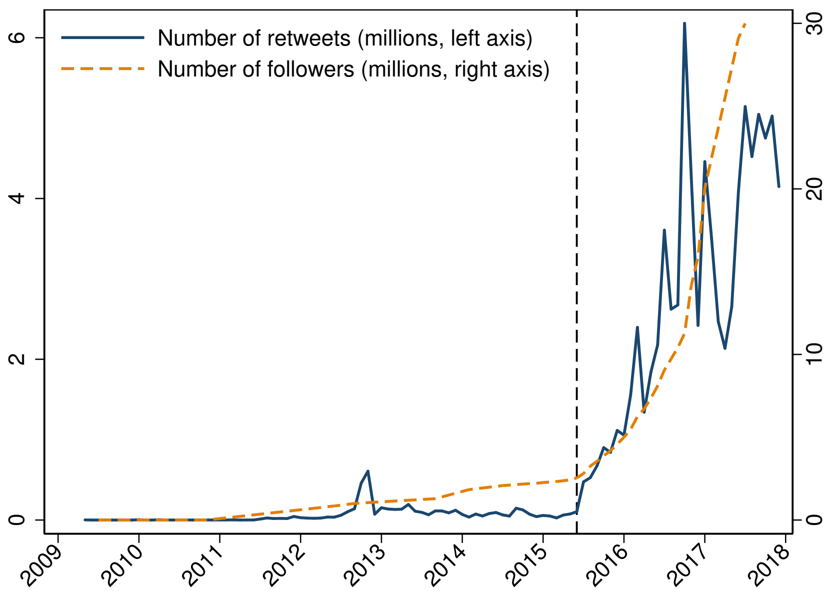
(b) Twitter Usage per Capita



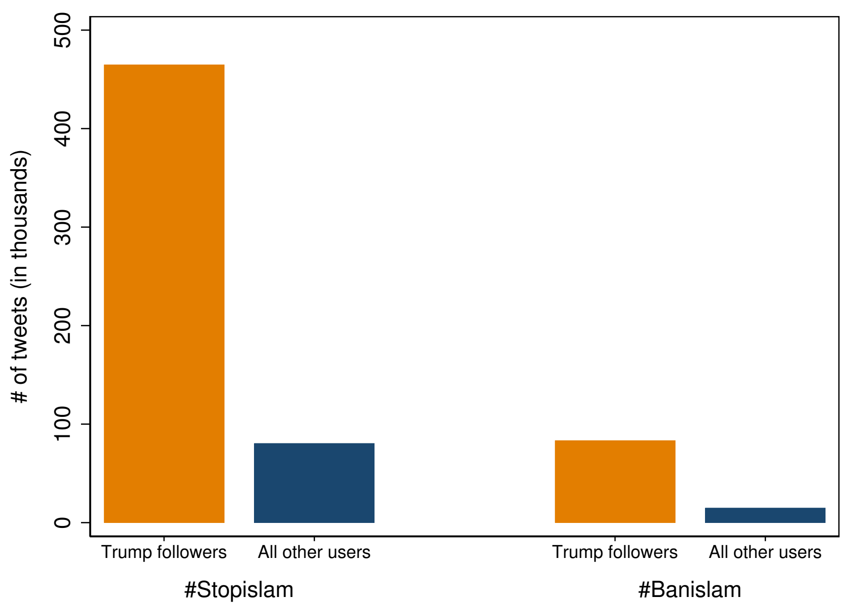
Notes: These maps plot the geographical distribution of the main variables of interest across the counties in the mainland US. Panel (a) plots quintiles of the total number of hate crimes per capita between 1990 and 2017 as reported by the FBI. Counties in grey never reported any hate crime. Panel (b) plots our measure of Twitter usage scaled by population.

Figure 3: Trump's Twitter Reach

(a) Trump's Retweets Over Time



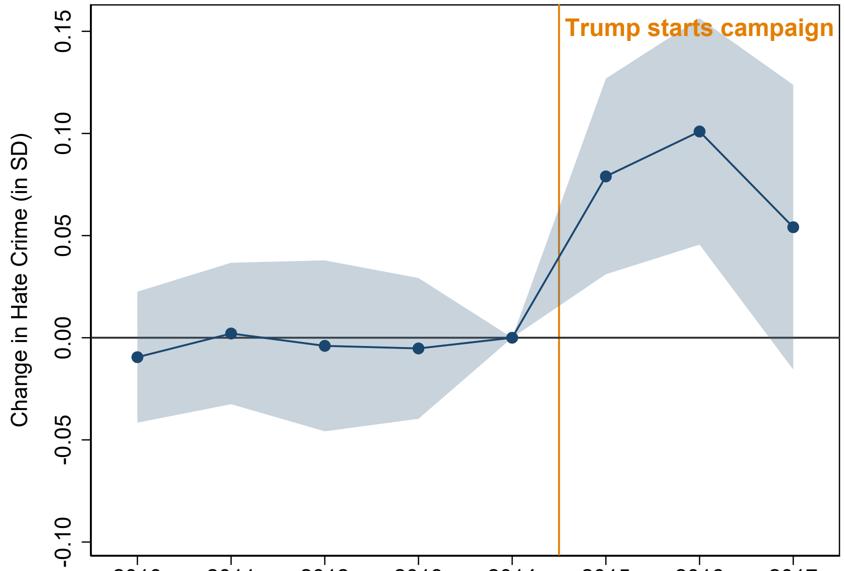
(b) Trump Followers and Anti-Muslim Tweets



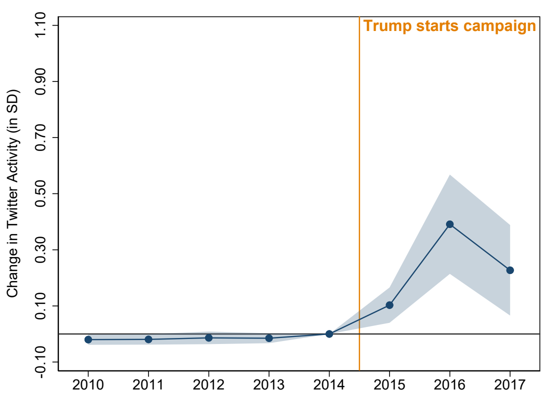
Notes: Panel (a) plots the number of monthly retweets (in millions) Trump's Twitter account received since he joined the site in 2009. The vertical line marks the start of his presidential campaign in June 2015. Panel (b) plots the number of tweets containing the hashtags #StopIslam or #BanIslam sent between 2010 and 2017, which we interpret as clearly expressing negative sentiment towards Muslims. The orange proportion of the bar indicates the number of these tweets posted by followers of Trump's Twitter account.

Figure 4: Twitter Usage and the Increase in Anti-Muslim Sentiments

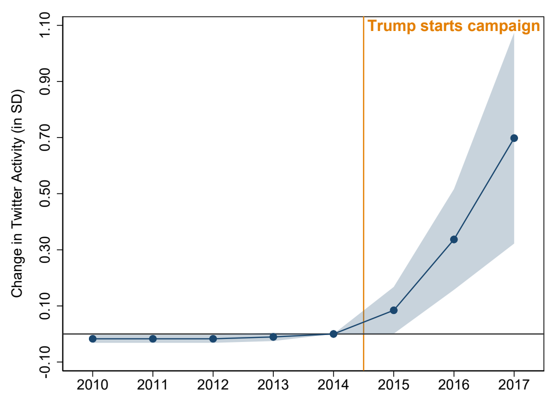
(a) Anti-Muslim Hate Crimes



(b) Tweets Containing #StopIslam



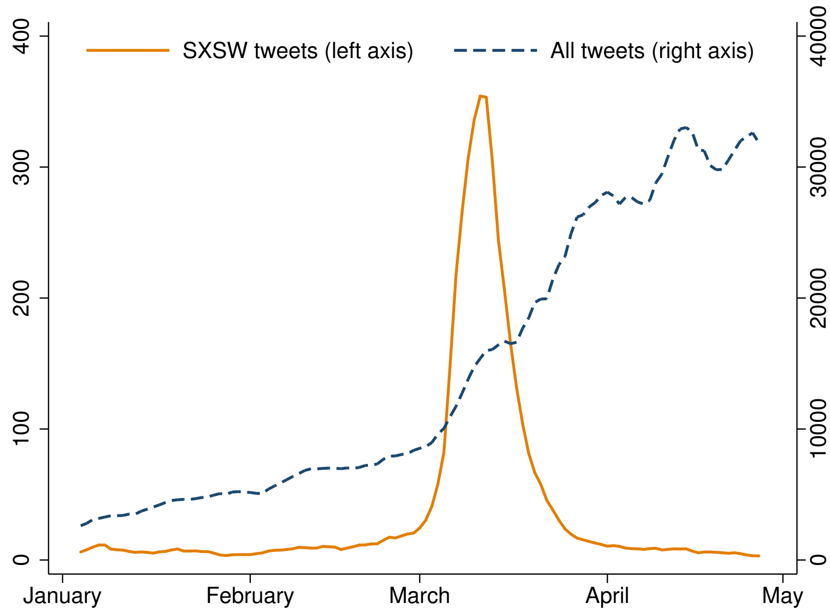
(c) Tweets Containing #BansIslam



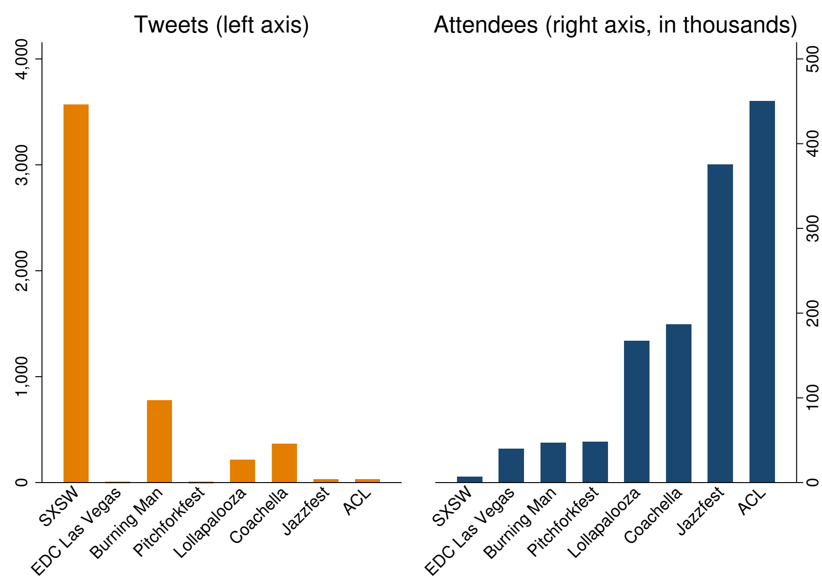
Notes: These figures plot the coefficients from running event study regressions as in Equation (1). The dependent variables are the natural logarithm of anti-Muslim hate crimes in panel (a) and the number of posts containing #StopIslam and #BanIslam in panels (b) and (c). We standardized the variables to have a mean of zero and standard deviation of one. The omitted category is the year leading up to Trump's presidential run. The vertical line indicates the approximate start of Trump's presidential campaign in June 2015. The shaded area indicates 95% confidence intervals.

Figure 5: South by Southwest (SXSW) 2007 and the Spread of Twitter

(a) Twitter Activity Around SXSW 2007

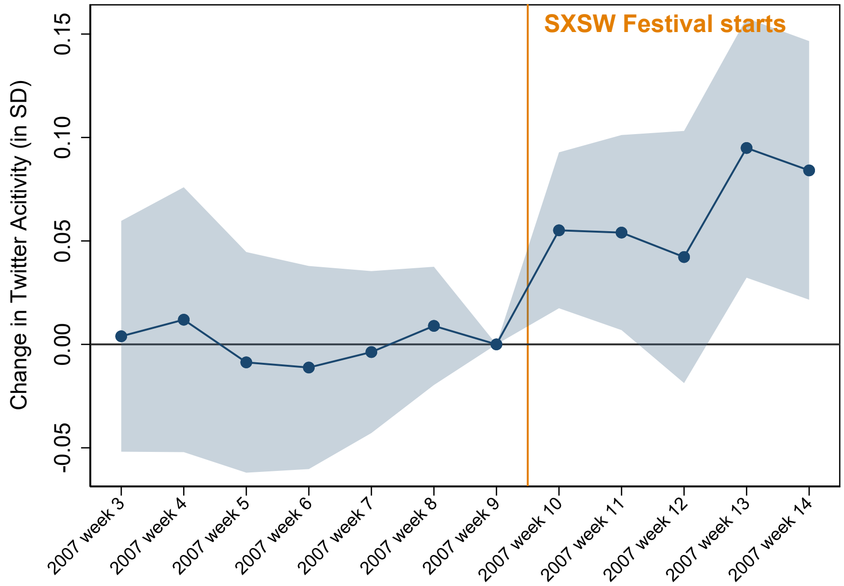


(b) Major Festivals in 2007: Tweets and Attendance



Notes: Panel (a) plot the total number of tweets and the number of tweets containing the term SXSW over time, smoothed using a 7-day moving average. The number of tweets on a given day is based on the 100 most common English words (see Table A.8). Panel (b) plots the number of tweets mentioning major festivals in 2007 in a 14 day window before and after the event.

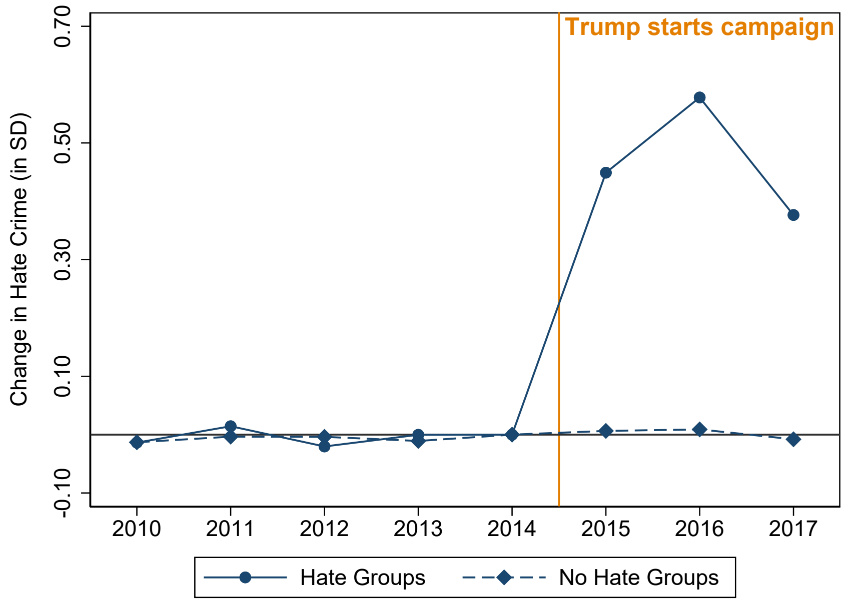
Figure 6: The Effect of SXSW on Twitter Adoption



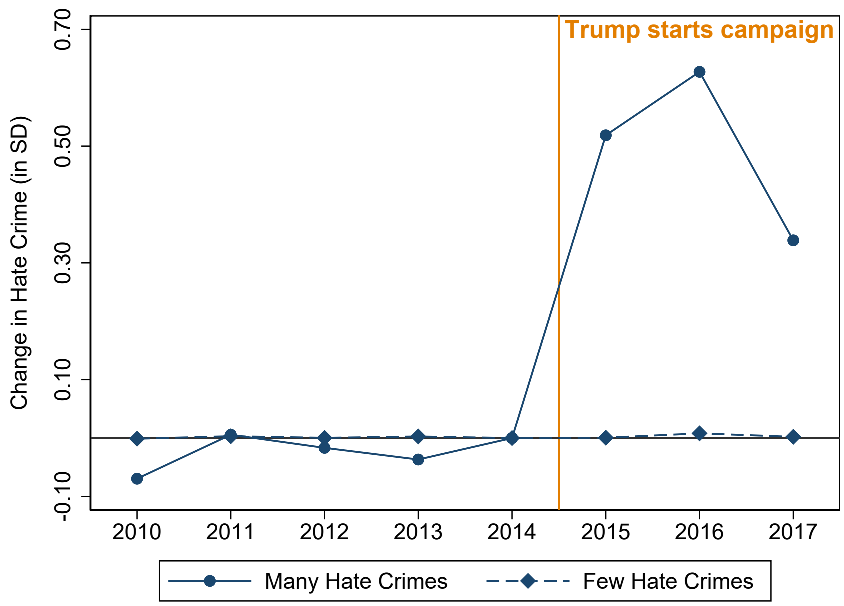
Notes: This figure plots the estimates of β_τ from the panel event study regression $Log(1 + \#\, of\ tweets) = \sum \beta_\tau SXSW\ followers, March\ 2007_c \times Week_\tau + \sum \delta_\tau SXSW\ followers,\ Pre_c \times Week_\tau + County\ FE + Week\ FE + \varepsilon_{cw}$. The number of tweets in a given county and week is based on the 100 most common English words. We standardize the variables to have a mean of zero and standard deviation of one. Standard errors are clustered by state.

Figure 7: Heterogenous Effects of Twitter Usage

(a) Split by Existing SPLC Hate Groups Share

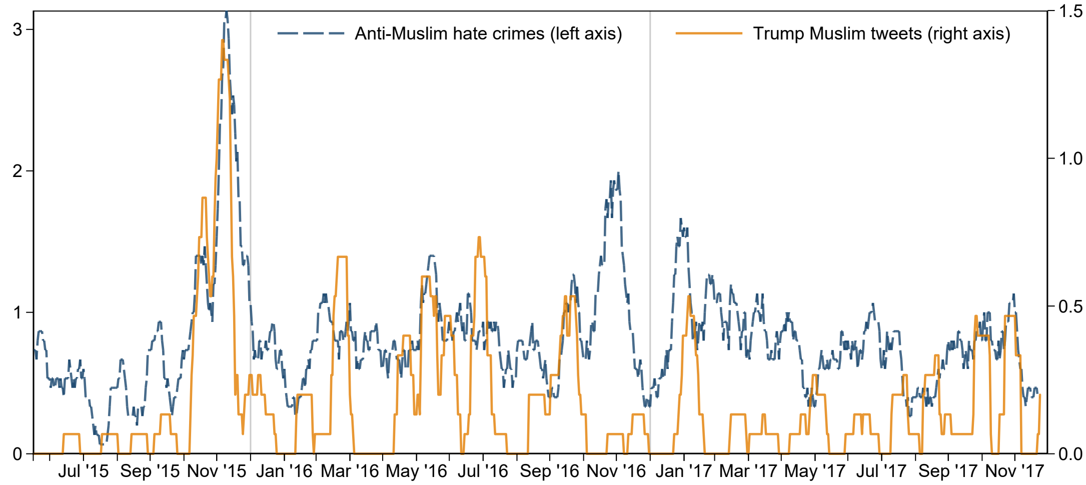


(b) Split by Frequency of Hate Crimes in Pre-Period



Notes: These figure plot the coefficients of running panel event study regressions as in Equation (1). We again standardized the variables to have a mean of zero and standard deviation of one. Equation (1) is estimated separately for counties with and without at least one hate group as defined by the Southern Poverty Law Center (SPLC). In panel (b) we split counties at the 90th percentile of the number of hate crimes per capita in the pre-period.

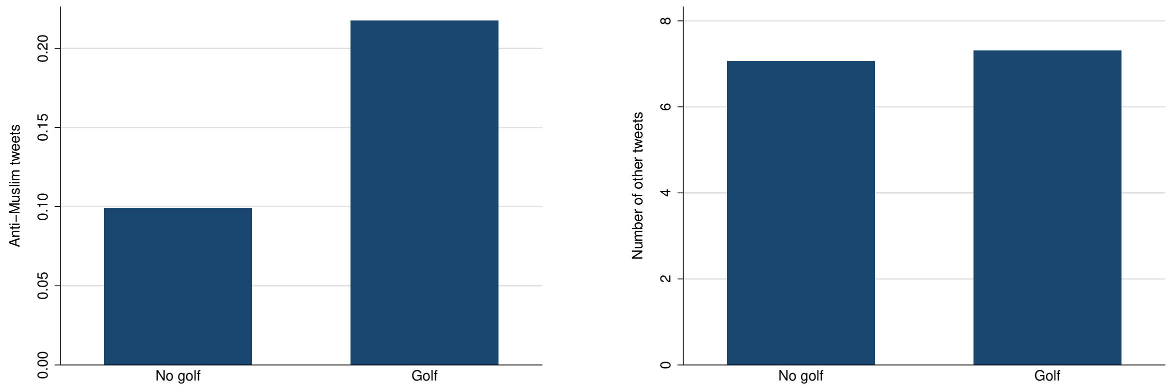
Figure 8: Trump's Tweets About Muslims and Anti-Muslim Hate Crime



Notes: This figure plots the daily time series of anti-Muslim hate crime and Trump's tweets about Muslims, smoothed using a 14-day moving average. The time period covers the start of Trump's presidential campaign in June 2015 until the end of 2017.

