Trump Effect: Hate Speech on Twitter and Hate Crime

Anina Sophie Jauris, Vanisa Vongphanakhone, Dung Duc Huynh

6/13/2020

Introduction

Over the last two decades the use of social media platforms like Facebook and Twitter became an important part of people's lives. Those platforms are not only used for social interaction, they also turned into a major source of information. Compared to other social media platforms Twitter stands out in particular, as 71% of all U.S. users get news on the site (2019a (n.d.)). A large share of news content is directly provided by political figures, with Donald Trump probably being the most active political Twitter user. This direct way of spreading information and opinions combines several advantages. For one thing, the reach and speed of news on Twitter is high, making it possible to instantly react to recent events. Secondly, social platforms are allowing biased statements, enabling politicians to promote their own agenda.

With the beginning of his presidential campaign in 2015, Trump started to regularly denounce several minority groups in order to support his arguments. Some of his tweets were highly prejudiced and often controversially discussed in the context of acceptable speech. Twitter's chief executive, Jack Dorsey, stated in an interview in 2019 that even though some statements of political figures like Trump are in violation of the terms and conditions of Twitter, they are of public interest and therefore remain unaffected by regulations even if they are promoting hate speech (Rogan (2019)). Hate speech broadly refers to every expression of hatred towards a stigmatized group, directed on specific characteristics like race, religion, ethnicity, etc (Álvarez-Benjumea and Winter (2018); Titley, Keen, and Földi (2014)). This immunity remained almost unchanged over the course of Trump's presidency, with a few recent exceptions in the end of May[1]. This shows that Trump had a privileged position in online environments without having to face any consequences for his prejudiced tweets.

This lack of regulation and Trump's elite status raised the question of a potential link between Trumps Twitter use and hate speech and hate crime, respectively. Yearly hate crime statistics provided by the FBI suggest that overall hate crimes in the U.S. have increased since the year of 2014 (Figure 1). The successive increase in hate crime is also observed when looking at anti-racial hate crimes seperately (Figure 2). This developments coincide with the start of Trump's political carrier as he anounced his run for president on June 16, 2015.

[1] On May 26th Twitter appended fact-checking labels to misleading tweets about mail-in voting. On May 29th Twitter again took action and labelled one of Trump's tweets as "glorifying violence" in which he threatened to shoot people if looting happens. The most recent intervention was the removal of a campaign video because of copyright complaints on June 3rd. In response, Trump issued an executive order to strip Twitter's legal protection to make them liable for the content on their platform (Conger and 2020 (n.d.))

Figure 1: Hate Crimes per Year

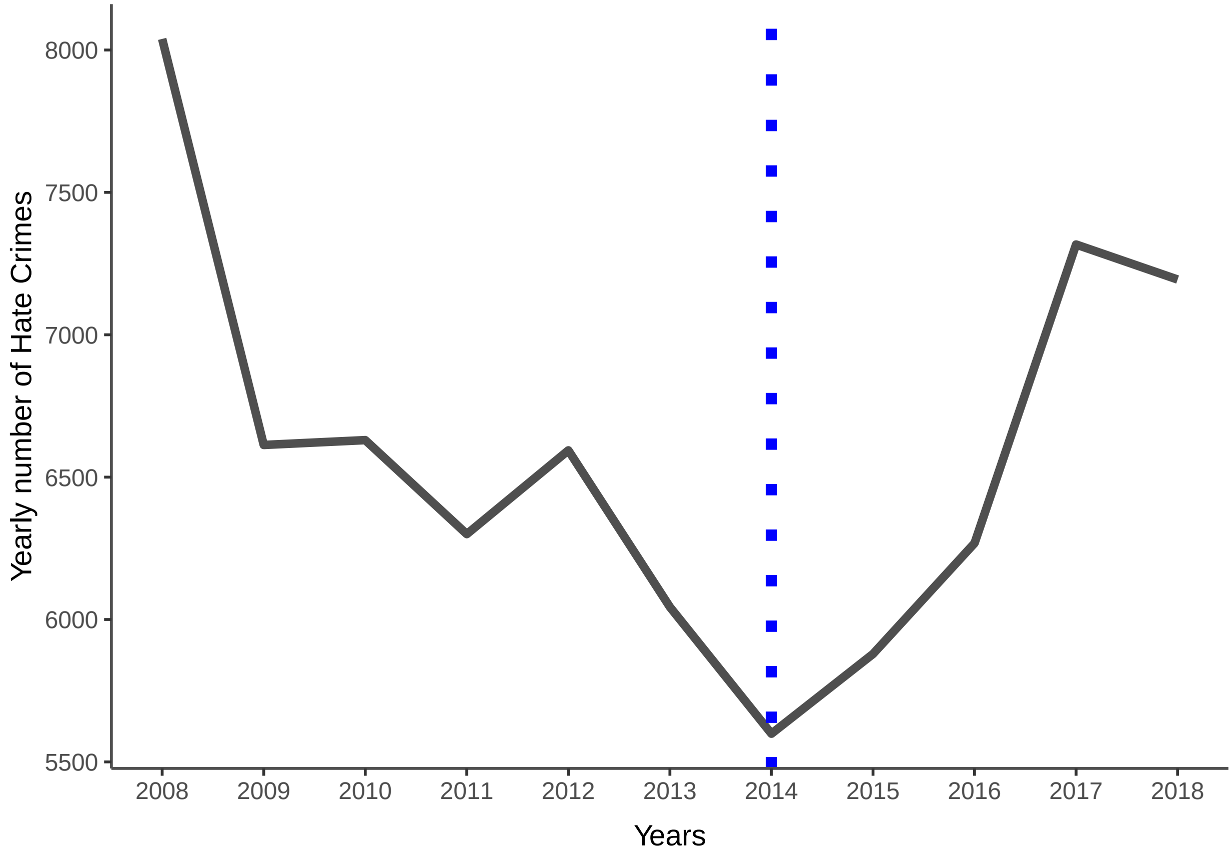
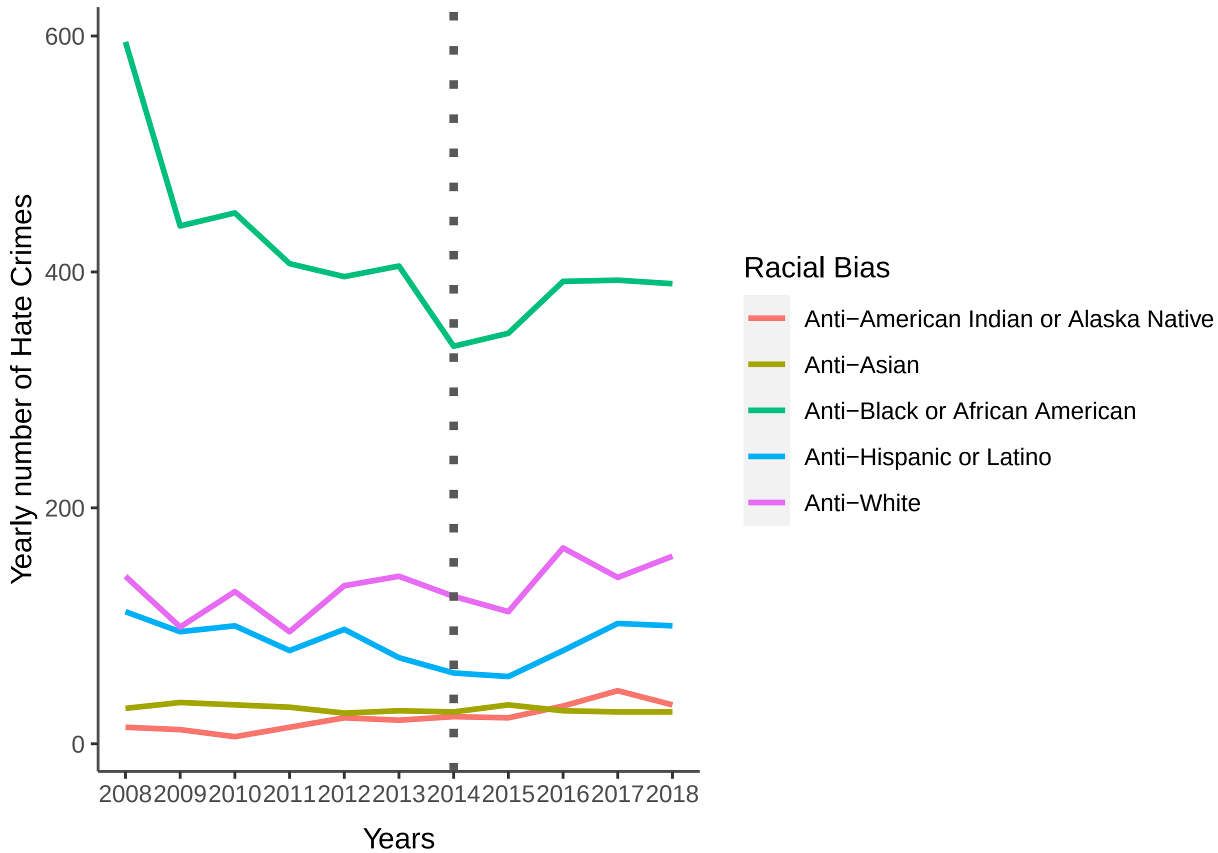


Figure 2: Yearly Hate Crimes per Ethnic Group



In this paper we are interested in uncovering a causal link between Trump's Tweets and hate speech on Twitter on the one hand and Trump's Tweets and hate crime on the other. We will focus on two different minority groups that both used to be Trump's main target group at one point in time. The first minority group is the population of Mexicans and Hispanics, who he addressed especially in the beginning of his presidential campaign. He often referred to them as illegal aliens, associating them with crime, violence and drug trafficking. For this group we have data on Trump tweets, tweets using the same keywords as Trump and data from hate crime statistics which enables us to draw a connection between both suggested links respectively.

The second minority group is the group of Chinese people. Starting with the Corona Virus pandemic Trump made Chinese people responsible for the health issues in the U.S.. Since this are rather new developments, we only have Trump's tweets and tweets using the same keywords as Trump. Data on hate crimes towards Chinese is not available yet, therefore we can only look at the first suggested link. To see if the Chinese minority group in the U.S. is about to face the same increase in hate crime incidences as the Hispanic minority group, we will compare the hate speech outcomes of the two groups. By finding similar or dissimilar results we might be able to make a conclusion for future trends of hate crimes towards Chinese.

We therefore try to answer three different questions. If Trump's online rhetoric is influencing other people's online expression, if Trump's tweets have an effect on hate crime and if Trump's tweets about Chinese might result in more future hate crime towards Chinese.

Theory & Hypotheses

People care to a large extent about what other people think of them. To navigate what behavior is socially accepted, people usually adjust their actions to social norms (Asch (1975); Cialdini and Goldstein (2004)). This norm conforming behavior prevents people from social exclusion and from being sanctioned. Over the last decades there has been a clear social norm against public expression of hate (Schaffner (2018)). Social norms therefore were a main motivation to hide one's prejudice and to not engage in hate speech or hate crime (Crandall, Eshleman, and O'brien (2002)). However, recent trends imply that there has been a disruption of these norms, leading to a greater acceptability of explicitly hateful and degrading behavior.

To learn about social norms people rely on cues from others to infer "normal behavior" (Zitek and Hebl (2007); Schaffner (2018)). These cues are often taken from observations of how other people act (Bicchieri and Xiao (2009); Krupka and Weber (2008)). Especially influential in these processes are people of public interest. Due to his social status as president of the United States and his large news coverage in online and offline media, Trump is likely to be one of the key figures shaping social norms. He might be seen as reference point to whom people adjust their behavior. When Trump uses hateful and clearly racist rhetoric the outer bounds of acceptable speech will expand. People with prejudiced thoughts who formerly suppressed hateful actions might now adjust their behavior to the new limits, resulting in more expression of prejudice. This could be either in the form of online commenting or in the form of physically or verbally assaulting people in real life.

This tendency was also found in prior experimental research. Gervais (2014) was able to show that, republicans are especially prone to negative commenting. When being exposed to like-minded uncivil political rhetoric, republicans were more likely to copy uncivil language in their own messages compared to democrats. Another experiment by Schaffner (2016) was focusing directly on the effect of tweets made by Trump. Individuals in the treatment group, who saw his negative tweets about Muslims, were significantly more likely to also engage in hate speech towards Muslims. Trump's remarks therefore served as a cue, constituting the bounds of acceptable speech and legitimizing uncivil talk. We argue that by creating an environment of higher tolerance for publicly displaying hate, this will not only show in online contexts but also transfer to real life behavior. This might result in open discrimination or even hate crime.

One of the reasons for Trump's considerable influence might be the existence of a general norm of authority compliance, even in the absence of strong incentives (Karakostas and Zizzo (2016)). The status as president might also make him thought of as more prestigious, more knowledgeable and more powerful. People might anticipate Trump to have greater access to crucial political information, unavailable to regular civilians, that informs him in making political statements. This expected information asymmetry might lead some people to think of his statements as more legitimate and trustworthy, regardless of whether this is true or not. Trump might also be seen as "voice of society". Since he won the 2016 elections, he resembles an aggregate of individuals opinion (Bursztyn, Egorov, and Fiorin (2017)). His electoral win thereby informed people about the opinions of the people surrounding them, suggesting that a large share of society actually approves of Trump's negative attitudes towards certain minority groups. This makes sanctioning or social exclusion for publicly expressing prejudice less likely and could therefore cause an increase in hate speech and hate crime.

Literature on framing and priming suggests, that politicians often use framing strategies to systematically manipulate public opinion (Schaffner and Sellers (2009); Chong and Druckman (2007)). By emphasizing certain aspects and leaving out others, opinions on controversial topics, can be modified to fall in line with the opinion held by the political actor. Likewise, Trump seems to use Twitter as political instrument, intentionally placing negative statements to strengthen his own agenda.