Figure 9: Trump's Twitter Activity, Split by Golf Days

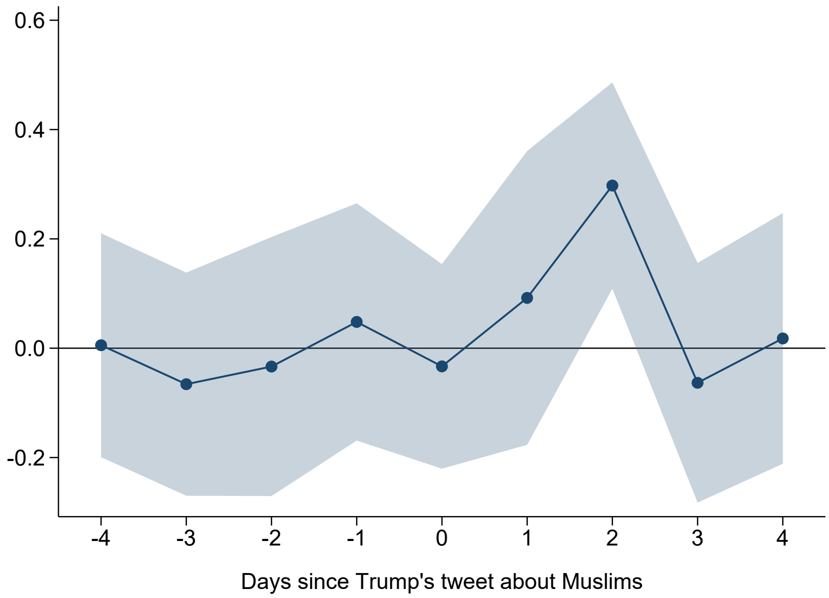
(a) Tweets about Muslims **(b) Total tweets**



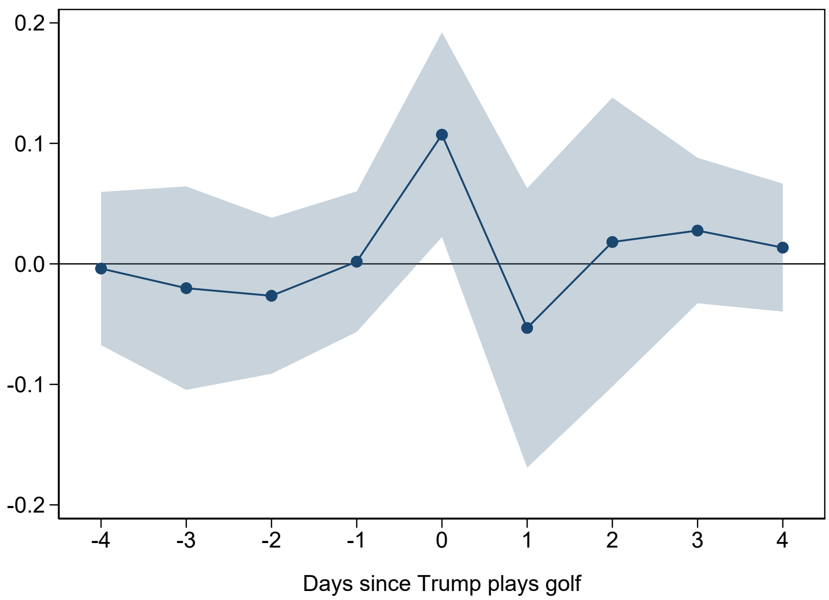
Notes: These figures plot the average daily number of Trump's tweets, split by whether he plays golf on a given day in 2017. Panel (a) reports the average number of tweets about Muslims, panel (b) reports the total number of tweets.

Figure 10: Time Series Correlations

(a) OLS - Trump Tweets about Muslims and Hate Crimes



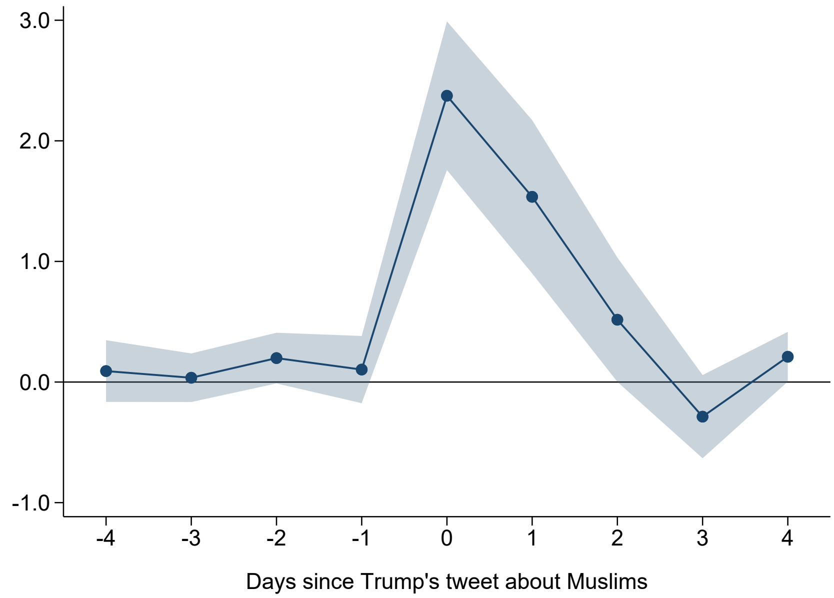
(b) First Stage - Golf and Trump Tweets about Muslims



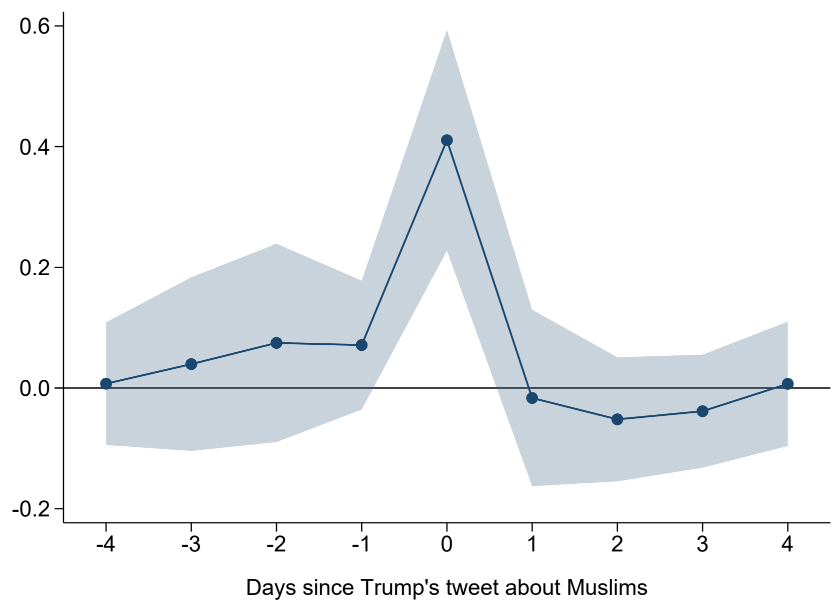
Notes: These figures plot the dynamic correlations for equations 4 and 5 for values of h ranging between -4 and 4. Panel (a) plots the correlation of Trump's tweets about Islam-related topics and anti-Muslim hate crimes (both in natural logarithm). Panel (b) plots the correlation of Trump's golf outings with the log number of his Islam-related tweets. T indicates the date of tweets about Muslims or golfing ($h = 0$). All regressions include time trends; a full set of day of week and year-month dummies; and four lags of dummies for the incidence of terror attacks in the US, Europe, and the rest of the world. The sample is 2017. The shaded areas are 95% confidence intervals based on Newey-West standard errors.

Figure 11: Spillovers of Trump's Tweets to His Followers

(a) Retweets of Trump's Tweets

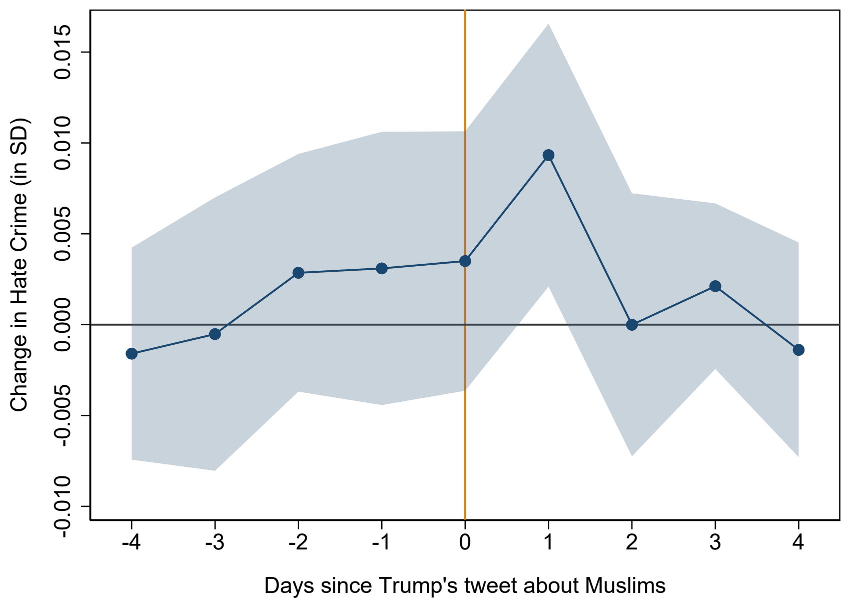


(b) New tweets about Muslims



Notes: These figures plot the dynamic correlations for equations 4 and 5 for values of h ranging between -4 and 4. Panel (a) plots the correlation of Trump's tweets about Islam-related topics and the retweets this tweets by Trump's followers (both in natural logarithm). Panel (b) plots the correlation of Trump's tweets about Islam-related topics and the self-produced anti-Muslim tweets by Trump's followers. T indicates the date of tweets about Muslims ($h = 0$). All regressions include a full set of day of week and year-month dummies; and four lags of dummies for the incidence of terror attacks in the US, Europe, and the rest of the world. The sample is 2017. The shaded areas are 95% confidence intervals based on Newey-West standard errors.

Figure 12: Panel Event Study – Trump Tweets, Twitter Usage, and Hate Crimes



Notes: These figures plot the dynamic correlations for equation 6 time periods ranging between -4 and 4 days around Trump's tweets in counties with high Twitter usage. The dependent variable is the log number of anti-Muslim hate crimes in county c on day d, which we standardized to have a mean of zero and standard deviation of one. T indicates the date of tweets about Muslims ($h = 0$). All regressions include population controls and county times month, day and county times day of month fixed effects. The shaded areas are 95% confidence intervals based on standard errors clustered at the state level.

Table 1: First Stage - South by Southwest 2007 and the Diffusion of Twitter Usage

	Log(Twitter usage)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(SXSW followers, March 2007)	0.505***	0.461***	0.440***	0.407***	0.403***	0.394***	0.371***	0.362***
	(0.061)	(0.061)	(0.064)	(0.054)	(0.052)	(0.053)	(0.056)	(0.056)
Log(SXSW followers, Pre)	0.153*	0.162*	0.120	0.112	0.104	0.102	0.090	0.086
	(0.077)	(0.091)	(0.089)	(0.084)	(0.083)	(0.081)	(0.081)	(0.077)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race and religion controls				Yes	Yes	Yes	Yes	Yes
Socioeconomic controls					Yes	Yes	Yes	Yes
Media controls						Yes	Yes	Yes
Election control							Yes	Yes
Crime controls								Yes
Geographical controls			Yes	Yes	Yes	Yes	Yes	Yes
Observations	3107	3107	3107	3107	3106	3105	3105	3105
Mean of DV	10	10	10	10	10	10	10	10
p-value: March 2007 = Pre	0.01	0.04	0.03	0.02	0.01	0.01	0.02	0.02

Notes: This table presents county-level regressions where the dependent variable is the number of tweets sent (in natural logarithm). *SXSW followers, March 2007* is the number of Twitter users who joined in March 2007 and follow South by Southwest (SXSW) *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. The bottom row reports p-values from F-tests for the equality of these coefficients. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Race and religion controls contains the share of people identifying as white, African American, Native American or Pacific Islander, Asian, Hispanic, or Muslim. Socioeconomic controls include the poverty rate, unemployment rate, local GINI index, the share of uninsured individuals, log median household income, the share of highschool graduates, the share of people with a graduate degree, as well as the employment shares in agriculture, information technology, manufacturing, nontradables, construction and real estate, utilities, business services, or other sectors. Media controls include the viewership share of Fox News, the cable TV spending to population ratio, and the prime time TV viewership to population ratio. Election control is the county-level vote share of the Republican party in 2012. Crime controls are the rates of violent or property crime from the FBI. Geographical controls include the linear distance from the SXSW festival location (Austin, Texas), population density, and the natural logarithm of county size. Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Reduced Form - South by Southwest 2007 and the Rise in Hate Crimes against Muslims

	ΔLog(Hate crimes against Muslims)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(SXSW followers, March 2007)	0.075**	0.074**	0.082***	0.075**	0.072**	0.072**	0.072**	0.072**
	(0.030)	(0.030)	(0.029)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
Log(SXSW followers, Pre)	0.033	0.034	0.050	0.025	0.025	0.026	0.026	0.027
	(0.054)	(0.054)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race and religion controls				Yes	Yes	Yes	Yes	Yes
Socioeconomic controls					Yes	Yes	Yes	Yes
Media controls						Yes	Yes	Yes
Election control							Yes	Yes
Crime controls								Yes
Geographical controls			Yes	Yes	Yes	Yes	Yes	Yes
Observations	3107	3107	3107	3107	3106	3105	3105	3105
Mean of DV	.019	.019	.019	.019	.019	.019	.019	.019

Notes: This table presents county-level regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *SXSW tweets* are the number of newly registered users in the indicated months of 2007 that tweeted about the South by Southwest (SXSW) festival. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Race and religion controls contains the share of people identifying as white, African American, Native American or Pacific Islander, Asian, Hispanic, or Muslim. Socioeconomic controls include the poverty rate, unemployment rate, local GINI index, the share of uninsured individuals, log median household income, the share of highschool graduates, the share of people with a graduate degree, as well as the employment shares in agriculture, information technology, manufacturing, nontradables, construction and real estate, utilities, business services, or other sectors. Media controls include the viewership share of Fox News, the cable TV spending to population ratio, and the prime time TV viewership to population ratio. Election control is the county-level vote share of the Republican party in 2012. Crime controls are the rates of violent or property crime from the FBI. Geographical controls include the linear distance from the SXSW festival location (Austin, Texas), population density, and the natural logarithm of county size. Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: 2SLS - Social Media and the Rise in Hate Crimes against Muslims

	ΔLog(Hate crimes against Muslims)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS - Hate crimes against Muslims								
Log(Twitter usage)	0.021***	0.019***	0.019***	0.015***	0.015***	0.016***	0.015***	0.015***
	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Panel B: 2SLS - Hate crimes against Muslims								
Log(Twitter usage)	0.148**	0.161**	0.187**	0.185**	0.178**	0.183**	0.194**	0.199**
	(0.064)	(0.069)	(0.075)	(0.082)	(0.080)	(0.083)	(0.091)	(0.093)
Weak IV 95% AR confidence set	[0.04; 0.27]	[0.04; 0.30]	[0.06; 0.35]	[0.04; 0.35]	[0.04; 0.34]	[0.04; 0.35]	[0.04; 0.39]	[0.04; 0.40]
Log(SXSW followers, Pre)	0.010	0.008	0.027	0.005	0.007	0.008	0.009	0.010
	(0.065)	(0.069)	(0.065)	(0.064)	(0.062)	(0.062)	(0.062)	(0.061)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race and religion controls				Yes	Yes	Yes	Yes	Yes
Socioeconomic controls					Yes	Yes	Yes	Yes
Media controls						Yes	Yes	Yes
Election control							Yes	Yes
Crime controls								Yes
Geographical controls			Yes	Yes	Yes	Yes	Yes	Yes
Observations	3107	3107	3107	3107	3106	3105	3105	3105
Mean of DV	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
Robust F-stat.	68.03	58.04	46.96	56.25	61.27	55.30	43.89	41.82

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Race and religion controls contains the share of people identifying as white, African American, Native American or Pacific Islander, Asian, Hispanic, or Muslim. Socioeconomic controls include the poverty rate, unemployment rate, local GINI index, the share of uninsured individuals, log median household income, the share of highschool graduates, the share of people with a graduate degree, as well as the employment shares in agriculture, information technology, manufacturing, nontradables, construction and real estate, utilities, business services, or other sectors. Media controls include the viewership share of Fox News, the cable TV spending to population ratio, and the prime time TV viewership to population ratio. Election control is the county-level vote share of the Republican party in 2012. Crime controls are the rates of violent or property crime from the FBI. Geographical controls include the linear distance from the SXSW festival location (Austin, Texas), population density, and the natural logarithm of county size. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Further Robustness - Social Media and the Rise in Hate Crimes against Muslims

	No state FE (1)	Change since 1990 (2)	Log hate crimes (3)	Drop zero change counties (4)	Drop potentially nonreporting counties (5)	Drop counties with few Muslims (6)	Population weights (7)	Only neighbouring counties (8)	Index dependent variable (9)
Panel A: OLS - Hate crimes against Muslims									
Log(Twitter usage)	0.012*	0.028***	0.067***	0.042	0.028***	0.057**	0.087**	0.044***	0.031**
	(0.006)	(0.008)	(0.017)	(0.032)	(0.009)	(0.024)	(0.040)	(0.013)	(0.015)
Panel B: Reduced form - Hate crimes against Muslims									
Log(SXSW followers, March 2007)	0.072**	0.125***	0.224***	0.112**	0.071**	0.080**	0.113***	0.079**	0.172**
	(0.030)	(0.031)	(0.045)	(0.047)	(0.032)	(0.034)	(0.038)	(0.033)	(0.072)
Panel C: 2SLS - Hate crimes against Muslims									
Log(Twitter usage)	0.154**	0.271***	0.487***	0.234**	0.142**	0.173**	0.210***	0.181**	0.373**
	(0.065)	(0.069)	(0.104)	(0.103)	(0.065)	(0.075)	(0.068)	(0.084)	(0.169)
Weak IV 95% AR confidence set	[0.03; 0.28]	[0.15; 0.41]	[0.30; 0.69]	[0.06; 0.43]	[0.02; 0.27]	[0.03; 0.32]	[0.08; 0.34]	[0.04; 0.36]	[0.09; 0.71]
Log(SXSW followers, Pre)	0.019	-0.021	0.051	0.019	0.032	0.041	-0.020	0.010	-0.066
	(0.066)	(0.071)	(0.117)	(0.089)	(0.064)	(0.070)	(0.058)	(0.074)	(0.156)
Observations	3108	3107	3107	381	2319	586	3107	1167	3107
Mean of DV	0.019	0.025	0.052	0.153	0.026	0.082	0.155	0.040	0.029
Robust F-stat.	80.40	58.04	58.04	64.79	80.13	61.35	44.91	37.76	58.04

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017 in all columns except columns 2 and 3. In column 2, the dependent variable is the log change between 1990 and 2017; in column 3, it is the log number of hate crimes against Muslims in a county after the start of Donald Trump's presidential run on June 16, 2015. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. All regressions control for population deciles, state fixed effects (except in column 1), and demographic controls that include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Column 4 drops all counties for which the change in hate crimes between 2010 and 2017 was zero. Column 5 drops all counties which never report a hate crime between 1990 and 2017. Column 6 drops all counties for which the (rounded) share of Muslims in the county population is zero according to Census data. Column 7 estimates all regressions using weighted least squares (WLS) with population weights. Column 8 only keeps neighbouring counties that differ in whether they have SXSW followers in March 2007 or not. Column 9 recodes the dependent variable into an index equal to 1 for increases in hate crimes, -1 for decreases, and 0 for no change. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Social Media and Other Hate Crimes

	FBI Data						ADL Data
	Total (1)	Hispanic (2)	Other ethnic (3)	Race (4)	Sexual Orientation (5)	Religion (excl. Muslims) (6)	Total (Levels) (7)
Panel A: OLS - Hate crimes							
Log(Twitter usage)	0.006	-0.000	-0.018***	0.005	-0.007	0.017*	0.129***
	(0.012)	(0.008)	(0.007)	(0.008)	(0.006)	(0.009)	(0.034)
Panel B: Reduced form - Hate crimes							
Log(SXSW followers, March 2007)	0.085**	0.079**	0.007	0.055	0.046	0.058	0.357***
	(0.042)	(0.034)	(0.033)	(0.048)	(0.043)	(0.041)	(0.110)
Panel C: 2SLS - Hate crimes							
Log(Twitter usage)	0.184*	0.171**	0.014	0.119	0.099	0.125	0.775***
	(0.100)	(0.068)	(0.071)	(0.109)	(0.096)	(0.084)	(0.192)
Weak IV 95% AR confidence set	[0.01; 0.40]	[0 .0; 0.29]	[0.10; 0.16]	[0.06; 0.34]	[0.07; 0.29]	[0.04; 0.27]	[0.38; 1.13]
Log(SXSW followers, Pre)	-0.052	-0.074	-0.039	-0.035	-0.025	-0.036	0.055
	(0.078)	(0.071)	(0.074)	(0.081)	(0.082)	(0.064)	(0.177)
Observations	3107	3107	3107	3107	3107	3107	3107
Mean of DV	-0.015	-0.012	-0.016	-0.011	-0.025	0.005	0.226
Robust F-stat.	58.04	58.04	58.04	58.04	58.04	58.04	58.04

Notes: This table presents county-level OLS, reduced form, and IV regressions where the dependent variable is the log change in hate crimes against the group in the top row between 2010 and 2017. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. The hate crime data from the Anti-Defamation League (ADL) is sparse prior to 2016, so we use the log-level of hate crimes in column 7. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Trump Tweets and Anti-Muslim Hate Crimes

	Baseline (1)	Add lagged dependent variable (2)	Add federal holiday control (3)	Add Google search control (4)	Add TV coverage control (5)	Add terror attack control (6)	Add total tweets control (7)
Panel A: OLS - Log(Hate crimes against Muslims) in t+2							
Log(Muslim Trump tweets)	0.130*	0.140**	0.132*	0.101	0.099	0.192**	0.116
	(0.069)	(0.066)	(0.068)	(0.062)	(0.063)	(0.077)	(0.074)
Panel B: First Stage - Log(Trump tweets about Muslims)							
Trump golfs	0.102***	0.098***	0.104***	0.103***	0.078***	0.086***	0.098***
	(0.027)	(0.026)	(0.027)	(0.027)	(0.025)	(0.025)	(0.027)
Panel C: Reduced form - Log(Hate crimes against Muslims) in t+2							
Trump golfs	0.165**	0.173**	0.158**	0.168**	0.157**	0.172**	0.163**
	(0.071)	(0.076)	(0.070)	(0.068)	(0.070)	(0.074)	(0.071)
Panel D: 2SLS - Log(Hate crimes against Muslims) in t+2							
Log(Muslim Trump tweets)	1.617**	1.756**	1.523**	1.626**	2.009*	2.011*	1.659**
	(0.779)	(0.892)	(0.736)	(0.761)	(1.198)	(1.050)	(0.842)
Weak IV 95% AR confidence set	[0.31; 4.01]	[0.43; 4.49]	[0.29; 3.64]	[0.50; 3.96]	[0.47; 6.87]	[0.48; 5.79]	[0.41; 4.41]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363	363
R^2	0.21	0.17	0.25	0.21	0.08	0.12	0.20
Robust F-stat.	13.15	12.97	13.55	13.54	9.487	10.90	11.87

Notes: This table presents OLS and IV regressions where the dependent variable is the number of hate crimes against Muslims on any given day based on FBI data. We use a dummy for days on which President Donald Trump golfs used as an instrument for his tweets about Muslims. Column 2 controls for one lag of the dependent variable and column 3 for a dummy that tags federal holidays. Column 4 controls for the first principal component of Google searches for Islam-related terms. Column 5 controls for the number of times Fox News, CNN or MSNBC mention Islam-related words in their reporting on a given day. Column 6 controls for the number of terror attacks in the US, Europe, or other countries. Column 7 controls for the total number of tweets by Donald Trump. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as a dummy for whether Trump's golfing is the first of a series of golf days. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Spillover Effects on Trump's Followers and Cable News Coverage

	Trump followers' Muslim tweets			Cable news coverage			
	Trump retweets (1)	New content (2)	Contains #BanIslam (3)	All stations (4)	Fox News (5)	CNN (6)	MSNBC (7)
Panel A: OLS - Log(Total number of Muslim TV mentions/tweets)							
Log(Muslim Trump tweets)	2.658***	0.680***	0.360***	0.677***	0.607***	0.808***	0.660***
	(0.346)	(0.105)	(0.094)	(0.089)	(0.117)	(0.109)	(0.084)
Panel B: Reduced Form - Log(Total number of Muslim TV mentions/tweets)							
Trump golfs	0.456**	0.117**	0.234***	0.273**	0.296**	0.285	0.185*
	(0.208)	(0.058)	(0.074)	(0.134)	(0.115)	(0.212)	(0.110)
Panel C: 2SLS - Log(Total number of Muslim TV mentions/tweets)							
Log(Muslim Trump tweets)	4.508***	1.151**	2.313**	2.701**	2.923***	2.813	1.830**
	(1.305)	(0.469)	(0.955)	(1.114)	(0.966)	(1.891)	(0.921)
Weak IV 95% AR confidence set	[1.01; 6.96]	[0.17; 2.21]	[0.89; 5.43]	[0.39; 5.24]	[1.11; 5.31]	[-1.49; 7.12]	[-0.27; 3.93]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	364	364	364	364	364	364	364
Robust F-stat.	13.02	13.02	13.02	13.02	13.02	13.02	13.02

Notes: This table presents OLS and IV regressions where the dependent variable is the number of tweets by Trump followers in columns 1 to 3 and the number of times Muslims are mentioned on cable news stations on a given day in columns 4 to 7. We use a dummy for days on which President Donald Trump golfs used as an instrument for his tweets about Muslims. *Trump retweets* are retweets by Trump followers of Trump's negative tweets about Muslims. *New content* refers to tweets by Trump followers mentioning Muslims that are no Trump retweets and do not mention Trump. *Contains #BanIslam* is the number of tweets by Trump followers containing the hashtag #BanIslam. *Cable news coverage* is based on the mentions of Muslim-related words on Fox News, CNN, and MSNBC, which are also reported separately. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as a dummy for whether Trump's golfing is the first of a series of golf days. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Robustness Bartik Interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel C: OLS – Log(Hate crimes against Muslims) in t+1							
Muslim Trump Tweet × Twitter Usage	0.013**	0.010**	0.014**	0.014**	0.015**	0.017***	0.015**
	(0.005)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Muslim Trump Tweet × Fox News Viewership						0.002**	
						(0.001)	
Muslim Trump Tweet × Republican Vote Share 2012							-0.000
							(0.001)
Panel B: Reduced Form – Log(Hate crimes against Muslims) in t+1							
Muslim Trump Tweet × SXSW Treat	0.010**	0.009**	0.010**	0.010**	0.010**	0.010**	0.010**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Muslim Trump Tweet × SXSW Pre	0.001	-0.000	-0.000	-0.001	-0.000	-0.000	-0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Muslim Trump Tweet × Fox News Viewership						0.001*	
						(0.001)	
Muslim Trump Tweet × Republican Vote Share 2012							-0.001
							(0.001)
Panel C: 2SLS – Log(Hate crimes against Muslims) in t+1							
Muslim Trump Tweet × Twitter Usage	0.143***	0.124**	0.137**	0.141**	0.147***	0.185***	0.193**
	(0.049)	(0.052)	(0.053)	(0.053)	(0.053)	(0.068)	(0.073)
Muslim Trump Tweet × SXSW Pre	-0.005	-0.006	-0.006	-0.007	-0.007	-0.008	-0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Muslim Trump Tweet × Fox News Viewership						0.021***	
						(0.007)	
Muslim Trump Tweet × Republican Vote Share 2012							0.024**
							(0.010)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop. deciles x Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County x Month FE		Yes	Yes	Yes	Yes	Yes	Yes
State x Day FE			Yes	Yes	Yes	Yes	Yes
County x Day of Week FE				Yes	Yes	Yes	Yes
County x Day of Month FE					Yes	Yes	Yes
Lag dep. variable	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2887332	2887332	2886403	2886403	2886403	2885474	2886403

Notes: This table presents OLS, reduced form and IV regressions where the dependent variable is the log number of anti-Muslims hate crime in county c on day d. The independent variable is either the interaction Trump's anti-Muslim tweet with county-level Twitter usage or a reduced form/2SLS specification with our SXSW variables. The variables are standardized to have a mean of zero and standard deviation of one. All regressions include population controls, one lag of the dependent variable, as well as county and day fixed effects. Some regressions further control for county × month, state × day, county × day-of-week, and county × day-of-month fixed effects (as indicated). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A Online Appendix:

A.1. Appendix 1: Additional Details on Data

Table A.1: Variable Descriptions (Part 1/2)

Variable	Description	Source
	Hate crime variables	
Hate crimes	Total number of hate crimes recorded in the FBI hate crime data.	FBI Hate Crime Data
Anti-Muslim hate crimes	Anti-Muslim hate crimes recorded in the FBI hate crime data, based on bias motivation code 24.	FBI Hate Crime Data
Anti-Hispanic hate crimes	Anti-Hispanic hate crimes recorded in the FBI hate crime data, based on the bias motivation codes 32.	FBI Hate Crime Data
Other ethnic-based hate crimes	Anti-ethnic hate crimes recorded in the FBI hate crime data, based on the bias motivation codes 33.	FBI Hate Crime Data
Anti-racial hate crimes	Racial hate crimes recorded in the FBI hate crime data, based on bias motivation codes 11, 12, 13, 14, 15, 16.	FBI Hate Crime Data
Anti-religious hate crimes	Anti-religious hate crimes (except anti-Muslim) recorded in the FBI hate crime data, based on bias motivation codes 21, 22, 23, 25, 26, 27, 28, 29, 81, 82, 83, 84, 85.	FBI Hate Crime Data
Anti-sexual orientation hate crimes	Hate crimes based on sexual orientation recorded in the FBI hate crime data, based on the bias motivation codes 41, 42, 43, 44, 45.	FBI Hate Crime Data
	Twitter data	
Trump tweets	The total number of tweets from Donald Trump's Twitter account.	Trump Twitter Archive
Muslim tweets	The number of tweets from Donald Trump's Twitter account about Islam-related topics. We start classifying these tweets by searching for the terms "sharia", "refugee", "mosque", "muslim", "islam" and "terror". We then read all tweets and verify that they indeed mention Muslims in a negative way.	Trump Twitter Archive
Twitter usage	The number of geolocated tweets per county that were collected using the Twitter streaming API in a 12 month period from June to November 2014 and June to November 2015.	Gesis Datatorium
SXSW followers, March 2007	The number of Twitter users following the SXSW account in each county that signed up to Twitter in March 2007.	Twitter Search API
SXSW followers, Pre	The total number of Twitter users following the SXSW account in each county that signed up to Twitter at any point in 2006.	Twitter Search API
Burning Man Twitter Users, August 2007	The number of Twitter users in each county that tweeted about the Burning Man festival in August 2007 and joined Twitter in August 2007.	Twitter Search API
Coachella Twitter Users, April 2007	The number of Twitter users in each county that tweeted about the Coachella festival in April 2007 and joined Twitter in April 2007.	Twitter Search API
Lollapalooza Twitter Users, August 2007	The number of Twitter users in each county that tweeted about the Lollapalooza festival in August 2007 and joined Twitter in August 2007.	Twitter Search API
	Trump golf data	
Trump golfs	A dummy variable for each day in 2017 Trump spent on a golf course and likely played golf.	NYT, trumpgolfcount.com and Pres. Schedule
Trump golfs (NYT only)	A dummy variable for each day in 2017 Trump spent on a Golf course and likely golfed, based solely on the information of the New York Times.	NYT
Trump golf (alternative)	A dummy variable for each day in 2017 Trump spent on a golf course and likely golfed, based on the information of trumpgolfcount.com and extended with information from the Pres. Schedule	trumpgolfcount.com and Pres. Schedule
Golf holiday	A dummy for any of Trump's golf outings that lasts longer than 3 days.	NYT and trumpgolfcount.com
Golf at any point in previous week	A dummy variable which is 1 if Trump golfed at any point in the previous week.	NYT and trumpgolfcount.com

Table A.2: Variable Descriptions (Part 2/2)

Variable	Description	Source
Other cross sectional controls		
Demographic controls	Contain the share of people in the age buckets 20-24, 25-29, 30-34, 40-44, 45-49 and 50+, and the percentage change in population between 2000 and 2016.	US Census
Education controls	Contains the share of people over 25 with at least a high school degree and the share of people over 25 with at least a graduate degree.	US Census
Race and religion controls	Contains population shares of Muslims, Whites, Blacks, Native Americans, Asians, and Hispanics.	US Census/Religious Census
Socioeconomic controls	Contains a county's poverty rate, unemployment rate, GINI coefficient, share of uninsured, log of median household income, and the share of the population employed in agriculture, manufacturing, accommodation/retail, utilities, information technologies services, and other industries.	US Census/Bureau of Labor Statistics
Media controls	Contains the ratio of prime time TV viewership to population, cable spending to population, and the share of Fox News viewership.	SimplyAnalytics
Election control	Contains the vote share of the Republican party in the 2012 presidential election.	MIT Election Lab
Crime controls	Contains the number of violent crime per capita as well as the number of property crimes per capita based on FBI data.	FBI UCR Data
Distance control	Contains the distance to Austin Texas, the population density, and the logarithm of the land area for each county.	US Census Tigerline File
Change in implicit bias against Muslims	The change in the county-level mean implicit association test score from the Arab-Muslim module between 2015-2017 compared to 2010-2014.	Project Implicit
Other time series variables		
Trump followers' retweets	The number of retweets of Trump's tweets about Muslims by his Twitter followers	Twitter
Trump followers' new content	The number of tweets by Trump followers containing the words "sharia", "refugee", "mosque", "muslim", "islam" or "terror".	Twitter
Contains #BanIslam	The number of tweets by Trump followers containing the term "#BanIslam".	Twitter
Muslim mentions (total)	The total number of cable news reports mentioning one of the following terms in their closed captions: "sharia", "refugee", "mosque", "muslim", "islam" and "terror".	Internet Archive
Muslim mentions (Fox News)	The total number of news reports on Fox News mentioning one of the following terms in their closed captions: "sharia", "refugee", "mosque", "muslim", "islam" and "terror".	Internet Archive
Muslim mentions (CNN)	The total number of news reports on CNN mentioning one of the following terms in their closed captions: "sharia", "refugee", "mosque", "muslim", "islam" and "terror".	Internet Archive
Muslim mentions (MSNBC)	The total number of news reports on MSNBC mentioning one of the following terms in their closed captions: "sharia", "refugee", "mosque", "muslim", "islam" and "terror".	Internet Archive
Google searches (PC)	The first princepal component of the rescaled Google trends for the following terms: "sharia", "refugee", "mosque", "muslim", "islam" and "terror".	Google Trends
Terror attack in the US	The number of Islamist terror attacks committed in the US.	Global Terrorism Database
Terror attack in Europe	The number of Islamist terror attacks committed in the Europe.	Global Terrorism Database
Terror attack elsewhere	The number of Islamist terror attacks committed outside of the US or Europe	Global Terrorism Database

A.1.1 FBI Hate Crime Data

As described in the Section 2, the FBI uses a two-tier decision making process for classifying hate crimes. FBI (2015) describes the decision making process in the following way:

> "Once the development of this collection was complete, the FBI UCR Program surveyed state UCR Program managers on hate crime collection procedures used at various law enforcement agencies which collected hate crime data employing a two-tier decision-making process. The first level is the law enforcement officer who initially responds to the alleged hate crime incident, i.e., the responding officer (or first-level judgment officer). It is the responsibility of the responding officer to determine whether there is any indication that the offender was motivated by bias. If a bias indicator is identified, the officer designates the incident as a suspected bias-motivated crime and forwards the case file to a second-level judgment officer/unit. (In smaller agencies this is usually a person specially trained in hate crime matters, while in larger agencies it may be a special unit.) It is the task of the second-level judgment officer/unit to review the facts of the incident and make the final determination of whether a hate crime has actually occurred. If so, the incident is to be reported to the FBI UCR Program as a bias-motivated crime." (FBI, 2015, pp. 2-3)

As indicated, all decisions by the responding officer will be passed on for review to a second examiner. The FBI manual also outlines criteria that have to be full-filled for a crime to be classified as a hate crime:

> "An important distinction must be made when reporting a hate crime. The mere fact the offender is biased against the victims actual or perceived race, religion, disability, sexual orientation, ethnicity, gender, and/or gender identity does not mean that a hate crime was involved. Rather, the offenders criminal act must have been motivated, in whole or in part, by his or her bias. Motivation is subjective, therefore, it is difficult to know with certainty whether a crime was the result of the offenders bias. For that reason, before an incident can be reported as a hate crime, sufficient objective facts must be present to lead a reasonable and prudent person to conclude that the offenders actions were motivated, in whole or in part, by bias. While no single fact may be conclusive, facts such as the following, particularly when combined, are supportive of a finding of bias:

1. The offender and the victim were of a different race, religion, disability, sexual orientation, ethnicity, gender, and/or gender identity. For example, the victim was African American and the offender was white.

2. Bias-related oral comments, written statements, or gestures were made by the offender indicating his or her bias. For example, the offender shouted a racial epithet at the victim.

3. Bias-related drawings, markings, symbols, or graffiti were left at the crime scene. For example, a swastika was painted on the door of a synagogue, mosque, or LGBT center.

4. Certain objects, items, or things which indicate bias were used. For example, the offenders wore white sheets with hoods covering their faces or a burning cross was left in front of the victims residence.

5. The victim is a member of a specific group that is overwhelmingly outnumbered by other residents in the neighborhood where the victim lives and the incident took place.

6. The victim was visiting a neighborhood where previous hate crimes had been committed because of race, religion, disability, sexual orientation, ethnicity, gender, or gender identity and where tensions remained high against the victims group.

7. Several incidents occurred in the same locality, at or about the same time, and the victims were all of the same race, religion, disability, sexual orientation, ethnicity, gender, or gender identity.

8. A substantial portion of the community where the crime occurred perceived that the incident was motivated by bias.

9. The victim was engaged in activities related to his or her race, religion, disability, sexual orientation, ethnicity, gender, or gender identity. For example, the victim was a member of the National Association for the Advancement of Colored People (NAACP) or participated in an LGBT pride celebration.

10. The incident coincided with a holiday or a date of significance relating to a particular race, religion, disability, sexual orientation, ethnicity, gender, or gender identity, e.g., Martin Luther King Day, Rosh Hashanah, or the Transgender Day of Remembrance.

11. The offender was previously involved in a similar hate crime or is a hate group member.
12. There were indications that a hate group was involved. For example, a hate group claimed responsibility for the crime or was active in the neighborhood.
13. A historically-established animosity existed between the victims and the offenders groups.
14. The victim, although not a member of the targeted racial, religious, disability, sexual orientation, ethnicity, gender, or gender identity group, was a member of an advocacy group supporting the victim group."