One of many examples is president Trump directly relating Mexican immigrants to higher crime rates, drugs, human trafficking and other bad behaviors, to strengthen his anti-immigration politics[2]. He often instrumentalized single incidents and portrayed them as being representative for the minority group as a whole. By ascribing bad characteristics and highlighting the distinctiveness between ingroup and outgroup, he

[2] On Oct 19, 2016 08:22:19 pm he posted for example „Druggies, drug dealers, rapists and killers are coming across the southern border. When will the U.S. get smart and stop this travesty?". On Apr 25, 2017 07:36:28 am he tweeted "Don't let the fake media tell you that I have changed my position on the WALL. It will get built and help stop drugs, human trafficking etc.". These are just some examples of his frequently tweeted aversion towards Mexican immigrants throughout his presidential campaign and presidency.

publicly defined group positions claiming that one group is superior to the other. This insinuated superiority and distinctiveness might elicit feelings of aversion or legitimize already existing aversion (Blumer (1958)). It might also unconsciously strengthen peoples' ingroup identification and trigger a feeling of threat and anger towards the outgroup members (Blumer (1958); Gervais (2013)). This threat is likely to be provoked by the supposed "bad" behavior or "bad" consequences that come with the outgroup members. The minority group of Mexicans might e.g. be perceived as an economic threat or a threat to inner security, since Trump often related them to crime and taking over jobs. When Trump used Terms like "invasion of illegal immigrants", these threats might feel even more severe. Given the current circumstances, with the Covid-19-pandemic, Chinese people on the other hand might have be perceived as a health threat. When the first consequences started to show in the U.S., Trump held China responsible, accusing them of purposely withholding information or even purposely spreading the virus. He regularly called the Covid-19-virus "Chinese Virus" or "China Virus". But even if Chinese people were not primarily seen as health threat themselves, people might have adopted Trump's argumentation, making Chinese people responsible for the critical economic and health situation in the U.S.

What might have intensified these mechanisms even further is the "sorting" mechanism in online environments. People tend to self-select into online platforms or groups used by like-minded people which leads to the formation of so-called echo chambers (Wahlström and Törnberg (2019); Gervais (2013)). By hearing or reading like-minded statements, peoples' beliefs are consolidated which might further increase polarization of opinions. As an analysis of the Pew Research Center (2019b (n.d.)) has shown, Trump's Twitter follower are more likely to be republican or republican-leaning. Thus, one can suggest that his followers already hold more conservative feelings and are more likely to be affected by Trumps tweets. However, tweets sent by the president probably reach a much larger share of people through either retweets, quoting tweets or media coverage. Therefore, a lot more Americans are exposed to Trump's tweets even if they don't follow him. That is why we are interested in the effect of Trump on twitter users in general and not on his followers in particular.

The above argumentation leads us to the following hypothesis.

- H1: If the weekly number of negative Trump tweets increases, the overall number of tweets concerning the same topics will increase.
- H2: If the weekly number of negative Trump tweets increases, the attitudes of tweets concerning the same topics will get more negative.
- H3: If the weekly number of negative Trump tweets increases, the weekly number of hate crime incidents will increase as well.

The first hypothesis looks at how prominent Trump tweets are in guiding public discussion and setting topics of interest, whereas the second hypothesis considers the negativity of the responding tweets. These hypotheses will be examined for the group of Mexicans and Chinese separately. The third hypothesis is focusing on hate crime. Since we only have data for hate crime towards Mexicans our analysis will restricted to this group only. A recent paper from Müller and Schwarz (2019) took a similar approach regarding the effect of Trump on hate crime. They were able to identify a causal link between Trump's prejudiced tweets towards Muslims and anti-Islamic crime. We wish to add to this finding by providing more evidence of Trump's influence on Twitter for other minority groups.

Data

We restricted the data collection for the dependent and explanatory variables on two particular timeframes. For Trump tweets, hate speech and hate crime towards Mexicans we set the timeframe to the 20th of January 2017 till the 20th of April 2017. For Chinese, our timeframe was the 16th of March 2020 till the 12th of June 2020. The decision for those timeframes will be discussed below. To achieve a higher comparability between the trends towards Mexicans and Chinese we only looked at data since Trump officially became a president. His social influence before and after his electoral victory might have been different. When being in office Trump was not only a very well-known public figure, he was also holding judicial power which gave him additional legal legitimacy. We therefore set the starting date for our data collection of Mexican tweets and hate crime to the day of his inauguration, the 20th of January 2017, even though he started commenting negatively a couple of months before. Roughly around the start of his presidential campaign. For the tweets against Chinese we set the starting date of data collection to the first day Trump mentioned the term "Chinese Virus" on Twitter. Since the Covid-19 pandemic is a relatively new topic we were only able to collect data until the 12th of June 2020. To have a similar timespan for both minority groups we also restricted the data collection for Mexicans to a time span of three months. Because the timespan is relatively short, we focus on weekly changes in our analysis.

For an overview over all collected variables and sources please have a look at Table A1 in the Appendix.

Explanatory variable – Frequency of negative Trump tweets

We rely on the Trump Twitter Archive which was made available by Brendan Brown to collect Trump's tweets. The Archive covers almost every tweets ever made by Trump and is updated in real time. The website provides an advantage over extracting Trump's tweets directly from Twitter as it archives the tweets that Trump deletes on his Twitter account as well.

For our analysis we are interested in the frequency of Trump's negative tweets. To be able to calculate how frequently Trump was using hate speech towards a specific group we collected the number of negative Trump tweets and the overall number of Trump tweets per week and divided them.

For the count of negative tweets we used a set of keywords which are likely to be related to the respective minority group. For details please see Table A2 in the Appendix. We looked at every tweet containing one of these keywords separately and made sure that they were addressing Mexicans / Chinese.

The categorization of Trump's tweets into positive and negative tweets was done using the 'Dictionary-Based Sentiment Analysis', which is accessible through the package 'SentimentAnalysis' in R. The sentiment of each tweet is calculated by sum of positive and negative words divided by the total number of words in a tweet. The result is a sentiment score which can range between + 1 and -1, where the score larger than 0 signifies positive and the score lower than 0 indicates negative sentiment. The weekly total number of Trump's tweets was calculated by aggregating his daily tweets into weeks.

Dependent variables

Hate speech

For our hate speech variables we collected data from Twitter. To collect tweets from the general users we utilize python codes 'GetOldTweets-python' from Jefferson Henrique which has been created in 2016. The codes allow us to collect tweets that are older than seven days from the day we collect the data, which is a constraint we face if we use Twitter's official API. The python codes can be found below.

To overcome the negative side of the official API, the python codes took advantage of Twitter Search on browsers. Whenever a client or user used the search bar of Twitter, the scroll loader started, and then if he or she scrolled down further, the scroll loader would start to obtain more tweets by sending requests to the server and receiving JSON responses. The problem was that we do not know what the true mechanism of the searching bar was, how it returned results to clients or whether we had differences between different users in different locations when they were making the searches. So we assumed that the result returned by the search bar was the whole population of the query - the search term.

Tweets for each USA state were collected within the area of a circle in which the center was formed by the longitude and latitude of the state and the radius was the square root of the total area of that state. The information about longitude, attitude and the total area was collected via the internet. The upper limit number of tweets for each state was 1000 tweets.

To make sure that the tweets were referring back to Trump statements, concerning the same topic we used the same keywords as we did for collecting Trump tweets.

For Hispanic tweets we collected, in total, 26155 tweets in 49 out of 50 states. More than 16% of tweets were from the State of California and there were no tweets collected in Alaska.

For Chinese related tweets, we were only able to collect 1000 tweets in total, which are only from one State - Wisconsin. That was a huge disadvantage for our further analysis. The reason for this could be that Twitter users were not willing to publish their location anymore.