(FBI, 2015, pp. 6-7)

We report the full list of FBI bias motivation categories in Table A.4. The hate crime categories we use in the paper are defined as follows:

Table A.3: FBI Hate Crimes Codes

Hate Crime Category	**FBI Codes**
Muslim	24
Hispanic	32
Other ethnic	33
Racial	11, 12, 13, 14, 15, 16
Sexual orientation	41, 42, 43, 44, 45
Religious (excluding Muslim)	21, 22, 23, 25, 26, 27, 28, 29, 81, 82, 83, 84, 85

Table A.4: **Full List of FBI Bias Motivation Categories**

Bias category	Bias motivation and code
Race/Ethnicity/Ancestry	Anti-American Indian or Alaska Native (13) Anti-Arab (31) Anti-Asian (14) Anti-Black or African American (12) Anti-Hispanic or Latino (32) Anti-Multiple Races, Group (15) Anti-Native Hawaiian or Other Pacific Islander (16) Anti-Other Race/Ethnicity/Ancestry (33) Anti-White (11)
Religion	Anti-Buddhist (83) Anti-Catholic (22) Anti-Eastern Orthodox (81) Anti-Hindu (84) Anti-Islamic (Muslim) (24) Anti-Jehovahs Witness (29) Anti-Jewish (21) Anti-Mormon (28) Anti-Multiple Religions, Group (26) Anti-Other Christian (82) Anti-Other Religion (25) Anti-Protestant (23) Anti-Sikh (85) Anti-Atheism/Agnosticism (27)
Sexual Orientation	Anti-Bisexual (45) Anti-Gay (Male) (41) Anti-Heterosexual (44) Anti-Lesbian (42) Anti-Lesbian, Gay, Bisexual, or Transgender (Mixed Group)
Disability	Anti-Mental Disability (52) Anti-Physical Disability (51)
Gender	Anti-Female (62) Anti-Male (61)
Gender Identity	Anti-Gender Nonconforming (72) Anti-Transgender (71)

Notes: This table reports the complete list of hate crime bias motivations as classified by the FBI. The table is reproduced from (FBI, 2015, p. 5).

A.1.2 Trump Twitter Data

Table A.5: Examples of Trump's Negative Tweets about Muslims

Date	Text	Retweets
12/10/2015	"mimi_saulino: seanhannity @FoxNews Syrian Muslims escorted into U.S. through Mexico. Now arriving to Oklahoma and Kansas! Congress?"	1223
14/11/2015	Why won't President Obama use the term Islamic Terrorism? Isn't it now, after all of this time and so much death, about time!	6924
15/11/2015	"thewatcher23579: One of Paris terrorist came as Syrian refugee. Donald Trump is right again. BOMB THEIR OIL - TAKE AWAY THEIR FUNDING"	2165
17/11/2015	Refugees from Syria are now pouring into our great country. Who knows who they are - some could be ISIS. Is our president insane?	16285
22/11/2015	We better get tough with RADICAL ISLAMIC TERRORISTS, and get tough now, or the life and safety of our wonderful country will be in jeopardy!	5172
25/11/2015	I LIVE IN NEW JERSEY; @realDonaldTrump IS RIGHT: MUSLIMS DID CELEBRATE ON 9/11 HERE! WE SAW IT! https://t.co/1SksZU9qlj	2252
07/12/2015	Obama said in his speech that Muslims are our sports heroes. What sport is he talking about, and who? Is Obama profiling?	9600
07/12/2015	Statement on Preventing Muslim Immigration: https://t.co/HCWU16z6SR https://t.co/d1dhaIs0S7	4716
10/12/2015	The United Kingdom is trying hard to disguise their massive Muslim problem. Everybody is wise to what is happening, very sad! Be honest.	6028
10/12/2015	In Britain, more Muslims join ISIS than join the British army. https://t.co/LQVNz7b2Eb	4325
17/01/2016	Far more killed than anticipated in radical Islamic terror attack yesterday. Get tough and smart U.S., or we won't have a country anymore!	4126
27/03/2016	Another radical Islamic attack, this time in Pakistan, targeting Christian women & children. At least 67 dead,400 injured. I alone can solve	11353
22/05/2016	Crooked Hillary wants a radical 500% increase in Syrian refugees. We cant allow this. Time to get smart and protect America!	9758
12/06/2016	Appreciate the congrats for being right on radical Islamic terrorism, I don't want congrats, I want toughness & vigilance. We must be smart!	27146
13/06/2016	In my speech on protecting America I spoke about a temporary ban, which includes suspending immigration from nations tied to Islamic terror.	13026
25/06/2016	We must suspend immigration from regions linked with terrorism until a proven vetting method is in place.	11726
28/07/2016	Hillary's refusal to mention Radical Islam, as she pushes a 550% increase in refugees, is more proof that she is unfit to lead the country.	20106
18/10/2016	Thank you Colorado Springs. If Im elected President I am going to keep Radical Islamic Terrorists out of our count https://t.co/N74UK73RLK	12904
19/10/2016	ISIS has infiltrated countries all over Europe by posing as refugees, and @HillaryClinton will allow it to happen h https://t.co/MmeW2qsTQh	16130
11/02/2017	Our legal system is broken! "77% of refugees allowed into U.S. since travel reprieve hail from seven suspect countries." (WT) SO DANGEROUS!	23082
17/08/2017	Study what General Pershing of the United States did to terrorists when caught. There was no more Radical Islamic Terror for 35 years!	30534
18/08/2017	Radical Islamic Terrorism must be stopped by whatever means necessary! The courts must give us back our protective rights. Have to be tough!	37669
15/09/2017	Loser terrorists must be dealt with in a much tougher manner.The internet is their main recruitment tool which we must cut off & use better!	21411
20/10/2017	Just out report: "United Kingdom crime rises 13% annually amid spread of Radical Islamic terror." Not good, we must keep America safe!	29854
01/11/2017	NYC terrorist was happy as he asked to hang ISIS flag in his hospital room. He killed 8 people, badly injured 12. SHOULD GET DEATH PENALTY!	43455

Notes: This table reports examples of Trump's negative tweets about Muslims, including the date of the tweet and the number of retweets the tweet received.

Table A.6: Misclassified Trump’s Anti-Muslim Tweets

Date	Text	Retweets
12/12/2012	Watching Pyongyang terrorize Asia today is just amazing!	77
26/03/2013	The Scottish windfarm was conceived by the same mind that released terrorist al-Megrahi for humanitarian reasons. ..	101
23/04/2013	Did the Boston terrorists register their guns? No. Another example of why gun control legislation is not the answer!	1192
22/09/2013	”@LebaneseKobe: @realDonaldTrump as a Muslim and as an American, i know for a fact that you Mr. Trump respect all people!	33
22/09/2013	”@mandem3: realDonaldTrump you hate muslims.” Wrong	48
10/10/2013	Obama has called @GOP terrorists during this showdown. Its a shame he really doesnt think it because then he would meet all @GOP demands.	432
29/01/2014	Remember when ”comedian” Bill Maher openly praised the disgusting terrorists who destroyed the World Trade Center-then got canned by ABC?	117
26/01/2015	”tomtumillo: What is worse, Geraldo screaming ’screw the terrorists’ or Kenya feeling she’s ’fabulous’? #CelebrityApprentice	56
15/08/2015	”javonniandjeno: realDonaldTrump AP nbc Donald Trump is Clint Eastwood, the perfect hero not scared of American terrorists. Vote Trump!”	1742
27/08/2015	”jp_sitles: realDonaldTrump HillaryClinton: she compared republicans to terrorist but will not call terrorists , terrorists. #OhMe”	2869
06/09/2015	”jasonusmc2017: blayne_troy @realDonaldTrump: He was right when he called Obama the 5 for 1 president. 5 terrorist for one no good traitor	1016
21/09/2015	”TheBrodyFile: On the Muslim issue: It might help @BarackObama if he actually supported Christians religious liberty rights.	1242
21/09/2015	”TheBrodyFile: On the Muslim issue: It might help @BarackObama if he didn’t take five years to visit Israel”	818
21/11/2015	”WayneDupreeShow: ”Its clear that Donald Trump was NOT even talking about a Muslim Database!” https://t.co/3tLDZj2WGV”	1020
31/12/2015	”SenSanders: I have a message for Donald Trump: No, were not going to hate Latinos, were not going to hate Muslims.” I fully agree!	1250
23/03/2016	Just watched Hillary deliver a prepackaged speech on terror. Shes been in office fighting terror for 20 years-and look where we are!	11115
23/03/2016	I will be the best by far in fighting terror. Im the only one that was right from the beginning, & now Lyin Ted & others are copying me.	7224
15/06/2016	I will be meeting with the NRA, who has endorsed me, about not allowing people on the terrorist watch list, or the no fly list, to buy guns.	13903
21/05/2017	Speech transcript at Arab Islamic American Summit https://t.co/eUWxJXJxbe nReplay https://t.co/VtmlSqciXx #RiyadhSummit #POTUSAbroad	11498
26/05/2017	Getting ready to engage G7 leaders on many issues including economic growth, terrorism, and security.	11322
27/05/2017	Big G7 meetings today. Lots of very important matters under discussion. First on the list, of course, is terrorism. #G7Taormina	9489
18/08/2017	Today, I signed the Global War on Terrorism War Memorial Act (#HR873.) The bill authorizes....cont https://t.co/c3zIkdtowc https://t.co/re6n0MS0cj	14892
07/09/2017	During my trip to Saudi Arabia, I spoke to the leaders of more than 50 Arab & Muslim nations about the need to confront our shared enemies.[...]	10156
11/11/2017	When will all the haters and fools out there realize that having a good relationship with Russia is a good thing, not a bad thing.[...]	39627

Notes: The table lists the tweets we excluded by hand from the set of negative Muslim tweets.

A.1.3 Geocoded Twitter Data

Table A.7: Search Terms Used to Identify Users Tweeting about Other Festivals

Festival	Search Term
Austin City Limited Festival	Austin City Limits Festival
Burning Man	Burningman Burning Man
Coachella	Coachella
Electric Daisy Festival	EDC Las Vegas Electric Daisy Carnival
New Orleans Jazz and Heritage Festival	New Orleans Jazz and Heritage Festival Jazzfest
Lollapalooza	Lollapalooza
Pitchfork Music Festival	Pitchfork Music Festival Pitchforkfest
South by Southwest Festival	South by Southwest SXSW
West by Southwest Festival	West by Southwest WXSW

Table A.8: Search Terms Used to Create a Proxy for Total Tweets

0	but	his	one	these	would
1	by	how	only	they	year
2	can	if	or	think	you
3	come	in	other	this	your
4	could	into	our	time	
5	day	it	out	two	
6	do	its	over	up	
7	even	just	people	us	
8	first	know	say	use	
9	for	like	see	want	
I	from	look	she	way	
about	get	make	so	we	
after	give	me	some	well	
all	go	most	take	what	
also	good	my	than	when	
any	have	new	that	which	
as	he	no	their	who	
at	he	not	them	with	
back	her	now	then	with	
because	him	on	there	work	

Notes: This table list the search terms we used to collect a proxy of all tweets sent from a given county.

A.1.4 Rescaling of Google trends

As described in Section 2, we use the weekly Google trends data to rescale the daily Google trend values. The daily Google trends data are scaled between 0-100 for each 90 day period, while the weekly Google trends data have a consistent scaling for the entire time period.

To arrive at consistent values, we use the following process. First, we create a scaling factor by dividing the weekly interest by the daily interest. We then multiply the daily interest data with the scaling factor. If the weekly interest is 100 and the daily interest is 25, the scaling factor will be 4 and values will be scaled up. On the other hand, if the weekly interest is low, for example 10, a daily interest of 25 would be scaled down. This way, the adjustment guarantees that daily interest will be on the same scale and thus comparable over time.

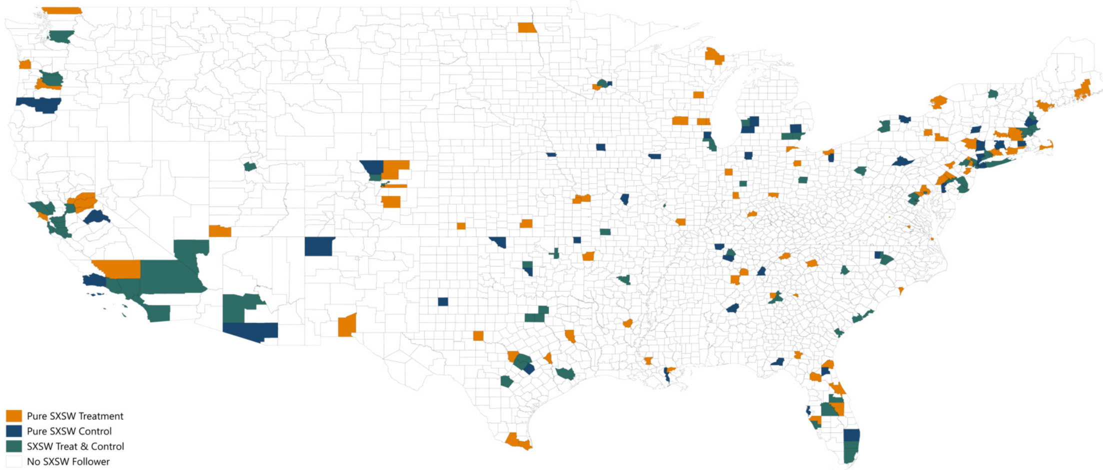
As a final step, we divide the rescaled values by their maximum and multiply them by 100. This is to re-normalize the Google trend values to take on values between 0 and 100.

A.1.5 Sources for Trump's golf activity

Table A.9: Sources for Golf Data

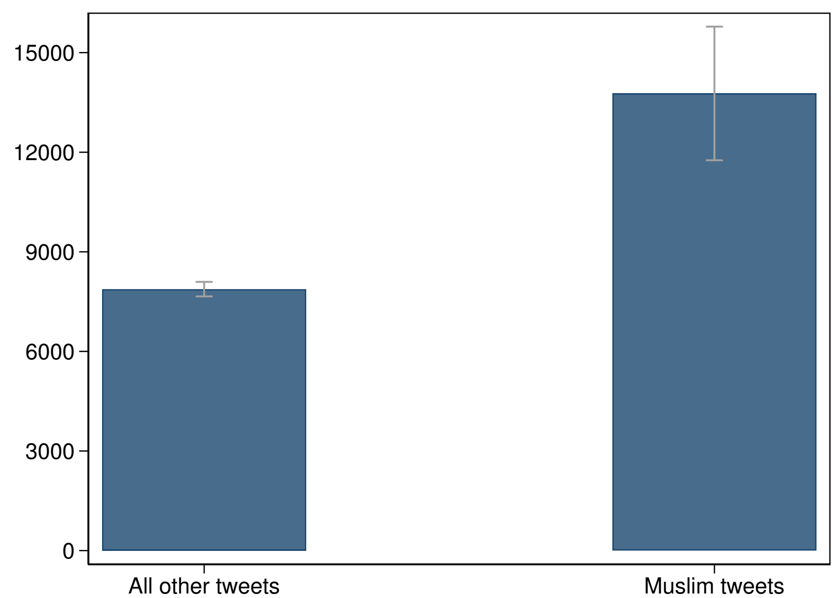
Source	**Description**
New York Times	The NYT tracks visits by Trump to his own properties. The data also track how often Trump visited a golf club.
trumpgolfcount.com	This website lists Trump's visits to golf clubs since his inauguration. It also provides additional analysis during which visits Trump likely played golf.
Presidential Schedule	The presidential schedule lists all past presidential journeys.

Figure A.1: Identifying Variation



Notes: This map plots counties with SXSW followers who joined Twitter in March 2007 in orange; counties with SXSW followers who joined prior to the 2007 event in blue; and counties in both categories in green.

Figure A.2: Average Retweets of Trump's Tweets, by Muslim Content



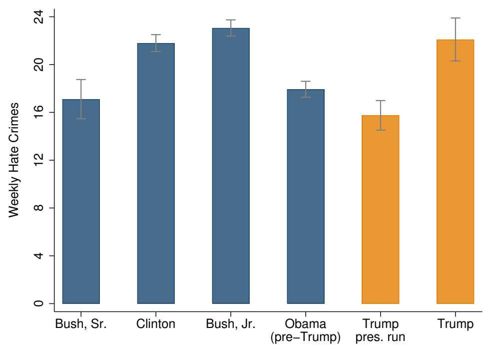
Notes: This figure plots the average number of retweets Donald Trump received on his tweets about Muslims compared to all other tweets. We also show 95% confidence intervals.

A.2. Appendix 2: Details on Trends in Hate Crimes by President

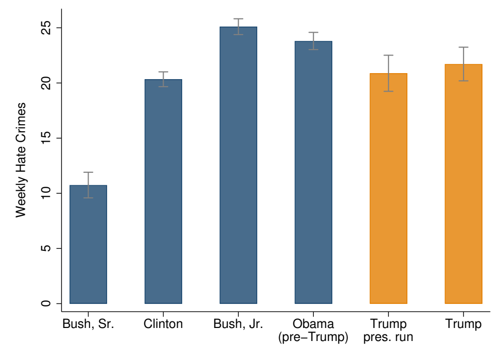
In this section, we provide some additional evidence on time series trends in hate crimes across US presidencies since 1990. A potential issue with the hate crime numbers we presented in Figure 1 might be that we consider all hate crimes jointly, which could hide underlying heterogeneous hate crime trends across groups. We thus reproduce the bar graphs using the other main categories of hate crimes in the FBI data (see Figure A.3). Overall, the results yield a qualitatively similar conclusion. Trump does not appear to be an outlier for any of the main categories except Muslims.

Figure A.3: Average Weekly Hate Crimes since 1990, by President and Motivating Bias

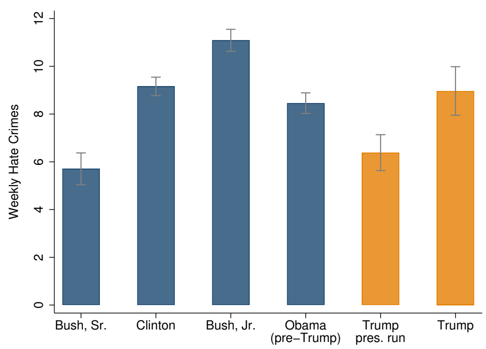
(a) Religious



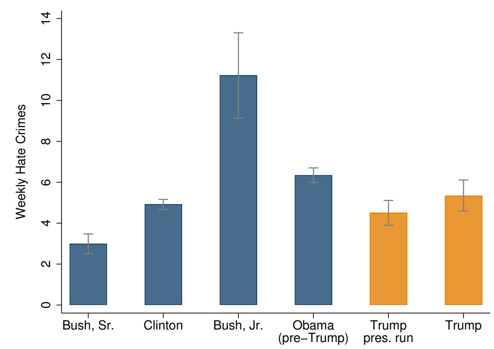
(b) Sexual orientation



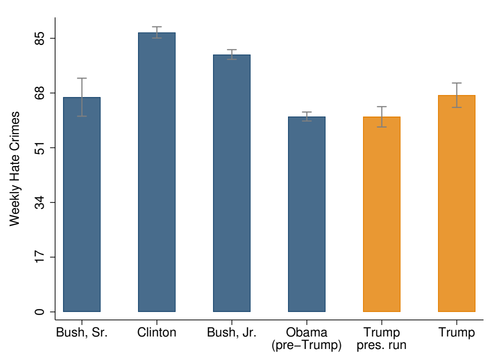
(c) Anti-Hispanic



(d) Other ethnic



(e) Racial



Notes: This figure plots the average weekly number of hate crimes, by president and type of hate crime (as defined by the FBI). The headings indicate which type of hate crime is plotted. The whiskers indicate the 95% confidence intervals.

A.3. Appendix 3: Additional Cross-sectional Evidence

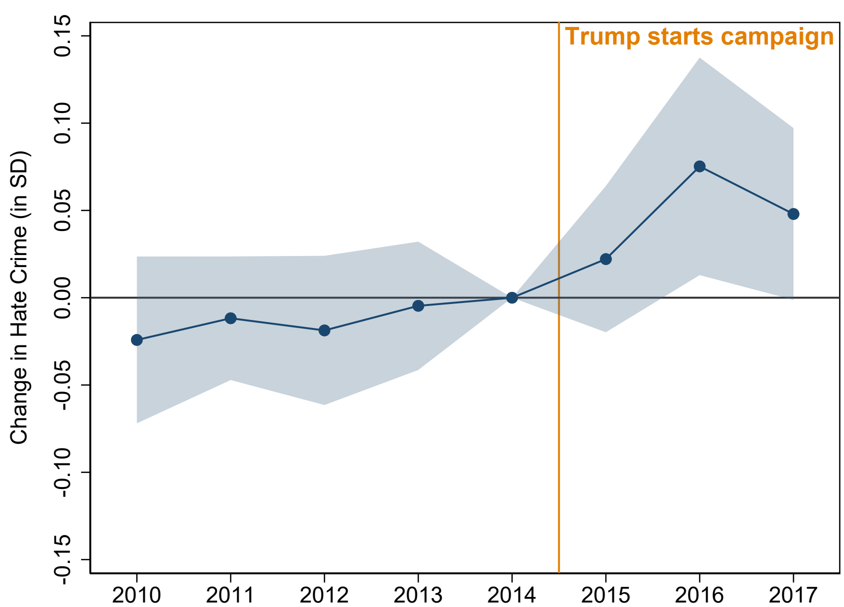
Table A.10: Descriptive Statistics (Main Variables)

	Mean	Std. Dev.	Min.	Median	Max.	N
Hate crime and Twitter variables						
Δ Log(Hate crimes against Muslims)	0.02	0.13	-0.71	0.00	1.26	3108
Log(Twitter usage)	10.03	1.91	3.33	9.94	16.90	3108
Log(SXSW followers, March 2007)	0.06	0.32	0.00	0.00	4.98	3108
Log(SXSW followers, Pre)	0.02	0.18	0.00	0.00	3.61	3108
Demographic controls						
% aged 20-24	0.06	0.02	0.01	0.06	0.27	3108
% aged 25-29	0.06	0.01	0.03	0.06	0.15	3108
% aged 30-34	0.06	0.01	0.03	0.06	0.12	3108
% aged 35-39	0.06	0.01	0.03	0.06	0.11	3108
% aged 40-44	0.06	0.01	0.02	0.06	0.10	3108
% aged 45-49	0.06	0.01	0.02	0.06	0.09	3108
% aged 50+	0.39	0.07	0.11	0.39	0.75	3108
Population growth, 2000-2016	0.06	0.18	-0.43	0.03	1.32	3108
Geographical controls						
Population density	261.27	1733.47	0.10	45.60	69468.40	3108
Log(County area)	6.53	0.86	0.69	6.47	9.91	3108
Distance from Austin, TX (in miles)	1450.64	612.61	5.04	1464.66	3098.88	3108
Race and religion controls						
% white	0.77	0.20	0.03	0.84	0.98	3108
% black	0.09	0.14	0.00	0.02	0.85	3108
% native American	0.02	0.06	0.00	0.00	0.90	3108
% Asian	0.01	0.02	0.00	0.01	0.37	3108
% Hispanic	0.09	0.14	0.01	0.04	0.96	3108
% Muslim	0.23	1.08	0.00	0.00	30.35	3108
Socioeconomic controls						
% below poverty level	16.74	6.58	1.40	16.00	53.30	3108
% unemployed	5.50	1.94	1.80	5.30	24.10	3108
Gini index	0.44	0.03	0.33	0.44	0.65	3108
% uninsured	13.32	5.28	1.80	12.80	49.00	3108
Log(Median household income)	10.72	0.24	9.87	10.71	11.72	3107
% employed in agriculture	0.01	0.03	0.00	0.00	0.58	3108
% employed in IT	0.01	0.01	0.00	0.01	0.21	3108
% employed in manufacturing	0.16	0.13	0.00	0.13	0.72	3108
% employed in nontradable sector	0.29	0.11	0.00	0.28	1.00	3108
% employed in construction/real estate	0.07	0.05	0.00	0.06	1.00	3108
% employed in utilities	0.04	0.05	0.00	0.03	1.00	3108
% employed in business services	0.16	0.07	0.00	0.15	0.95	3108
% employed in other services	0.25	0.10	0.00	0.24	1.00	3108
% adults with high school degree	34.77	7.07	7.50	35.20	54.80	3108
% adults with graduate degree	7.05	4.12	0.00	5.80	44.40	3108

Table A.11: **Descriptive Statistics (Main Variables, Continued)**

	Mean	Std. Dev.	Min.	Median	Max.	N
Media controls						
% watching Fox News	0.26	0.01	0.23	0.26	0.30	3107
% watching prime time TV	0.43	0.01	0.40	0.43	0.47	3107
Election control						
Republican vote share, 2012	0.60	0.15	0.06	0.61	0.96	3108
Crime controls						
Violent crime rate	0.00	0.00	0.00	0.00	0.02	3108
Property crime rate	0.02	0.01	0.00	0.01	0.10	3108
Other hate crime variables						
Δ Log(Total hate crimes)	-0.01	0.36	-2.28	0.00	2.04	3108
Δ Log(Hate crimes against Hispanics)	-0.01	0.17	-1.65	0.00	1.21	3108
Δ Log(Other ethnicity-based hate crimes)	-0.02	0.16	-2.60	0.00	1.09	3108
Δ Log(Racially motivated hate crimes)	-0.01	0.31	-1.69	0.00	1.74	3108
Δ Log(Hate crimes based on sexual orientation)	-0.03	0.22	-1.46	0.00	1.20	3108
Δ Log(Hate crimes against other religions)	0.00	0.21	-1.58	0.00	1.59	3108
Log(Total hate crimes, ADL data)	0.23	0.63	0.00	0.00	5.38	3108

Figure A.4: Change in Anti-Muslim Hate Crimes by Twitter Usage (Reduced Form)



Notes: This figure plots the coefficients from running panel event study regressions as in Equation (1), where $\log(Twitter\ Usage)$ is replaced by $\log(SXSW\ followers, March$ 2007. The dependent variable is the log number of hate crimes in a county. We standardized the variables to have a mean of zero and standard deviation of one. The vertical line indicates the start of Trump's presidential campaign start. The shaded areas are 95% confidence intervals.

Table A.12: Comparing Counties with SXSW Followers, March 2007 vs. Pre

	March 2007 *and* Pre (1)	March 2007 *only* (2)	Pre *only* (3)	Difference in means (2) - (3)	t-stat
Demographic controls					
% aged 20-24	0.07	0.08	0.08	0.00	0.13
% aged 25-29	0.09	0.07	0.07	-0.00	-0.57
% aged 30-34	0.08	0.07	0.07	-0.00	-0.45
% aged 35-39	0.07	0.06	0.06	-0.00	-0.21
% aged 40-44	0.06	0.06	0.06	0.00	0.25
% aged 45-49	0.07	0.06	0.06	0.00	0.14
% aged 50+	0.32	0.35	0.35	-0.00	-0.03
Population growth, 2000-2016	0.18	0.18	0.15	0.03	0.67
Race and religion controls					
% white	0.50	0.65	0.67	-0.02	-0.53
% black	0.18	0.12	0.08	0.04	2.04**
% native American	0.01	0.01	0.02	-0.02	-1.03
% Asian	0.10	0.05	0.05	-0.01	-0.44
% Hispanic	0.20	0.16	0.15	0.01	0.32
% Muslim	1.31	0.81	0.75	0.05	0.20
Socioeconomic controls					
% below poverty level	15.71	15.82	13.69	2.14	1.94*
% unemployed	4.86	5.05	4.51	0.54	1.76*
Gini index	0.48	0.46	0.45	0.01	1.22
% uninsured	12.87	12.40	11.21	1.19	1.08
Log(Median household income)	11.00	10.91	10.99	-0.09	-1.57
% employed in agriculture	0.00	0.00	0.00	0.00	1.99$*$
% employed in IT	0.04	0.02	0.02	-0.00	-0.02
% employed in manufacturing	0.07	0.09	0.09	0.01	0.55
% employed in nontradable sector	0.23	0.26	0.27	-0.01	-0.62
% employed in construction/real estate	0.06	0.07	0.07	0.01	1.02
% employed in utilities	0.04	0.04	0.03	0.00	0.53
% employed in business services	0.29	0.25	0.24	0.01	0.35
% employed in other services	0.27	0.26	0.28	-0.02	-0.94
% adults with high school degree	21.76	25.99	25.77	0.22	0.13
% adults with graduate degree	16.15	13.08	14.34	-1.26	-0.64
Media controls					
% watching Fox News	0.25	0.26	0.26	-0.00	-0.13
% watching prime time TV	0.42	0.43	0.43	0.00	0.11
Election control					
Republican vote share, 2012	0.33	0.46	0.47	-0.02	-0.43
Crime controls					
Violent crime rate	0.01	0.00	0.00	0.00	0.02
Property crime rate	0.03	0.02	0.02	0.00	1.09
Geographical controls					
Population density	5192.27	1021.39	1998.35	-976.96	-0.91
Log(County area)	6.30	6.63	6.54	0.09	0.31
Distance from Austin, TX (in miles)	1775.99	1749.38	1626.64	122.74	0.68

Notes: This table plots the mean values of the control variables for the three types of counties relevant for the cross-sectional results: (1) counties with new SXSW followers in March 2007 *and* the pre-period; (2) counties with new SXSW followers in March 2007 but no new followers in the pre-period; and (3) counties with new SXSW followers in the pre-period but no new followers in March 2007. $t-stat$ reports the result from a simple t-test for the equality of means between the counties with the key identifying variation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

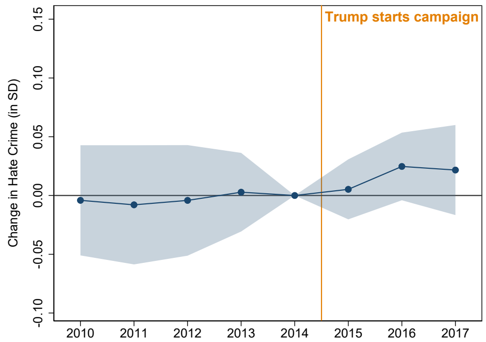
Table A.13: Balancedness SXSW Counties Individual Characteristics

First names (Corr. = 0.69)		**Terms used in bio (Corr. = 0.92)**	
Pre-Period	Treatment Period	Pre-Period	Treatment Period
michael	michael	http	http
mike	john	founder	com
paul	chris	com	digital
chris	jeff	co	founder
ryan	matt	tech	medium
eric	brian	design	director
david	david	director	tech
matthew	alex	product	music
john	jason	digital	social
jeff	kevin	designer	marketing
robert	paul	medium	design
mark	mike	music	co
andrew	dan	social	writer
daniel	andrew	love	love
james	peter	marketing	lover
kevin	jim	web	dad
jay	tom	geek	creative
jonathan	jennifer	writer	tweet
rob	steve	technology	author
rachel	todd	dad	designer

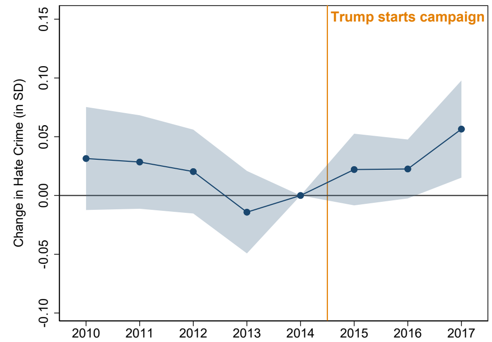
Notes: This table plots the ranking of the most common first names and terms used in a Twitter user's "bio" among users who follow "South by Southwest" on Twitter, depending on whether they signed up during the SXSW 2007 event or in the pre-period.

Figure A.5: Change in Other Hate Crimes, by Twitter Usage (OLS)

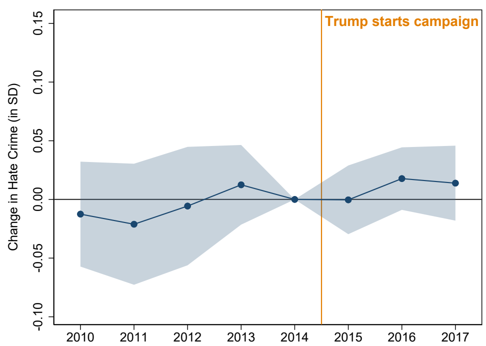
(a) All Hate Crimes



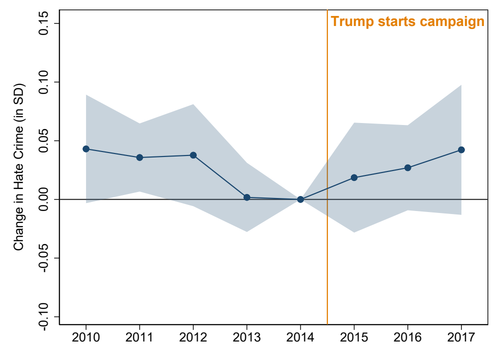
(b) Religious Hate Crimes



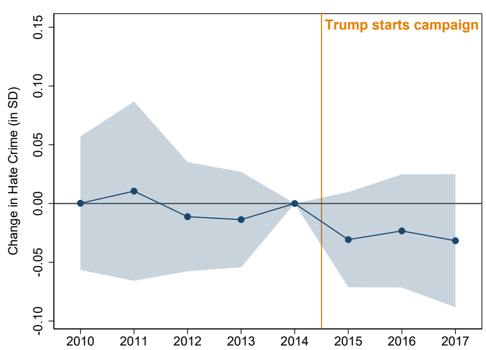
(c) Racial Hate Crimes



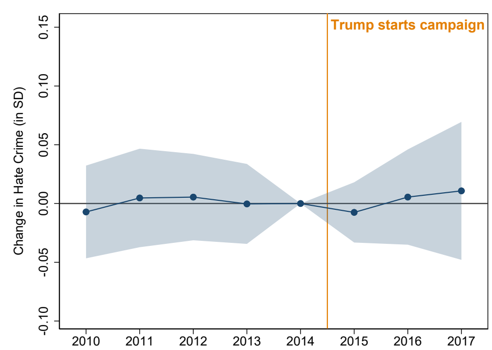
(d) Anti-Hispanic Hate Crimes



(e) Other Ethnic Hate Crimes



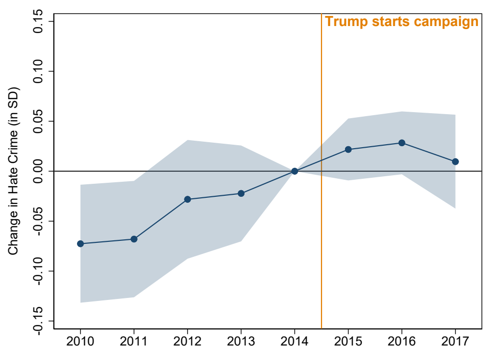
(f) Sexual Orientation Hate Crimes



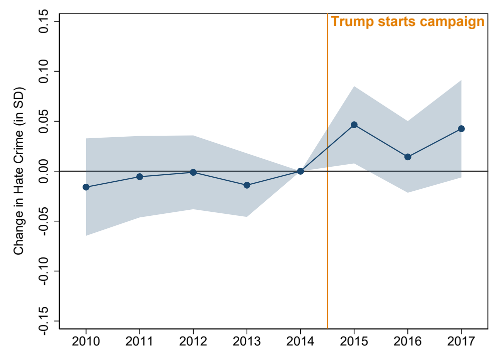
Notes: These figures plot the coefficients of running panel event study regressions as in Equation (1) for different types of hate crimes. We standardized the variables to have a mean of zero and standard deviation of one. The vertical line indicates the start of Trump's presidential campaign. The shaded areas are 95% confidence intervals. The excluded category is the year 2014.

Figure A.6: Change in Other Hate Crimes, by Twitter Usage (Reduced Form)

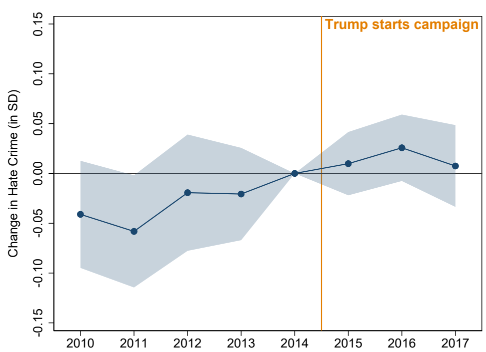
(a) All Hate Crimes



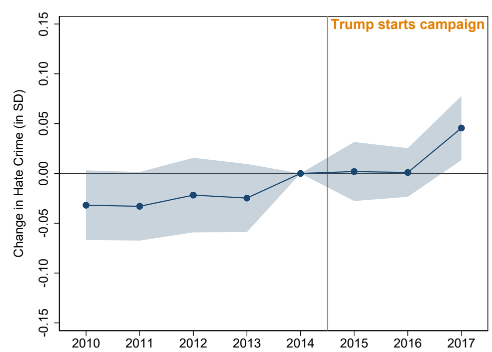
(b) Religious Hate Crimes



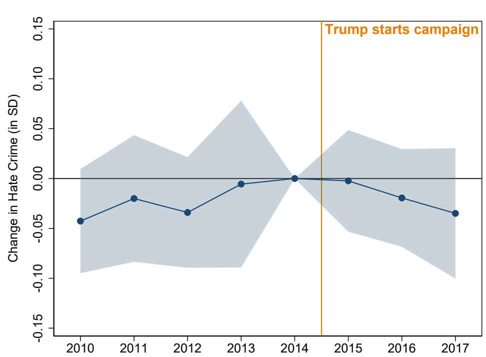
(c) Racial Hate Crimes



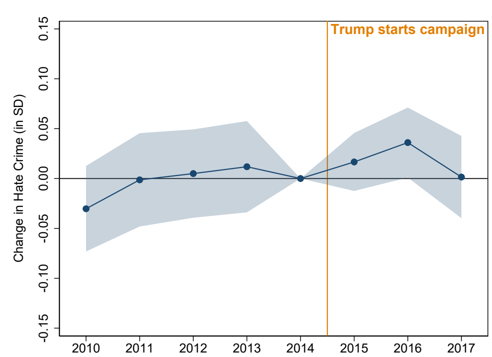
(d) Anti-Hispanic Hate Crimes



(e) Other Ethnic Hate Crimes



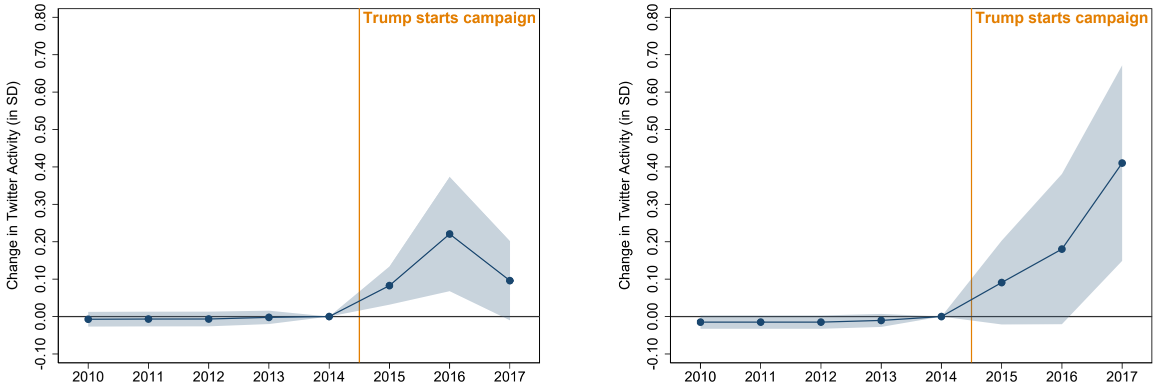
(f) Sexual Orientation Hate Crimes



Notes: These figures plot the coefficients of running panel event study regressions as in Equation (1) for different types of hate crimes, where $\log(Twitter\ usage)$ is replaced with $\log(SXSW\ followers, March$ 2007. We standardized the variables to have a mean of zero and standard deviation of one. The vertical line indicates the start of Trump's presidential campaign. The shaded areas are 95% confidence intervals. The excluded category is the year 2014.