The first variable we generated was the total number of tweets including at least one of the keywords. The collected tweets could either be positive or negative. This variable was supposed to represent how influential Trump is in defining relevant subjects of discussion and therefore guiding the overall opinion formation. The more people talked about a topic the more controversial it seems to be.

For our second variable we computed the average sentiments over all the collected tweets using the same keywords as Trump. For doing that we used a dictionary-based sentiment analysis. The sentiments were calculated by using the Harvard-IV dictionary as used in the General Inquirer software. Each word in the dictionary contains a value which can be positive or negative depending on its meaning.

Hate Crime

The data on hate crime was provided to us by the FBI. The dataset included all hate crimes that were reported as part of the Uniform Crime Reporting (UCR) program during the period of 1991 to 2018. The data set described the daily hate crime incidents within each state, including the type of the racial bias. Identical to the Trump tweets and hate speech we also focused on the number of weekly incidents within each state.

We would like to add that the number of reported incidents is likely to be lower than the actual number of hate crimes. In order to be included in the hate crime statistic, crimes had to undergo a two-stage decision process and had to meet a list of objective criteria, characterizing them as hate crime. In addition to this detailed process, incidents that would categorize as hate crime, might not have been reported to the police at all.

Another point we would like to mention is that the yearly number of incidents in each state and the yearly number of incidents in the whole U.S. were a little bit higher than the published reports on the official website of the FBI. We do not know how they aggregated the data for the public reports. So there was inconsistency between our data and FBI's public reports.

Control variables

Twitter Use

For measuring the Twitter use within each state we rely on data from GESIS datorium. Upon request we were able to access geotagged Twitter posts from the U.S., collected in a 6 months period, both in 2014 and 2015. The files contained the aggregated number of hashtags used per day and state. One difficulty we were facing was the large volume of the data. The data was structured in zip-files per month containing several zip-files per day (Gesis datorium). To be able to calculate the total number of hashtags per State we had to join the txt-files first. Luckily R can handle big data well. Therefore we only had to load each txt-file independently and process it directly before moving to another txt-files. Doing so prevented problems with an overflowing memory in R.

Election Results 2016

As a first control variable we collected data on the election outcome for the 2016 election within each state. We extracted the data from a Wikipedia table. We used the count of electoral votes to identify if the majority of the population voted for Trump or for Hillary Clinton. We recorded the values to a dummy variable with 1 reporting a higher voting share for the republicans and 0 reporting a majority for the democratic party.

Demographics

For descriptive and controlling purposes we included a list of demographic variables, which were all collected from the US Census Bureau website (see Figure A1). The data was taken from four different online tables and had to be cleaned afterwards. We chose to use 5-year estimates instead of yearly data to have a more stable representation. We assumed that 5-year estimates are less vulnerable to yearly fluctuation and more likely to represent the general demographic structure in each state. Until now the latest year of reporting is 2018, our data therefore covers the years from 2013 until 2018.

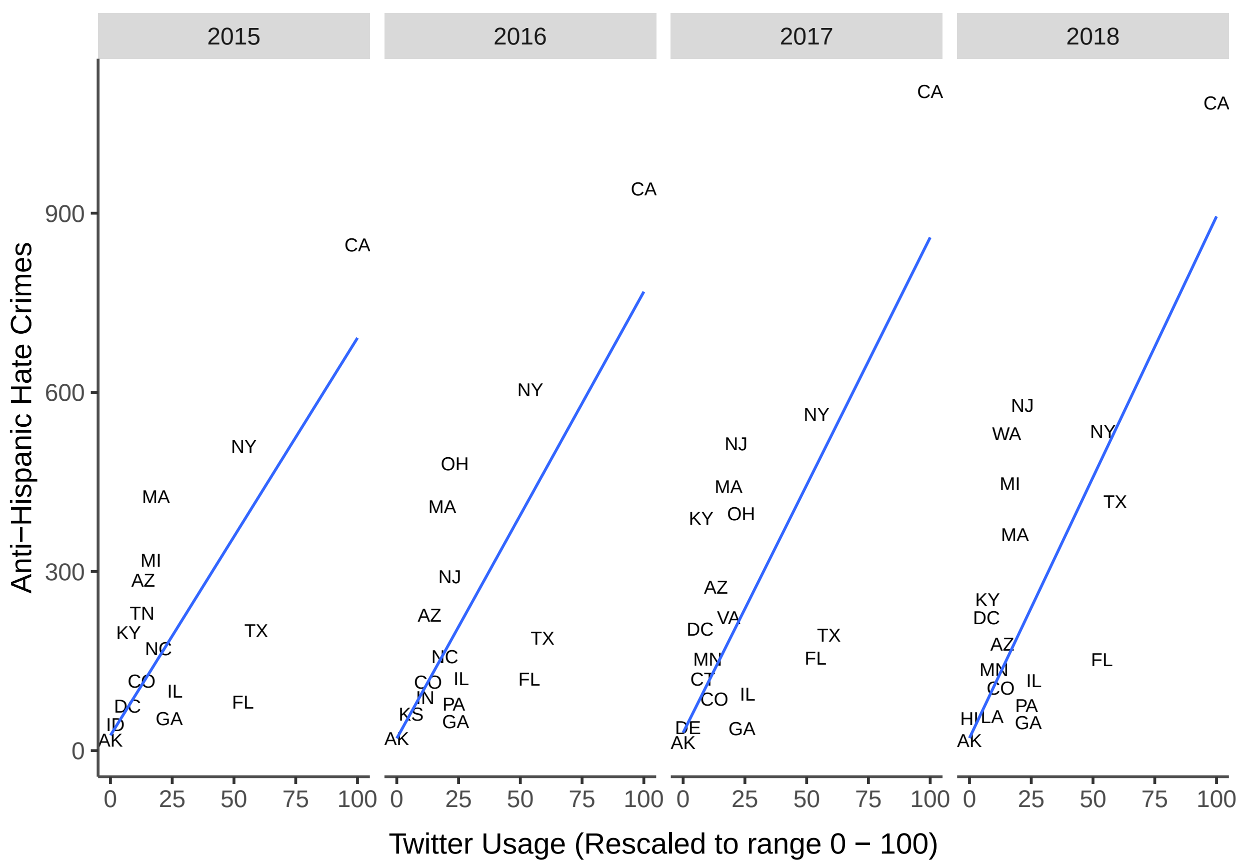
Some of our other variables were taken from the year 2020, which was not represented in the statistics. We argue that by using averaged 5-year data the demographic structure is likely to remain constant and can be seen as a good estimate for 2020 as well.