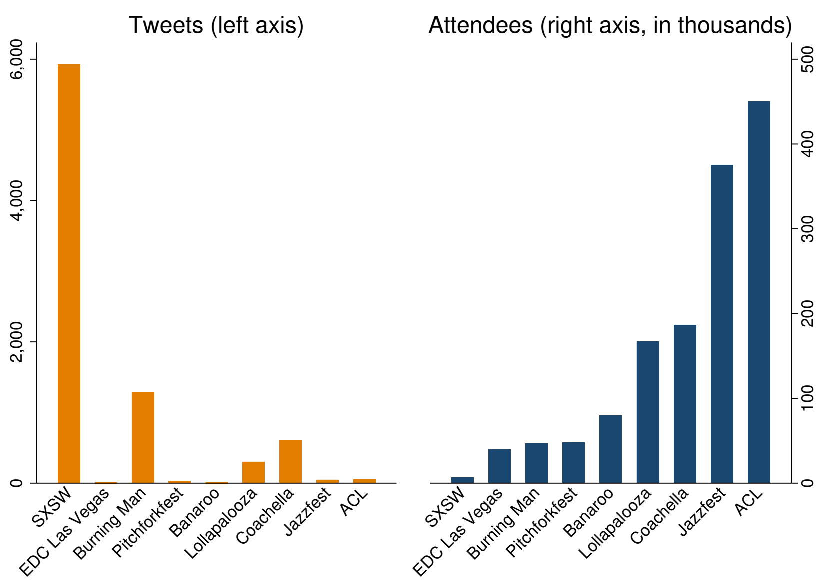
Figure A.7: Change in Anti-Muslim Tweets (Reduced Form)

(a) #StopIslam (b) #BansIslam



Notes: These figures plot the coefficients of running panel event study regressions as in Equation (1). The dependent variables are the log number of tweets containing the terms #BanIslam in panel (a) and #StopIslam in panel (b). We standardized the variables to have a mean of zero and standard deviation of one. The vertical line indicates the start of Trump's presidential campaign. The shaded areas are 95% confidence intervals. The excluded category is the year 2014.

Figure A.8: Number of Tweets and Attendees for Different Festivals (Full Year)



Notes: This figure plots the number of tweets mentioning major festivals in 2007.

Table A.14: Correlation of Log(Twitter Users) across Events

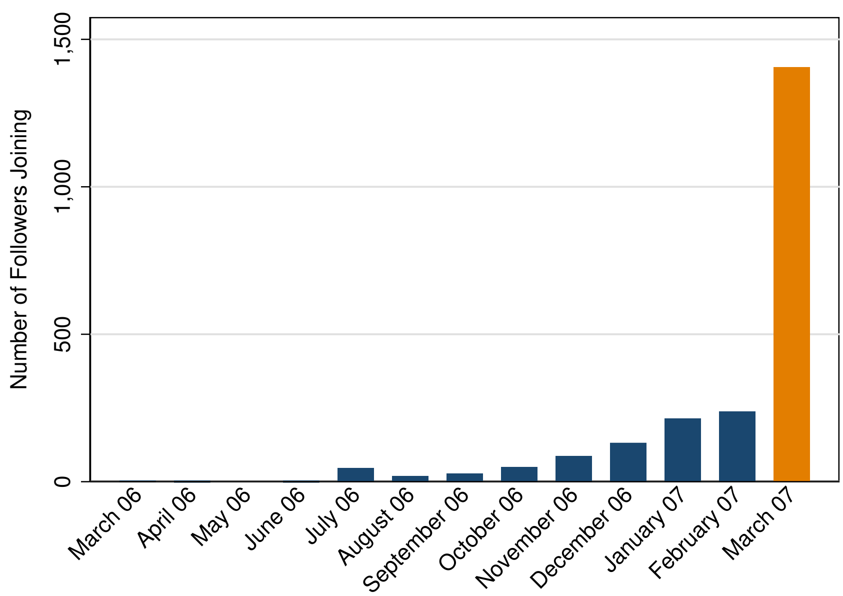
	SXSW March 2007	SXSW Pre	Coachella April 2007	Burning Man August 2007	Lollapalooza August 2007
SXSW followers, March 2007	1				
SXSW followers, Pre	0.77	1			
Coachella users, April 2007	0.44	0.48	1		
Burning Man users, August 2007	0.52	0.56	0.54	1	
Lollapalooza users, August 2007	0.03	0.06	0.00	0.00	1

Notes: This table reports the Pearson correlation coefficients between the main measure of interest (*SXSW followers, March 2007*) and different control variables. “Followers” are based on the locations of people who started following SXSW in a given month; “users” are based on people who tweeted at least once about a festival. We take the natural logarithm of these numbers with one added inside.

Table A.15: Number of Counties With Any Twitter Users at SXSW or Other Festivals

	SXSW March 2007	SXSW Pre	Coachella April 2007	Burning Man August 2007	Lollapalooza August 2007
No followers	2953	2987	3091	3098	3105
At least 1 follower	155	121	17	10	3

Figure A.9: Number of SXSW Followers Joining Each Month



Notes: This figure plots the number of SXSW followers who joined Twitter each month running up to the 2007 SXSW Festival. The orange bar marks the main instrument used in the paper.

Table A.16: Robustness - Twitter Penetration Controls Based on Other Festivals in 2007

	ΔLog(Hate crimes against Muslims)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS - Hate crimes against Muslims								
Log(Twitter usage)	0.021***	0.019***	0.019***	0.015***	0.015***	0.016***	0.015***	0.015***
	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Panel B: Reduced form - Hate crimes against Muslims								
Log(SXSW followers, March 2007)	0.081***	0.081***	0.091***	0.080***	0.076***	0.076***	0.076***	0.076***
	(0.025)	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Panel C: 2SLS - Hate crimes against Muslims								
Log(Twitter usage)	0.153***	0.168***	0.198***	0.189***	0.187***	0.194***	0.205***	0.210***
	(0.045)	(0.046)	(0.049)	(0.055)	(0.055)	(0.057)	(0.063)	(0.064)
Weak IV 95% AR confidence set	[0.06; 0.23]	[0.08; 0.25]	[0 .1; 0.28]	[0.08; 0.29]	[0.08; 0.28]	[0.08; 0.29]	[0.08; 0.33]	[0.09; 0.34]
Log(Burning Man users, August 2007)	-0.003	-0.003	0.031	-0.016	-0.021	-0.020	-0.009	0.008
	(0.083)	(0.084)	(0.084)	(0.078)	(0.077)	(0.077)	(0.075)	(0.072)
Log(Coachella users, April 2007)	0.007	-0.005	0.018	0.002	-0.005	-0.011	-0.014	-0.020
	(0.100)	(0.106)	(0.117)	(0.108)	(0.111)	(0.112)	(0.112)	(0.112)
Log(Lollapalooza users, August 2007)	0.263	0.261	0.251	0.243	0.240	0.241	0.240	0.234
	(0.187)	(0.198)	(0.194)	(0.197)	(0.191)	(0.196)	(0.194)	(0.194)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race and religion controls				Yes	Yes	Yes	Yes	Yes
Socioeconomic controls					Yes	Yes	Yes	Yes
Media controls						Yes	Yes	Yes
Election control							Yes	Yes
Crime controls								Yes
Geographical controls			Yes	Yes	Yes	Yes	Yes	Yes
Observations	3107	3107	3107	3107	3106	3105	3105	3105
Mean of DV	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
Robust F-stat.	158.32	117.39	101.21	112.09	87.87	80.59	67.50	66.25

Notes: This table presents county-level OLS, first stage, and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. The other variables count the number of users tweeting about any of the three largest US music festivals in 2007: Cochella, Burning Man, and Lollapalooza. All regressions control for population deciles, state fixed effects, and the full set of controls as in column 8 of Table 3 (not shown). Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.17: Robustness - Alternative Measures of Twitter Usage

	Survey # households using Twitter (1)	Survey % households using Twitter (2)	GESIS Tweets (Pre-Trump) (3)	GESIS Twitter users (4)
Panel A: OLS - Hate crimes against Muslims				
Twitter usage measure	0.059***	0.024**	0.017***	0.003**
	(0.020)	(0.010)	(0.006)	(0.001)
Panel B: First stage - Twitter usage				
Log(SXSW followers, March 2007)	0.440***	0.080***	0.443***	0.634***
	(0.041)	(0.018)	(0.061)	(0.157)
Panel C: 2SLS - Hate crimes against Muslims				
Twitter usage measure	0.169**	0.926**	0.167**	0.117**
	(0.067)	(0.387)	(0.072)	(0.057)
Weak IV 95% AR confidence set	[0.04; 0.29]	[0.28; 1.87]	[0.04; 0.31]	[0.03; 0.27]
Log(SXSW followers, Pre)	0.014	-0.021	0.008	-0.014
	(0.062)	(0.090)	(0.070)	(0.077)
Observations	3106	3106	3107	3107
Mean of DV	0.019	0.019	0.019	0.019
Robust F-stat.	114.10	20.59	53.15	16.35

Notes: This table presents county-level OLS, reduced form, and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *Twitter usage measure* is the measure listed in the top row, instrumented using the number of users who started following SXSW in March 2007 (in log with 1 added inside). *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006 (in log with 1 added inside). All regressions control for population deciles and state fixed effects, as well as demographic controls including population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.18: 2SLS - Alternative SXSW Controls

SXSW measure	Followers	**Followers**	Followers	Followers	Followers	Followers	Tweets	Tweets	Tweets	Tweets	Tweets
Transformation	Log	**Log**	Log	Log	Log	Log	Dummies	Dummies	Dummies	Dummies	Dummies
Control variable(s)	None	**Pooled**	Pooled	Individual	Individual	Individual	Pooled	Pooled	Individual	Individual	Individual
Control period	–	**2006**	2006-Feb. 2007	Feb. 2007	2006	2006-Feb. 2007	2006	2006-Feb. 2007	Feb. 2007	2006	2006-Feb. 2007
Control counties	–	**67**	121	59	67	121	109	154	55	109	154
Corr(March 2007, Control), average	–	**0.77**	0.83	0.72	0.49	0.54	0.45	0.44	0.42	0.28	0.30
	(1)	**(2)**	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Reduced form - Hate Crimes against Muslims											
SXSW measure, March 2007	0.089***	**0.074****	0.077**	0.088**	0.071**	0.070**	0.092***	0.103***	0.090***	0.064**	0.066**
	(0.021)	**(0.030)**	(0.037)	(0.033)	(0.028)	(0.035)	(0.026)	(0.026)	(0.028)	(0.027)	(0.027)
SXSW measure, control (linear combination)	–	**0.034**	0.019	0.002	-0.241	-0.237	0.032	-0.008	0.058	-0.106	-0.117
		(0.054)	(0.041)	(0.053)	(0.246)	(0.247)	(0.031)	(0.025)	(0.050)	(0.095)	0.107
Panel B: 2SLS - Hate Crimes against Muslims											
Log(Twitter usage)	0.167***	**0.161****	0.272**	0.189***	0.155**	0.204**	0.319***	0.344***	0.344***	0.297**	0.362**
	(0.036)	**(0.069)**	(0.131)	(0.061)	(0.062)	(0.099)	(0.099)	(0.102)	(0.102)	(0.131)	(0.173)
Weak IV 95% AR confidence set	[0.10; 0.23]	**[0.05; 0.30]**	[0.03; 0.59]	[0.06; 0.30]	[0.04; 0.28]	[0.02; 0.41]	[0.16; 0.56]	[0.18; 0.59]	[0.18; 0.59]	[0.08; 0.72]	[0.11; 1.15]
Observations	3,107	**3,107**	3,107	3,107	3,107	3,107	3,105	3,105	3,105	3,105	3,105
Robust F-stat.	165.7	**58.04**	16.67	48.02	76.74	34.02	24.34	26.59	26.59	10.63	7.257

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *Log(Twitter usage)* is instrumented using the measure described in the top rows; column 2 plots the baseline specification. *SXSW measure, control (linear combination)* is the estimate for the SXSW control variable. "Pooled" controls refer to one variable for the entire control period; "individual" to a vector of individual variables for each control period (e.g. one variable for March 2006, one variable for April 2006, etc.). For the case of individual controls, we plot the linear combinations of the coefficients and associated standard errors. In those cases, we also plot the *average* of the correlation of the individual controls with the March 2007 measure. "Followers" are based on Twitter users in a county that started following SXSW in a given month. "Tweets" are based on whether we can identify any user that tweeted about SXSW in a given month. All regressions control for population deciles, state fixed effects and demographic controls that include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. The specifications in columns 7 through 11 include the full vector of control variables. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.19: Social Media and Types of Hate Crimes

	Any (1)	Vandalism (2)	Theft (3)	Burglary (4)	Robbery (5)	Assault (6)
Panel A: OLS - Hate crimes against Muslims						
Log(Twitter usage)	0.019***	0.008	0.001*	0.001	0.001	0.018***
	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)	(0.006)
Panel B: Reduced form - Hate crimes against Muslims						
Log(SXSW followers, March 2007)	0.074**	0.031	0.003	0.007	0.000	0.067**
	(0.030)	(0.022)	(0.005)	(0.010)	(0.004)	(0.029)
Panel C: 2SLS - Hate crimes against Muslims						
Log(Twitter usage)	0.161**	0.068	0.007	0.014	0.001	0.146**
	(0.069)	(0.047)	(0.011)	(0.021)	(0.008)	(0.066)
Weak IV 95% AR confidence set	[0.04; 0.30]	[0.01; 0.15]	[0.01; 0.03]	[0.02; 0.05]	[0.01; 0.01]	[0.03; 0.28]
Log(SXSW followers, Pre)	0.008	0.036	-0.004	-0.016	0.017	0.016
	(0.069)	(0.051)	(0.008)	(0.017)	(0.021)	(0.060)
Observations	3107	3107	3107	3107	3107	3107
Mean of DV	0.019	0.008	0.000	0.000	0.001	0.014
Robust F-stat.	58.04	58.04	58.04	58.04	58.04	58.04

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the log change in hate crimes against Muslims of the type in the top row between 2010 and 2017. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Race and religion controls contains the share of people identifying as white, African American, Native American or Pacific Islander, Asian, Hispanic, or Muslim. Socioeconomic controls include the poverty rate, unemployment rate, local GINI index, the share of uninsured individuals, log median household income, the share of highschool graduates, the share of people with a graduate degree, as well as the employment shares in agriculture, information technology, manufacturing, nontradables, construction and real estate, utilities, business services, or other sectors. Media controls include the viewership share of Fox News, the cable TV spending to population ratio, and the prime time TV viewership to population ratio. Election control is the county-level vote share of the Republican party in 2012. Crime controls are the rates of violent or property crime from the FBI. Geographical controls include the linear distance from the SXSW festival location (Austin, Texas), population density, and the natural logarithm of county size. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.20: Social Media and Hate Crimes – Alternative Standard Errors

	Robust SE (1)	Bootstrap robust SE (2)	Bootstrap state cluster SE (3)	Spatial SE (4)
Panel A: OLS - Hate crimes against Muslims				
Log(Twitter usage)	0.019***	0.019***	0.019***	0.019***
	(0.005)	(0.005)	(0.006)	(0.005)
Panel B: Reduced form - Hate crimes against Muslims				
Log(SXSW followers, March 2007)	0.074***	0.074**	0.074***	0.074***
	(0.029)	(0.031)	(0.027)	(0.028)
Panel C: 2SLS - Hate crimes against Muslims				
Log(Twitter usage)	0.161**	0.161**	0.161**	0.161**
	(0.066)	(0.069)	(0.071)	(0.067)
Weak IV 95% AR confidence set	[0.05; 0.30]			
Log(SXSW followers, Pre)	0.008	0.008	0.008	0.008
	(0.057)	(0.057)	(0.077)	(0.064)
Observations	3107	3107	3107	3107
Mean of DV	0.019	0.019	0.019	0.019
Robust F-stat.	39.37	39.37	57.15	52.14

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the log change in hate crimes against Muslims between 2010 and 2017. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. All regressions control for population deciles and state fixed effects (not shown). Demographic controls include population growth between 2000 and 2016 as well as age cohort controls for the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50. Spatial standard errors are based on the method proposed in Colella et al. (2019), implemented in Stata as *acreg*, using a 200 miles cutoff. For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Standard errors are computed as indicated in the top row. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.21: Heterogeneous Effects – Hate Groups and Hate Crimes

Dependent variable: Log(Anti-Muslim hate crimes)	(1) No hate groups	(2) Any hate group	(3) Few hate crimes	(4) Many hate crimes
Panel A: OLS				
Log(Twitter Usage) x Year=2010	-0.01	-0.01	-0.00	-0.07
	(0.01)	(0.09)	(0.00)	(0.11)
Log(Twitter Usage) x Year=2011	-0.00	0.01	0.00	0.01
	(0.01)	(0.11)	(0.00)	(0.13)
Log(Twitter Usage) x Year=2012	-0.00	-0.02	0.00	-0.02
	(0.01)	(0.14)	(0.00)	(0.15)
Log(Twitter Usage) x Year=2013	-0.01	-0.00	0.00	-0.04
	(0.01)	(0.11)	(0.00)	(0.13)
Log(Twitter Usage) x Year=2015	0.01	0.45***	0.00	0.52***
	(0.01)	(0.14)	(0.00)	(0.15)
Log(Twitter Usage) x Year=2016	0.01	0.58***	0.01**	0.63***
	(0.01)	(0.17)	(0.00)	(0.18)
Log(Twitter Usage) x Year=2017	-0.01	0.38	0.00	0.34
	(0.01)	(0.23)	(0.00)	(0.25)
Panel B: Reduced form				
Log(SXSW followers) x Year=2010	-0.07**	-0.01	-0.00	-0.03
	(0.03)	(0.04)	(0.00)	(0.03)
Log(SXSW followers) x Year=2011	-0.04*	0.00	-0.00	0.00
	(0.02)	(0.03)	(0.00)	(0.03)
Log(SXSW followers) x Year=2012	-0.03	-0.02	0.00	-0.02
	(0.02)	(0.03)	(0.01)	(0.03)
Log(SXSW followers) x Year=2013	-0.05*	0.02	-0.00	0.01
	(0.03)	(0.03)	(0.00)	(0.03)
Log(SXSW followers) x Year=2015	-0.01	0.03	-0.00	0.10***
	(0.03)	(0.03)	(0.00)	(0.03)
Log(SXSW followers) x Year=2016	0.02	0.09*	-0.01	0.14***
	(0.03)	(0.05)	(0.01)	(0.04)
Log(SXSW followers) x Year=2017	-0.01	0.06*	-0.00	0.13***
	(0.03)	(0.03)	(0.01)	(0.05)
County FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Pop. deciles x Year FE	Yes	Yes	Yes	Yes
Observations	1145248	147680	1156896	136032

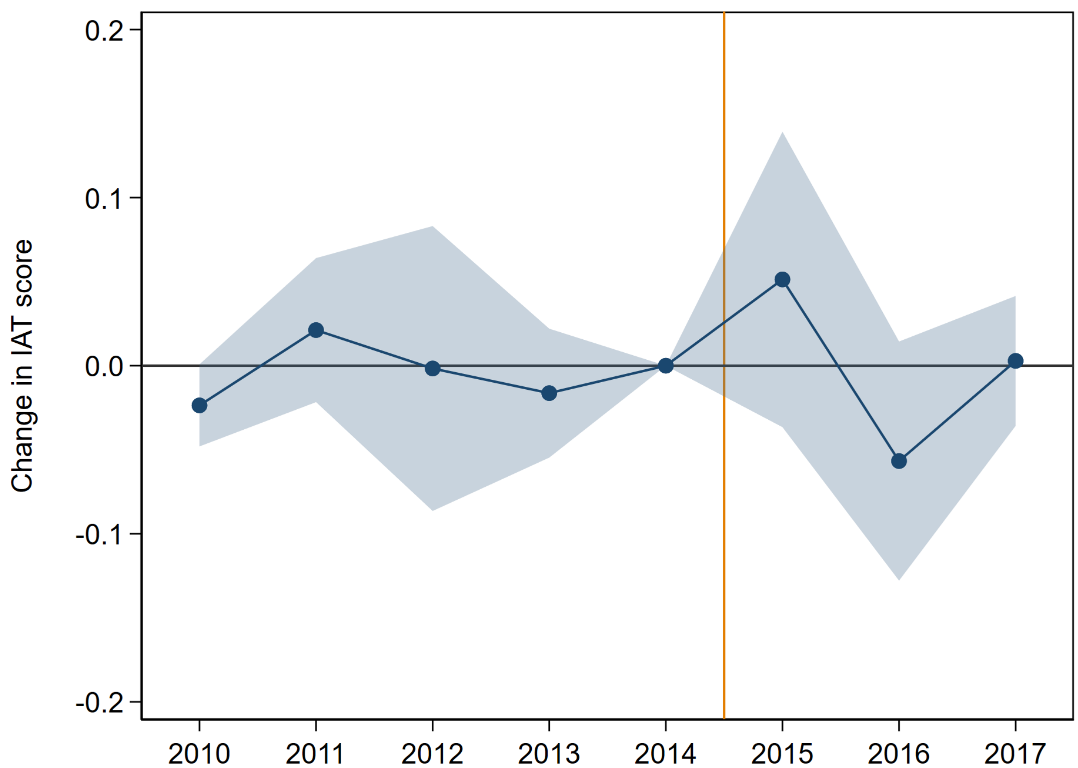
Notes: This table presents panel event study regressions where the dependent variable is the log number of hate crimes against Muslims (with one added inside). We standardized the variables to have a mean of zero and standard deviation of one. The sample period is 2010 to 2017. 2014 is the excluded period. $Log(SXSW followers)$ is the number of local SXSW followers that joined Twitter in March 2007. The existence of hate groups is based on data from the Southern Poverty Law Center (SPLC). The number of hate crimes in the pre-period is based on the total number of hate crimes per capita the FBI registered in a county from 2010 until 2015, split at the 90th percentile. All regressions control for the interaction of population deciles with year dummies. Standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.22: Social Media and Changes in Implicit Bias against Muslims

	Raw IAT scores (1)	Residual IAT scores (2)	Only conservatives (3)	Only whites (4)	Only Christians (5)	Only non-Muslims (6)	Only obligatory tests (7)	At least 10 tests (8)
Panel A: OLS - Change in implicit bias against Muslims								
Log(Twitter usage)	0.026*	0.023	-0.023	-0.012	0.002	0.021	0.017	0.004
	(0.014)	(0.014)	(0.022)	(0.014)	(0.012)	(0.014)	(0.021)	(0.006)
Panel B: Reduced form - Change in implicit bias against Muslims								
Log(SXSW followers, March 2007)	-0.016	-0.014	-0.023	-0.012	-0.027	-0.017	-0.003	0.006
	(0.016)	(0.014)	(0.035)	(0.022)	(0.020)	(0.016)	(0.022)	(0.008)
Panel C: 2SLS - Change in implicit bias against Muslims								
Log(Twitter usage)	-0.043	-0.039	-0.061	-0.035	-0.077	-0.048	-0.007	0.017
	(0.046)	(0.039)	(0.096)	(0.066)	(0.064)	(0.048)	(0.058)	(0.024)
Weak IV 95% AR confidence set	[-0.14; 0.03]	[-0.11; 0.02]	[-0.27; 0.11]	[-0.18; 0.07]	[-0.21; 0.02]	[-0.15; 0.03]	[-0.12; 0.09]	[-0.02; 0.06]
Log(SXSW followers, Pre)	0.024	0.011	-0.036	-0.027	0.040	0.020	-0.015	-0.001
	(0.019)	(0.017)	(0.051)	(0.029)	(0.028)	(0.019)	(0.025)	(0.017)
Observations	2251	2222	1303	1945	1987	2230	1759	571
Mean of DV	-0.038	-0.013	-0.007	-0.053	-0.032	-0.039	-0.039	-0.044
Robust F-stat.	49.42	51.78	34.80	36.57	36.79	38.15	64.48	28.31

Notes: This table presents county-level OLS and IV regressions where the dependent variable is the change in average Implicit Association Test (IAT) scores that measures implicit bias against Muslims between 2010 and 2017. Higher scores reflect more bias. *Log(Twitter usage)* is instrumented using the number of users who started following SXSW in March 2007. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006. All regressions control for population deciles and state fixed effects (not shown). In column 2, IAT scores are residualized with respect to age and its squared term, as well as a full set of fixed effects for educational attainment, race, sex, and ethnicity. In columns 3 through 6, the sample is restricted to respondents as indicated in the top row. Column 7 only includes tests that are obligatory, e.g. as part of a work program. Column 8 restricts the sample to counties with at least 10 IAT tests before and after Trump's presidential run. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) using the Stata package from Sun (2018). For the just-identified case we study here, the "robust" F-stat. is equivalent to the "Kleibergen-Paap" or the "effective" F-statistic of Olea & Pflueger (2013). Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

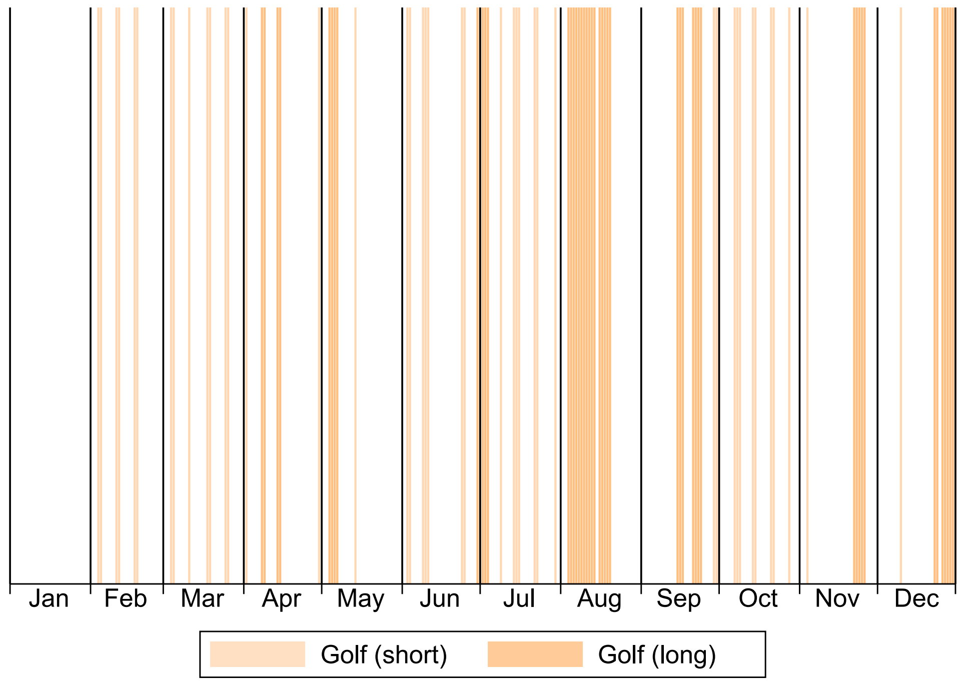
Figure A.10: Change in Implicit Bias (Reduced Form)



Notes: These figures plot the coefficients of running a panel event study regression as in Equation (1). The dependent variable is the mean county-level IAT score that measures implicit bias against Muslims. We standardize the variables to have a mean of zero and standard deviation of one. The vertical line indicates the start of Trump's presidential campaign. The shaded areas are 95% confidence intervals. The excluded category is the year 2014.

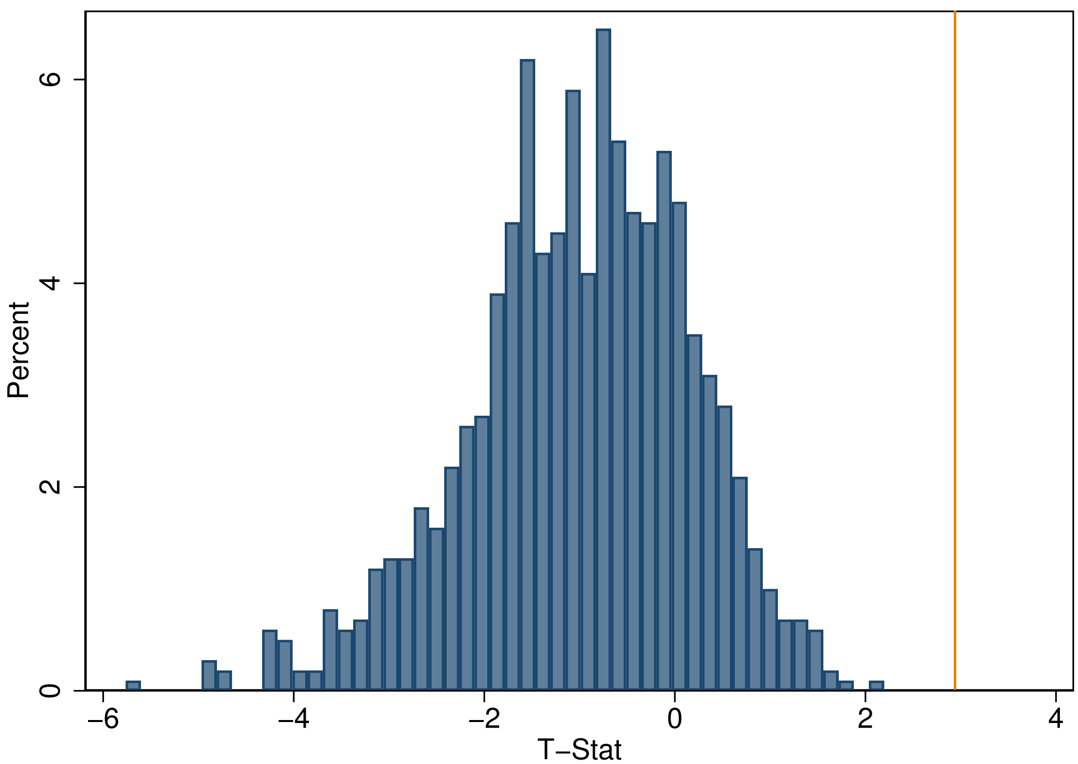
A.4. Appendix 4: Additional Time Series Evidence

Figure A.11: Trump's Golf Days in 2017



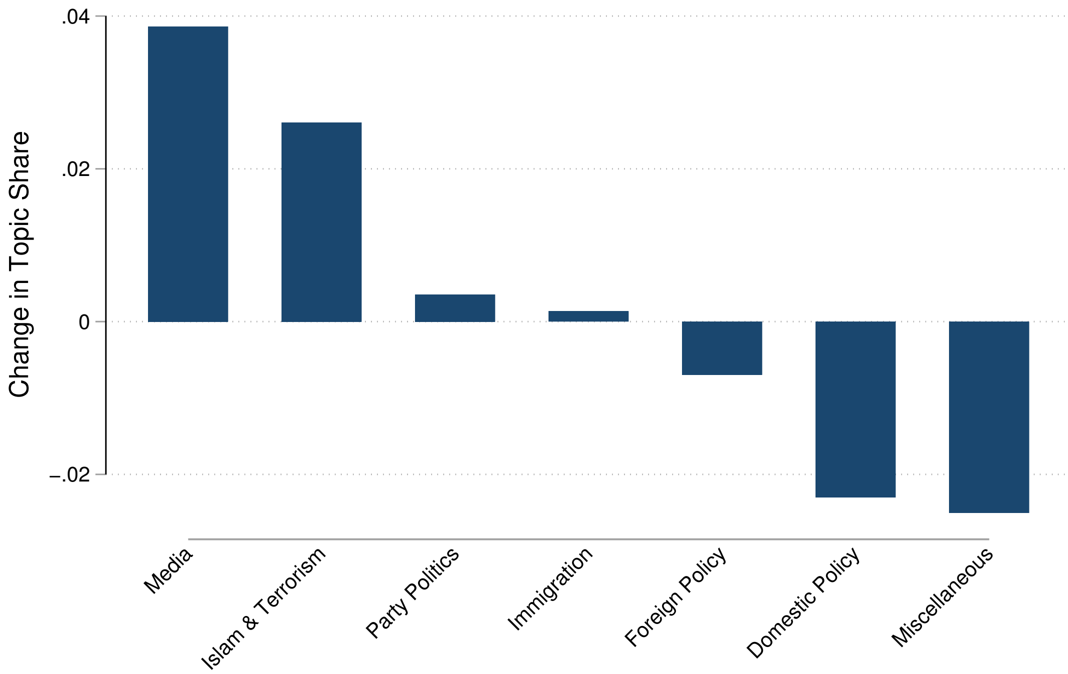
Notes: This figure plot the days in 2017 when Donald Trump played golf. Golf (long) indicates three or more consecutive days of golf.

Figure A.12: Randomization Test for Golf Days



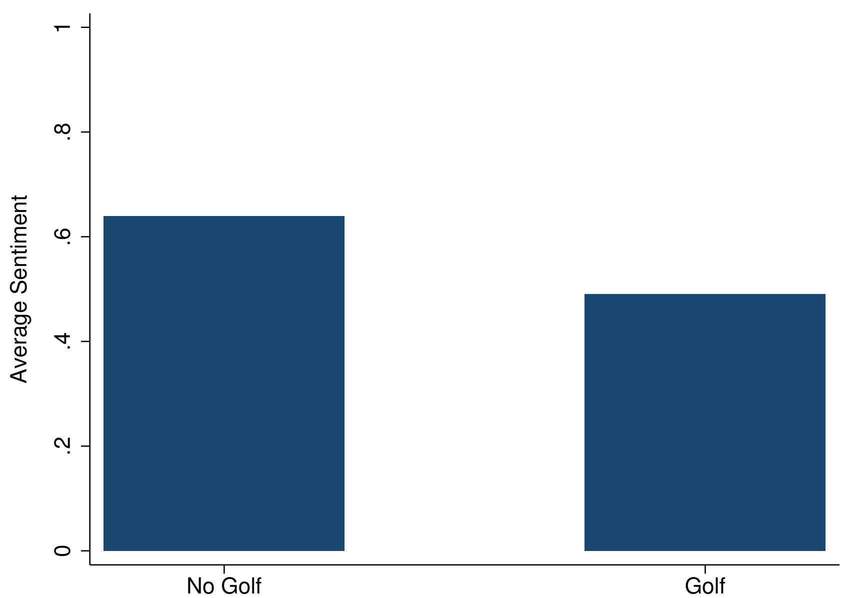
Notes: This figure visualizes the distribution of *t*-statistics from a randomization test of the first stage regression of Trump's tweets about Muslims on placebo golf days. In particular, we create 1,000 placebo sets of 92 golf days, which is the number of times Trump golfed in 2017. We then regress the log number of Trump's tweets about Muslims on these dummies using the baseline specification in Equation (4) and report the distribution of the resulting *t*-statistics. The orange line marks our baseline point estimate.

Figure A.13: Shift in Topics of Trump's Tweets on Golf Days



Notes: This figure shows how the content of Trump's tweets changes on days when he plays golfs. These topics were hand-coded using Amazon Mechanical Turk.

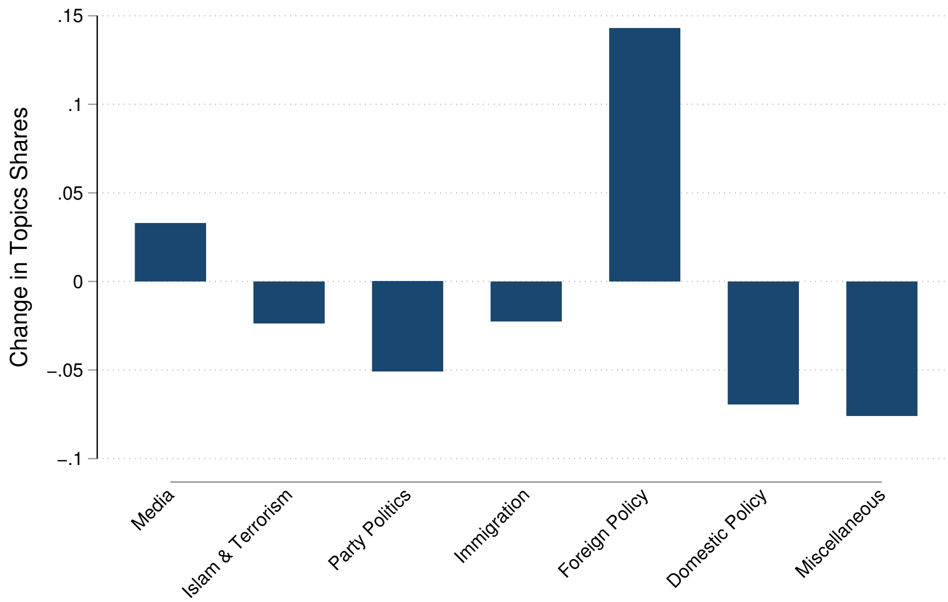
Figure A.14: Trump's Tweets Are More Negative on Golf Days



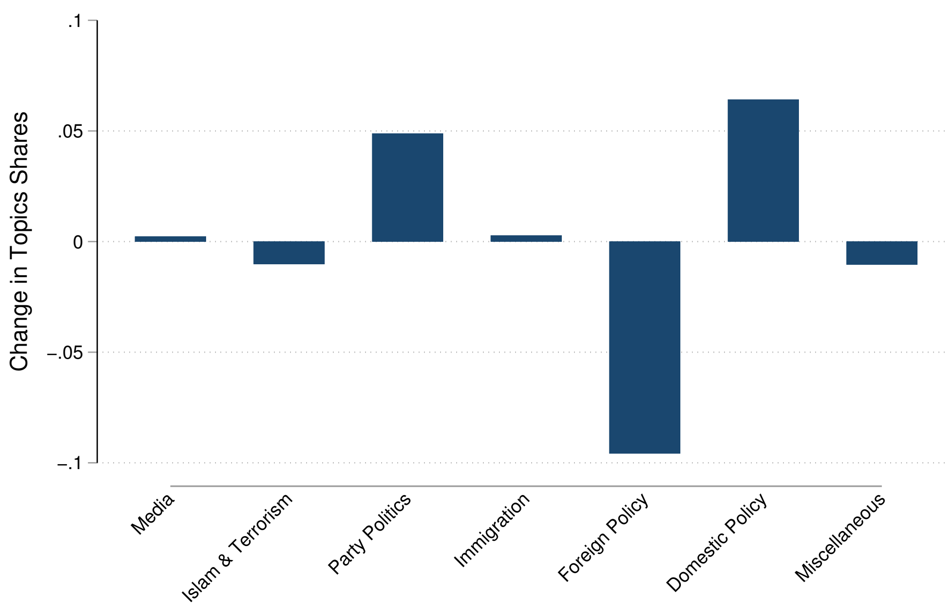
Notes: This figure plots the average sentiment of Trump's tweets on golf and non-golf days. Lower values mean more negative sentiment. The sentiment was hand-coded using Amazon Mechanical Turk on a scale from -2 to 2.

Figure A.15: Shift in Topics of Trump's Tweets During Other Events

(a) Travel Abroad



(b) Policy Briefing



Notes: This figure shows how the content of Trump's tweets changes on days when he is traveling abroad (panel a) or receives a policy briefing (panel b). These topics were hand-coded using Amazon Mechanical Turk.