Graphical data exploration

Hate Crime

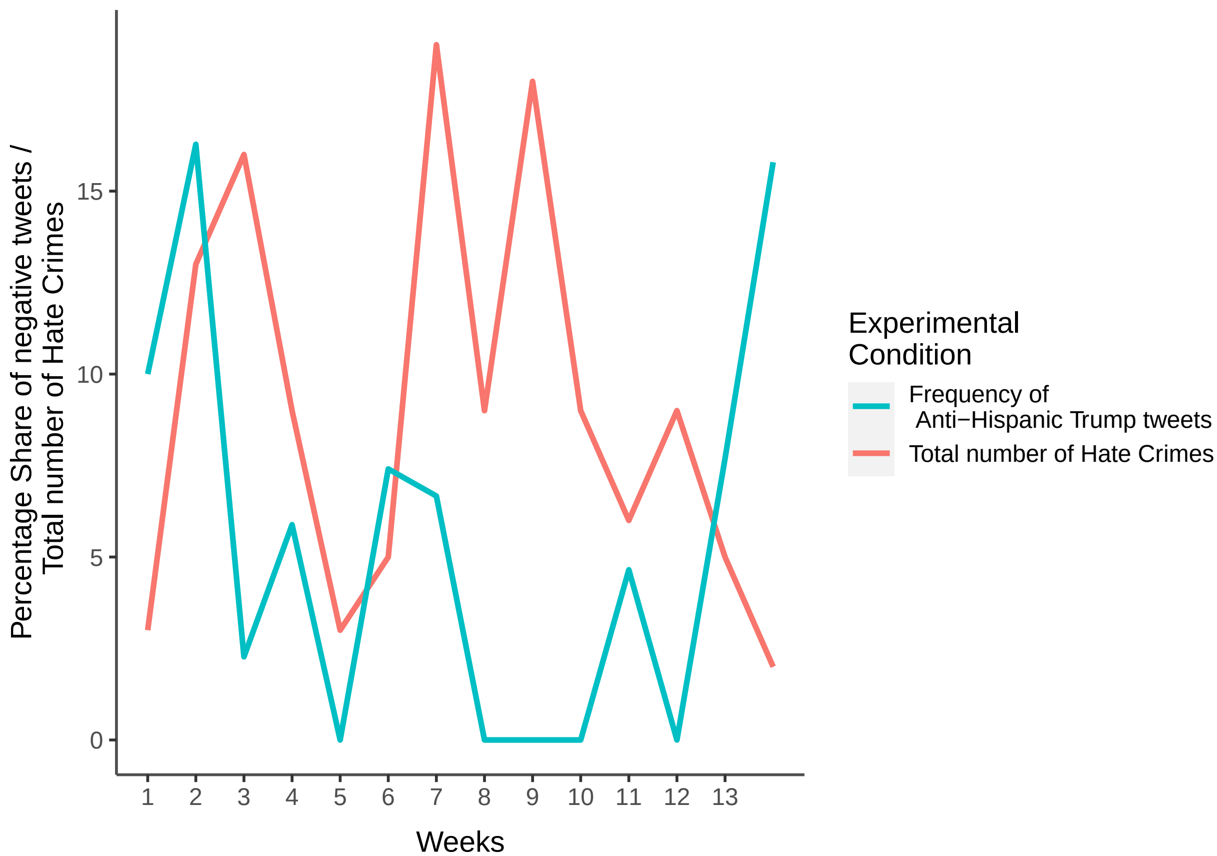
Figure 3 looks at the Twitter Usage and the number of Anti-Hispanic hate crime per State. In this plot we can see, that States with a large Twitter use also seem to have a higher number in hate crimes. This suggests that there is a correlation between Twitter in general and Anti-Hispanic hate crime. To assess if this correlation could be directly induced by Trump's tweets, we will have a closer look on Trumps Twitter activity and Hate Crime in Figure 4. States which show a high Twitter Usage and a high number of crime, tend to be States with a higher overall population and a higher population density, like California or New York. The graphical relation depicted here might therefore also just be a consequence of this larger population size.

Figure 3: Twitter Use and Hate Crime (per State)



With the next graph (Figure 4) we get a glimpse at how Trump's Twitter activity and the number of hate crimes changed over time. The y-axes describes two different variables with different scales. This might be a little problematic, but since it's just an explanatory graph only a minor issue. In the graph we can see that there is a coinciding tendency of the two trends, with hate crime incidents shifted a little to the right. This would suggest that weeks, in which Trump used a lot of negative anti-Hispanic rhetoric in relation to his overall number of tweets, seem to be followed by weeks with more hate crimes. We can see that in the second week after his inauguration aroung 16 % out of all his tweets were directed against Hispanics. This peak in negative tweets was followed by a peak in anti-Hispanic hate crime one week later. The subsequent drop in anti-Hispanic rhetoric in week 3 was again followed by a large decline in hate crimes in week 4. Trump's tweets therefore seem to preceed the hate crime trend by a week. Of course this is just a visual assessment. To get a better idea of the causal relationship we conducted regression analysis in the next chapter.

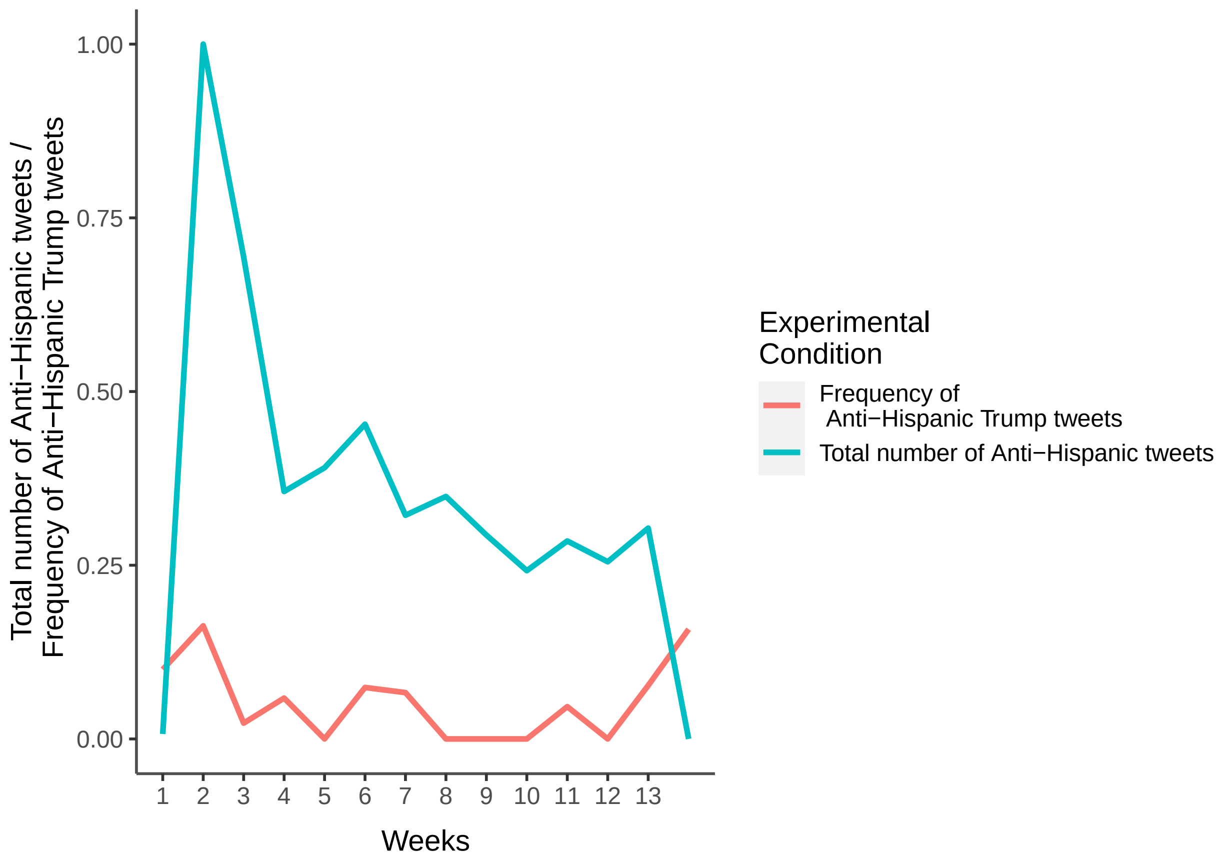
Figure 4: Anti-Hispanic Trump tweets and anti-Hispanic hate crime (U.S.)