Table A.23: Summary Statistics for Time Series

Variable	Mean	SD	p50	Min	Max	N
Trump tweets						
Muslim Trump tweets (1+log)	0.08	0.25	0.00	0.00	1.79	365
Total Trump tweets (1+log)	1.95	0.58	1.95	0.00	3.30	365
Muslim Trump tweets (dummy)	0.09	0.29	0.00	0.00	1.00	365
Hate crimes against Muslims (1 + natural logarithm)						
All types	0.45	0.47	0.69	0.00	1.79	365
Assault	0.31	0.42	0.00	0.00	1.61	365
Vandalism	0.15	0.30	0.00	0.00	1.39	365
Theft	0.01	0.09	0.00	0.00	1.10	365
Burglary	0.01	0.07	0.00	0.00	0.69	365
Robbery	0.01	0.09	0.00	0.00	0.69	365
Other hate crimes (1 + natural logarithm)						
All hate crimes	2.99	0.27	3.00	2.08	3.74	365
Ethnicity (incl. Hispanic)	0.44	0.47	0.69	0.00	1.79	365
Race	2.27	0.37	2.30	0.69	3.00	365
Sexual orientation	1.32	0.46	1.39	0.00	2.40	365
Religion (excl. Muslims)	1.40	0.50	1.39	0.00	2.89	365
TV news coverage (1 + natural logarithm)						
Muslim mentions (total)	3.71	0.64	3.69	0.69	5.26	365
Muslim mentions (Fox News)	2.75	0.66	2.77	0.00	4.29	365
Muslim mentions (CNN)	2.24	0.94	2.30	0.00	4.29	365
Muslim mentions (MSNBC)	2.75	0.66	2.77	0.00	4.26	365
Trump's golfing						
Trump golfs	0.25	0.43	0.00	0.00	1.00	365
Trump golfs (NYT only)	0.24	0.43	0.00	0.00	1.00	365
Trump golf (alternative)	0.25	0.44	0.00	0.00	1.00	365
Golf holiday	0.16	0.37	0.00	0.00	1.00	365
Golf at any point in previous week	0.71	0.45	1.00	0.00	1.00	365
Other control variables						
Google searches (PC)	-0.19	1.59	-0.48	-1.47	11.94	365
Terror attack in the US	0.00	0.05	0.00	0.00	1.00	365
Terror attack in Europe	0.03	0.17	0.00	0.00	1.00	365
Terror attack elsewhere	0.08	0.28	0.00	0.00	2.00	365

Notes: This table presents descriptive statistics for the IV sample. The sample year is 2017, for which we have information on Trump's golfing. *1+log* or *1+natural logarithm* means that the logarithm of any variable is calculated with 1 added inside. The data on hate crimes come from the FBI hate crime statistics. Data on Trump's golfing come from the New York Times, the official White House presidential schedule, and trumpgolfcount.com. *Google searches (PC)* is the first principal component of Google trends for the key words "islam", "mosque", "muslim", "refugee", "sharia", and "terror". We use these same keywords as measures of TV news attention based on data from the internet archive. The sources for the number of terror attacks is the Global Terrorism Database. See the online appendix for more details on data and variable construction.

Table A.24: Summary Statistics by Day of Week (2017 only)

Day of week		Hate crimes against Muslims	Tweets about Muslims	Trump golfs
Monday	Sum	43	3	4
	Mean	0.83	0.06	0.08
Tuesday	Sum	33	6	3
	Mean	0.63	0.12	0.06
Wednesday	Sum	43	10	4
	Mean	0.83	0.19	0.08
Thursday	Sum	43	6	6
	Mean	0.83	0.12	0.12
Friday	Sum	36	12	13
	Mean	0.69	0.23	0.25
Saturday	Sum	36	4	30
	Mean	0.69	0.08	0.58
Sunday	Sum	42	6	32
	Mean	0.79	0.11	0.60
Total	Sum	276	47	92
	Mean	0.76	0.13	0.25

Notes: This table presents descriptive statistics by day of week for the number of anti-Muslim hate crimes, the number of Trump's tweets about Muslims and the number of Trump's golf outing for the sample used in the instrumental variable regressions (2017 only).

Table A.25: Robustness Time Series Regressions

	Baseline (1)	Add 7 lagged dependent variables (2)	Add golf holiday control (3)	Add previous week golf control (4)	Use Trump Tweet dummy (5)	Use only NYT golf count (6)	Use alternative golf count (7)
Panel A: OLS - Log(Hate crimes against Muslims) in t+2							
Log(Muslim Trump tweets)	0.130*	0.148**	0.128*	0.127*	0.106	0.130*	0.130*
	(0.069)	(0.069)	(0.069)	(0.069)	(0.074)	(0.069)	(0.069)
Panel B: First Stage - Log(Trump tweets about Muslims)							
Trump golfs	0.102***	0.098***	0.129***	0.094***	0.118***	0.095***	0.098***
	(0.027)	(0.027)	(0.031)	(0.027)	(0.033)	(0.028)	(0.027)
Panel C: Reduced form - Log(Hate crimes against Muslims) in t+2							
Trump golfs	0.165**	0.164**	0.163**	0.163**	0.165**	0.168**	0.155**
	(0.071)	(0.080)	(0.078)	(0.072)	(0.071)	(0.068)	(0.071)
Panel D: 2SLS - Log(Hate crimes against Muslims) in t+2							
Log(Muslim Trump tweets)	1.617**	1.682*	1.269**	1.631**	1.398*	1.764**	1.571*
	(0.779)	(0.935)	(0.633)	(0.821)	(0.716)	(0.824)	(0.809)
Weak IV 95% AR confidence set	[0.31; 4.01]	[0.29; 4.55]	[0.20; 2.96]	[0.27; 4.9]	[0.34; 3.74]	[0.54; 4.62]	[0.21; 4.05]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	358	363	363	363	363	363
R^2	0.213	0.209	0.339	0.207	0.204	0.149	0.231
Robust F-stat.	13.15	12.24	15.95	13.10	11.76	10.61	12.61

Notes: This table presents OLS and IV regressions where the dependent variable is the number of hate crimes against Muslims on any given day based on FBI data. We use a dummy for days on which President Donald Trump golfs used as an instrument for his tweets about Muslims. Column 2 controls for seven lags of the dependent variable. Column 3 controls for golf days that are part of a golf "holiday", which we define as Trump golfing for more than three consecutive days. Column 4 controls for whether Trump golfed in the previous week. Column 5 replaces the number of Muslim Trump tweets with a dummy for whether Trump sends any tweet about Muslims. Column 6 replaces the main measure $Trump\ golfs$ with one that only uses information from the New York Times (ignoring that contained in his presidential schedule). Column 7 uses an alternative golf count that incorporates information from $trumpgolfcount.com$. Column 8 presents an alternative specification where we cluster standard errors by week (ignoring serial correlation). The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as a dummy for whether Trump's golfing is the first of a series of golf days. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses except in column 8. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.26: Time Series - Split by Type of Hate Crime

	Any (1)	Vandalism (2)	Theft (3)	Burglary (4)	Robbery (5)	Assault (6)
Panel A: OLS - Log(Hate crimes against Muslims) in t+2						
Log(Muslim Trump tweets)	0.130*	0.023	0.023	0.093**	0.011	0.033
	(0.069)	(0.053)	(0.033)	(0.042)	(0.014)	(0.061)
Panel B: Reduced form - Log(Hate crimes against Muslims) in t+2						
Trump golfs	0.165**	0.139**	-0.003	0.022	-0.007	0.075
	(0.071)	(0.057)	(0.014)	(0.016)	(0.013)	(0.069)
Panel C: 2SLS - Log(Hate crimes against Muslims) in t+2						
Log(Muslim Trump tweets)	1.617**	1.363**	-0.033	0.216	-0.065	0.741
	(0.779)	(0.629)	(0.132)	(0.148)	(0.131)	(0.692)
Weak IV 95% AR confidence set	[0.31; 4.01]	[0.30; 3.29]	[-0.31; 0.27]	[-0.09; 0.58]	[-0.44; 0.16]	[-0.56; 2.59]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363
R^2	0.213	-0.697	0.032	-0.026	0.004	0.291
Robust F-stat.	13.15	13.15	13.15	13.15	13.15	13.15

Notes: This table presents OLS and IV regressions where the dependent variable is the number of hate crimes against Muslims on any given day based on FBI data. We use a dummy for days on which President Donald Trump golfs used as an instrument for his tweets about Muslims. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as a dummy for whether Trump's golfing is the first of a series of golf days. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.27: Robustness Time Series Regressions - Timing of Effect

	t-1	t	t+1	**Baseline** **t+2**	t+3	t+4	t+5	t+6	t+7
	(1)	(2)	(3)	**(4)**	(5)	(6)	(7)	(8)	(9)
Panel A: OLS - Log(Hate crimes against Muslims)									
Log(Muslim Trump tweets)	0.112	0.008	0.084	**0.192****	-0.126*	-0.036	-0.162*	-0.047	0.030
	(0.100)	(0.111)	(0.102)	**(0.077)**	(0.075)	(0.100)	(0.085)	(0.081)	(0.093)
Panel B: Reduced form - Log(Hate crimes against Muslims)									
Trump golfs	0.079	0.048	0.077	**0.165****	0.097	0.085	-0.022	0.149**	0.144***
	(0.064)	(0.071)	(0.074)	**(0.071)**	(0.081)	(0.073)	(0.066)	(0.065)	(0.054)
Panel C: 2SLS - Log(Hate crimes against Muslims)									
Log(Muslim Trump tweets)	0.774	0.472	0.759	**1.617****	1.059	0.912	-0.224	1.500**	1.450**
	(0.642)	(0.648)	(0.725)	**(0.779)**	(0.854)	(0.729)	(0.682)	(0.729)	(0.692)
Weak IV 95% AR confidence set	[-0.43; 2.49]	[-1.13; 1.69]	[-0.60; 2.55]	**[0.31; 4.01]**	[-0.72; 3.34]	[-0.75; 2.72]	[-2.18; 1.06]	[0.27; 3.88]	[0.42; 3.99]
Fixed effects (month, day of week)	Yes	Yes	Yes	**Yes**	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	**Yes**	Yes	Yes	Yes	Yes	Yes
Observations	363	364	364	**363**	362	361	360	359	358
R^2	0.430	0.482	0.469	**0.213**	0.324	0.399	0.519	0.181	0.222
Robust F-stat.	13.08	13.02	13.02	**13.15**	9.467	9.876	10.28	10.65	10.62

Notes: This table presents OLS and IV regressions where the dependent variable is the number of hate crimes against Muslims on any given day based on FBI data. Each column presents the results from a different regression, where the dependent variable is defined for the period in the top column. Column 4 is equivalent with column 6 in Table 6. We use a dummy for days on which Trump golfs used as an instrument for his tweets about Muslims. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends, dummies for terror attacks in the US, Europe or the rest of the world, as well as a dummy for whether Trump's golfing is the first of a series of golf days. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.28: Robustness Controls

	Baseline (1)	Lagged dependent variable (2)	Federal holiday control (3)	Google search control (4)	Terror attack control (5)	Total tweets control (6)
Panel A: OLS - Log(Total number of Muslim TV mentions) in t+1						
Log(Muslim Trump tweets)	0.700***	0.301***	0.702***	0.631***	0.575***	0.700***
	(0.095)	(0.073)	(0.095)	(0.088)	(0.100)	(0.092)
Panel B: Reduced form - Log(Total number of Muslim TV mentions) in t+1						
Trump golfs	0.299**	0.142**	0.296**	0.311**	0.278**	0.297**
	(0.131)	(0.070)	(0.131)	(0.123)	(0.119)	(0.128)
Panel C: 2SLS - Log(Total number of Muslim TV mentions) in t+1						
Log(Muslim Trump tweets)	2.958***	2.108*	2.869***	3.028***	3.276**	3.042***
	(1.014)	(1.136)	(0.995)	(0.941)	(1.433)	(1.082)
Weak IV 95% AR confidence set	[0.85; 5.27]	[0.20; 6.27]	[0.80; 5.13]	[1.07; 5.36]	[0.91; 7.70]	[0.79; 5.72]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	364	363	364	364	364	364
R^2	0.961	0.976	0.963	0.960	0.956	0.960
Robust F-stat.	13.02	8.928	13.39	13.39	10.77	12.05

Notes: This table presents OLS and IV regressions where the dependent variable is the number of times Muslims are mentioned on TV on a given day. We use a dummy for days on which Trump golfs used as an instrument for his tweets about Muslims. The results for *Total Coverage* shown here are based on Fox News, CNN, and MSNBC. The results for the individual channels are available upon request. Column 2 controls for one lag of the dependent variable and column 3 for a dummy that tags federal holidays. Column 4 controls for the first principal component of Google searches for Islam-related terms. Column 5 controls for the number of terror attacks in the US, Europe, or other countries. Column 6 controls for the total number of tweets by Trump. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as a dummy for whether Trump's golfing is the first of a series of golf days. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.29: Summary Statistics for Time Series – Split at Campaign Announcement

	Before campaign announcement						After campaign announcement					
	Mean	SD	p50	Min	Max	N	Mean	SD	p50	Min	Max	N
Trump tweets												
Muslim Trump tweets (1+log)	0.03	0.16	0.00	0.00	1.39	2234	0.10	0.30	0.00	0.00	1.79	930
Total Trump tweets (1+log)	1.56	1.37	1.61	0.00	5.00	2234	2.27	0.72	2.30	0.00	4.54	930
Muslim Trump tweets (dummy)	0.04	0.20	0.00	0.00	1.00	2234	0.11	0.31	0.00	0.00	1.00	930
Hate crimes (1 + natural logarithm)												
Muslims	0.26	0.39	0.00	0.00	1.61	2234	0.47	0.48	0.69	0.00	1.95	930
All hate crimes	2.84	0.31	2.89	1.10	3.61	2234	2.89	0.30	2.89	1.79	3.76	930
Ethnicity (incl. Hispanic)	0.51	0.48	0.69	0.00	2.30	2234	0.40	0.45	0.00	0.00	2.08	930
Race	2.13	0.40	2.20	0.00	3.14	2234	2.17	0.40	2.20	0.69	3.04	930
Sexual orientation	1.35	0.51	1.39	0.00	2.56	2234	1.28	0.50	1.39	0.00	2.40	930
Religion (excl. Muslims)	1.22	0.55	1.39	0.00	2.71	2234	1.26	0.53	1.39	0.00	2.89	930
Other control variables												
Terror attack in the US	0.00	0.02	0.00	0.00	1.00	2234	0.01	0.07	0.00	0.00	1.00	930
Terror attack in Europe	0.00	0.04	0.00	0.00	1.00	2234	0.04	0.19	0.00	0.00	1.00	930
Terror attack elsewhere	0.02	0.14	0.00	0.00	3.00	2234	0.15	0.43	0.00	0.00	3.00	930

Notes: This table presents descriptive statistics for the OLS sample. The sample is split into the period before and after June 16, 2015 when Trump announced his presidential campaign. *1+log* or *1+natural logarithm* means that the logarithm of any variable is calculated with 1 added inside. The data on hate crimes come from the FBI hate crime statistics. The sources for the number of terror attacks is the Global Terrorism Database. See the online appendix for more details on data and variable construction.

Table A.30: Time Series Regression Full Period

	Baseline (1)	Add lagged dependent variable (2)	Add terror attack control (3)	Add total tweets control (4)	Use Trump Tweet dummy (5)
Panel A: Before campaign announcement					
Log(Muslim Trump tweets)	0.017	0.018	0.019	0.015	0.053
	(0.018)	(0.018)	(0.018)	(0.019)	(0.098)
Observations	2,234	2,232	2,233	2,234	2,234
R^2	0.026	0.027	0.028	0.026	0.026
Panel B: After campaign announcement					
Log(Muslim Trump tweets)	0.108**	0.104***	0.090**	0.094**	0.307**
	(0.042)	(0.039)	(0.041)	(0.041)	(0.132)
Observations	930	928	929	930	930
R^2	0.079	0.082	0.092	0.082	0.077
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS regressions where the dependent variable is the number of hate crimes against the group in the top row on any given day based on FBI data. The sample is split into the period before and after June 16, 2015 when Trump announced his presidential campaign. All regressions include day-of-week and year-month dummies as well as linear and quadratic time trends. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.31: Time Series - Split by Motivating Bias

	All (1)	Hispanic (2)	Other Ethnicity (3)	Race (4)	Sexual Orientation (5)	Religion (excl. Muslims) (6)
Panel A: OLS - Log(Hate crimes)						
Log(Muslim Trump tweets), t-1	0.077*	0.007	0.247***	0.004	0.009	0.090
	(0.047)	(0.076)	(0.095)	(0.067)	(0.076)	(0.074)
Panel B: Reduced form - Log(Hate crimes)						
Trump golfs	0.013	-0.095	-0.017	0.043	0.024	0.011
	(0.045)	(0.084)	(0.084)	(0.065)	(0.070)	(0.069)
Panel C: 2SLS - Log(Hate crimes)						
Log(Muslim Trump tweets), t-1	0.105	-0.739	-0.129	0.333	0.183	0.089
	(0.330)	(0.619)	(0.636)	(0.482)	(0.530)	(0.516)
Weak IV 95% AR confidence set	[-0.71; 0.73]	[-2.15; 0.55]	[-1.70; 1.07]	[-0.67; 1.34]	[-0.92; 1.39]	[-1.19; 1.06]
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363
R^2	0.993	0.618	0.479	0.977	0.898	0.901
Robust F-stat.	15.95	15.95	15.95	15.95	15.95	15.95

Notes: This table presents OLS and IV regressions where the dependent variable is the number of hate crimes against the group in the top row on any given day based on FBI data. We use a dummy for days on which Trump golfs used as an instrument for his tweets about Muslims. The sample year is 2017, for which we have information on Trump's golfing. All regressions include day-of-week and year-month dummies, linear and quadratic time trends as well as dummies for whether Trump's golfing is the first of a series of golf days or part of a "golf holiday" (longer than three days). See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. Weak IV 95% Anderson-Rubin (AR) confidence sets are calculated using the two-step approach of Andrews (2018) with the Stata package from Sun (2018). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.32: Time Series Regression Full Post-Campaign Period: Split by Motivating Bias

	All (1)	Muslim (2)	Ethnicity (3)	Race (4)	Sexual Orientation (5)	Religion (excl. Muslims) (6)
Panel A: Before campaign announcement						
Log(Muslim Trump tweets)	0.013	0.017	-0.001	0.005	-0.012	0.015
	(0.020)	(0.018)	(0.018)	(0.022)	(0.021)	(0.022)
Observations	2,234	2,234	2,234	2,234	2,234	2,234
R^2	0.232	0.026	0.016	0.153	0.107	0.064
Panel B: After campaign announcement						
Log(Muslim Trump tweets)	0.027	0.108**	-0.030	0.027	-0.006	-0.056
	(0.039)	(0.042)	(0.030)	(0.028)	(0.033)	(0.039)
Observations	930	930	930	930	930	930
R^2	0.196	0.079	0.034	0.155	0.077	0.119
Fixed effects (month, day of week)	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS regressions where the dependent variable is the number of hate crimes against the group in the top row on any given day based on FBI data. The sample is split into the period before and after June 16, 2015 when Trump announced his presidential campaign. All regressions include day-of-week and year-month dummies. See online appendix for more details on data and variable construction. Newey-West standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.5. Appendix 5: Additional Bartik Evidence

Table A.33: Bartik Timing Results

	(1) OLS	(2) OLS	(3) Reduced Form	(4) Reduced Form
F4.Muslim Trump Tweet × Twitter Usage	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.002)	(0.002)
F3.Muslim Trump Tweet × Twitter Usage	-0.001	-0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
F2.Muslim Trump Tweet × Twitter Usage	0.003	0.002	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
F.Muslim Trump Tweet × Twitter Usage	0.003	0.003	0.002	0.002
	(0.004)	(0.004)	(0.005)	(0.005)
Muslim Trump Tweet × Twitter Usage	0.003	0.004	0.007	0.007
	(0.004)	(0.004)	(0.004)	(0.004)
L.Muslim Trump Tweet × Twitter Usage	0.009**	0.010**	0.007*	0.008**
	(0.004)	(0.004)	(0.004)	(0.004)
L2.Muslim Trump Tweet × Twitter Usage	-0.000	0.001	0.002	0.002
	(0.004)	(0.004)	(0.003)	(0.003)
L3.Muslim Trump Tweet × Twitter Usage	0.002	0.003	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.003)
L4.Muslim Trump Tweet × Twitter Usage	-0.001	-0.000	-0.005	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
County FE	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes
County x Month FE	Yes	Yes	Yes	Yes
County X Day of Month FE	Yes	Yes	Yes	Yes
Pop. deciles x Day FE	Yes	Yes	Yes	Yes
7 lags dep. variable		Yes		Yes
Observations	2865576	2856252	2865576	2856252

Notes: This table presents OLS and reduced form regressions where the dependent variable is the log number of anti-Muslims hate crime in county c on day d. The independent variable is either the interaction Trump's anti-Muslim tweet with county-level Twitter usage or a reduced form/IV specification with our SXSW variables. The variables are standardized to have a mean of zero and standard deviation of one. All regressions include 4 leads and lags of Trump's anti-Muslim tweets. All regressions include population controls, county, day, county time month and county times day of month fixed effects. Later regression control also control for 7 lags of the dependent variable. Robust standard errors in parentheses are clustered by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
This table presents OLS and IV regressions where the dependent variable is We standardized the variables to have a mean of zero and standard deviation of one
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.