Hate Speech - Hispanic

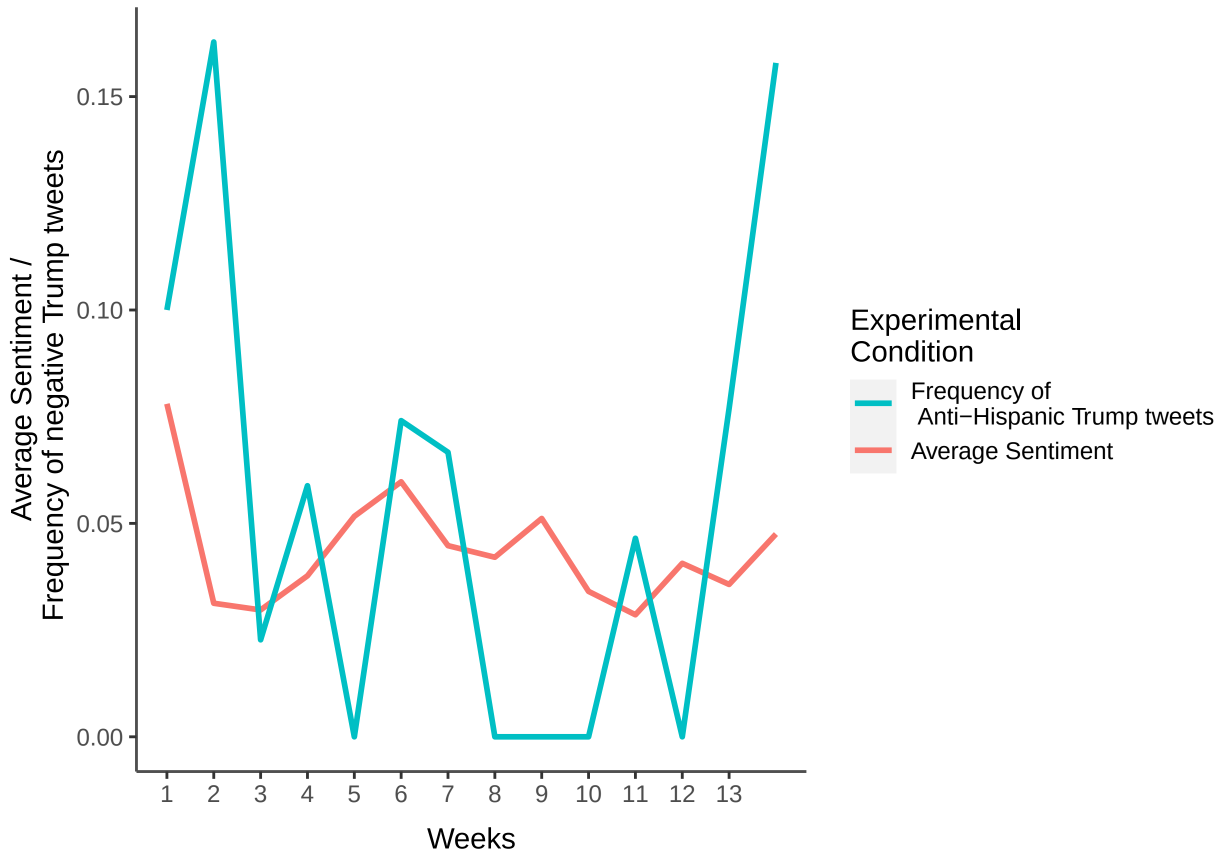
Figure 5 displays the relationship between anti-Hispanic Trump tweets and the total number of tweets from other users using similar keywords. The timeline suggests similar trends in upwards and downwards shifts until week 7. Afterwards the visual relation is a bit more ambiguous. Nevertheless, we still conclude that there is at least a slight correlation. The number of tweets covering the topic of Hispanic immigration is especially high in the beginning of Trump's presidency, as is the frequency of Trump's negative tweets. Both variables are therefore experiencing a peak right after the official inauguration of Trump. The simultaintiy of the trends might be due to preceding negative Trump tweets resulting in more tweets with the similar topic in response to his tweeting activity. Judging from the graph it is possible, that Trump's tweets were determining the relevance of the topic. However, the higher interest in Hispanic migration might have also been induced by his official announcement as president rather than his Twitter activity.

Figure 5: Change in Total number of Anti-Hispanic tweets compared to frequency of Trump tweets



We further looked at the average sentiments of Hispanic related tweets from American Twitter users. We wanted to see if the frequency of negative Trump tweets causes Twitter users to develope stronger negative feelings and to use more offensive language. The average attitudes within the tweets are relatively stable, with only minor variations over time. The small changes appear to be rather random without following Trump's Twitter activity. Since we took the average out of all collected tweets covering Hispanic migration, offensive tweets might have been counterbalanced by positive tweets. With out data we are not able to analyse for heterogeneous treatment effects of Trump's tweets for different groups. But since prior research suggested a stronger effect of negativity for republicans and people with already pre-existing prejudice the results might have been different if we only focused on Trump followers.

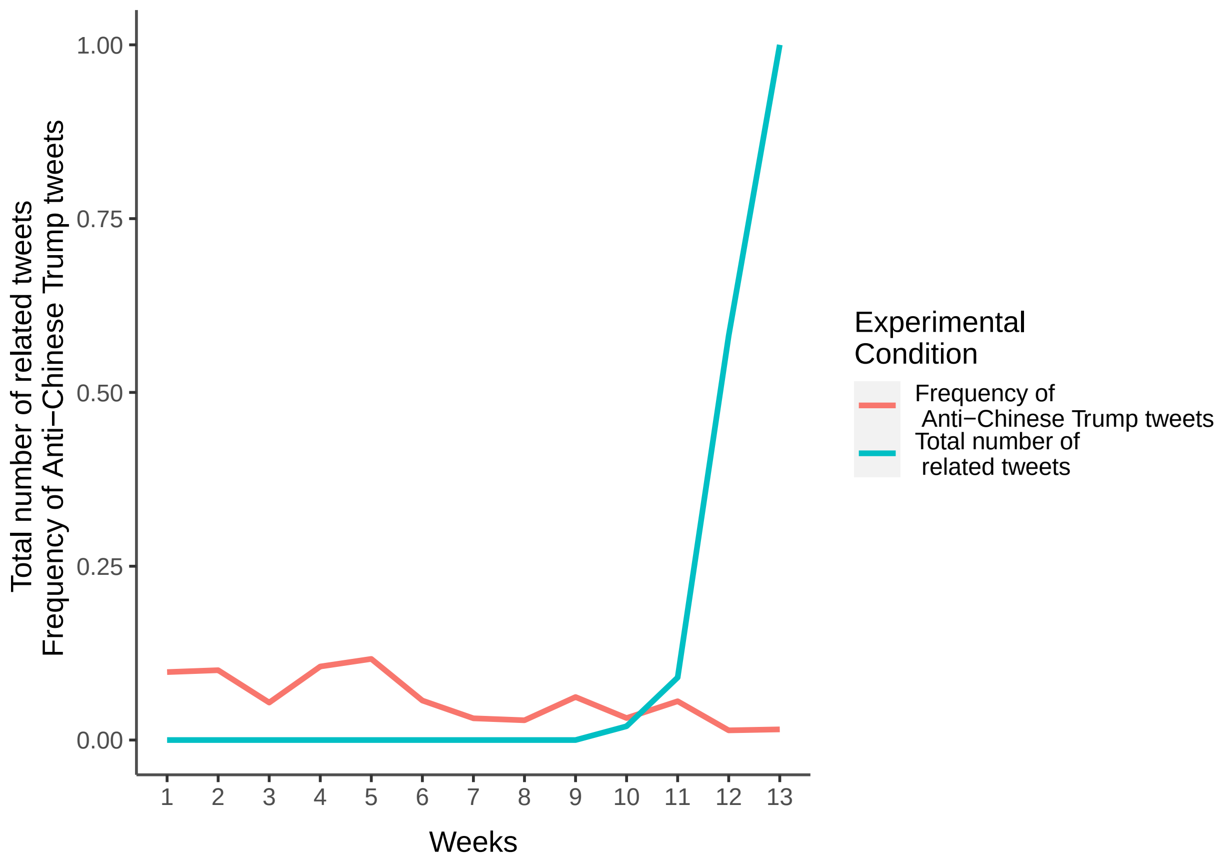
Figure 6: Change in sentiments in relation to negative Trump tweets



Hate Speech - Chinese

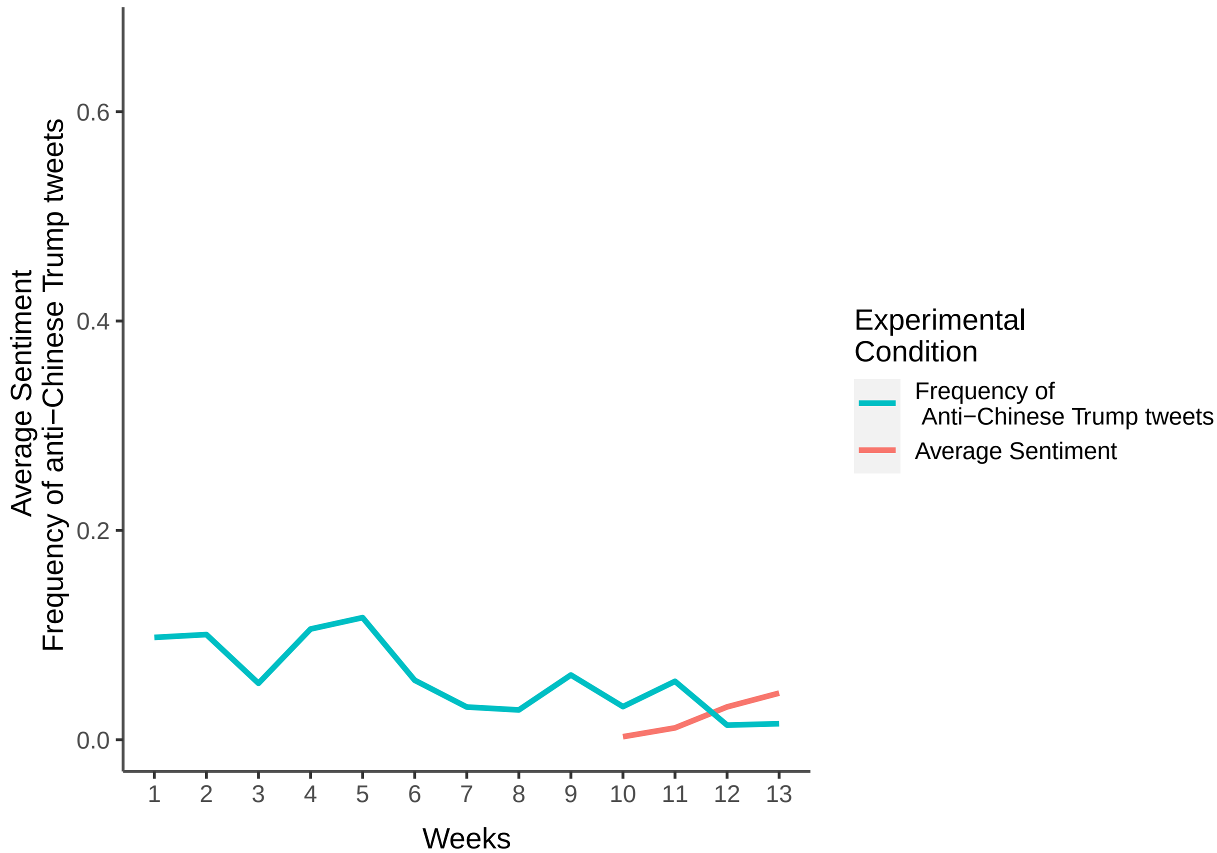
We generated the same graphs that we showed for anti-Hispanic hate speech also for anti-Chinese hate speech. Figure 7 displays the frequency of anti-Chinese Trump tweets in relation to the number of Corona and China related tweets from other Twitter users. We adjusted the scale of the user tweets to fit onto the same scale as Trump tweets from 0 to 1. The graph itself doesn't change by this adjustment. Unfortunately we were not able to collect any data for U.S. tweets prior to week 10. We thus can only look at the trends for user tweets for a short period of time, starting with week 10 up until week 13. During this period Trump's Twitter activity remained relatively low and even decreased. The number in user tweets on the other hand is increasing exponentially. These opposing trends give reason to believe that there is no effect of Trump's Twitter use on the number of overall Corona and China related tweets. Though the result might have been different, if we had more data.

Figure 7: Change in Total number of Corona/ China related tweets compared to frequency of negative Trump tweets



In the next Figure we look at the average sentiment of the user tweets. Again we are only able to look at changes starting from week 10. The change in the overall number of tweets (Figure 7) seems to be accompanied by a general increase in sentiments. Tweets that concern the topic of Corona seem to get more positive along the way. Thus this doesn't support our hypothesis of increasing negative feelings towards Chinese.

Figure 8: Change in sentiments in relation to Anti-Chinese Trump tweets



Originaly we were hoping to be able to compare the trends for anti-Hispanic and anti-Chinese hate speech on Twitter. Given the little amount of available data for Corona and China related user tweets we can not draw a comparison. We can therefore only compare Trump's Twitter behavior in the two timeframes, since we have data for both periods. Trumps's anti-Hispanic tweeting frequency (Figure 5 and 6) is fluctuating from high to low. His interest in the topic therefore doesn't seem to fade and he keeps refering to Hispanics in a negative way. For anti-Chinese rhetoric we see a descending Twitter activity. His interest was especially high in the beginning of the period, when he first started using the term Chinese Virus, but then seems to fade out. We therefore conclude that his anti-Hispanic feelings are more stable and therefore were more influential in influencing people's Twitter use. As Corona is more a temporary issue we expect Trump's Twitter behaviour to further decrease. Eventually this also speaks against a long-term effect on anti-Chinese hate crime.

Regression Analysis

To test our hypotheses in a more causal approach we performed Fixed Effect models. Fixed Effect models address one of the main problems when using longitudinal data. They rule out time-constant unobserved heterogeneity problems and therefore, control for an omitted variable bias. Since our units of interest are states, our models control for all constant state-related characteristics that could have biased the results. Under the assumption that there are no time-varying events influencing the outcome variables, we are able to identify the causal effect of Trump's tweets.

$$Hate\ Crimes = \beta\ Trump\ Tweets \times Twitter\ Usage + \epsilon$$

Since we hypothesized that Trump's rhetoric spreads through Twitter, there should be a higher increase in hate crimes and hate speech in states with larger Twitter usage. We therefore included Twitter usage to create an interaction effect with the frequency of Trump's negative tweets, which allowed us to model the effect heterogeneity and make statements about how Trump's effect varies with Twitter use. This approach was chosen for all outcome variables (hate crime and hate speech). Unfortunately, we were unable to perform any models for anti-Chinese tweets due to the lack of data. In our regression analysis we thus only evaluated the influence of Trump's anti-Hispanic rhetoric.

Hate Crime

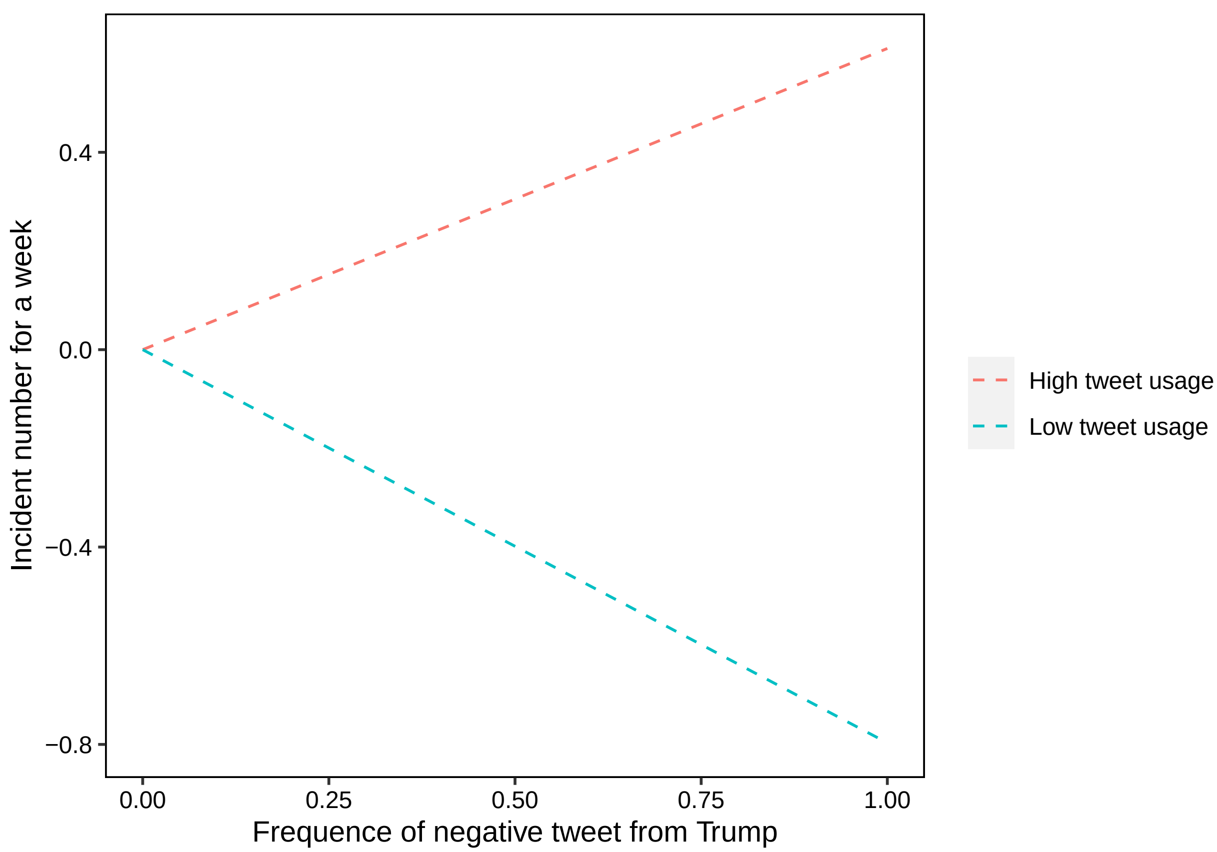
For hate crimes, we assumed Trump's influence to be higher in states with greater Twitter use. Model 1 provided two coefficients. The first coefficient (freq_trump) described a situation where the Twitter usage was zero. Since there were no states in America that did not use Twitter, interpreting this coefficient did not provide any meaningful interpretation.

The interaction effect shown by the second coefficient was positive and proved to be significant on a 1% significance level. From this, we conclude that the influence of Trump's tweets is higher in states with a large Twitter use, causing an increase in the number of hate crimes. This interpretation therefore supports our graphical assessment made earlier and supports hypotheses 3.

Model 1

	Incident number for a week	
	Estimate	P-value
Freq. Trump NegTweet	-1.1487	0.0374 *
Freq. Trump NegTweet x Tweet Usage	0.0035	1e-04 * * *

Interaction Plot



Number of overall tweets

In the second model we were interested in the effect of Trump's tweets on the overall number of tweets related to Hispanics. As can be seen in model 2, there did not seem to be an influence. Both estimates were insignificant. This means that neither Trump's tweets nor Twitter Usage were influencing the volume of Hispanic related tweets. When looking at the direction of the effects we see that it contradicts the direction of our suggested link in hypothesis 1. Our first hypothesis, suggested an increase in the volume of tweets concerning Hispanics, caused by Trump's anti-Hispanic tweets. Due to our results, H1 has to be rejected.

Model 2

	Total users'tweet for a week	
	Estimate	P-value
Freq. Trump NegTweet	-7.2829	0.7448
Freq. Trump NegTweet x Tweet Usage	-0.0449	0.5927

Sentiments of user tweets

Next we will test for the causal effect of the frequency of Trump's negative tweets on the sentiments of Twitter users. The findings revealed that there was no significant effect. As before, we further tested for the interaction effect between the frequency of Trump's negative tweets and the Twitter usage in each state. We expected that states with higher Twitter usage would be affected more strongly by the frequency of Trump's negative tweets. The results suggested that the direct and the interaction effects were both insignificant. In other words, the frequency of Trump's negative tweets towards the Hispanic population did not influence the negative sentiment of the Twitter users towards the Hispanic population; even for the States with high Twitter usage. Hypotheses 2 therefore has to be rejected.

Model 3

	Mean sentiment per week	
	Estimate	P-value
Freq. Trump NegTweet	0.0127	0.8808
Freq. Trump NegTweet x Tweet Usage	0.0000	0.9291

Conclusion

Our study aimed to investigate the influence that Trump has on the public. More specifically, we planned to examine how Trump's Twitter activity had an effect on hate speech on Twitter and the incidents of hate crimes. We chose to look at two time periods: at the start of his presidency where he made degrading tweets about the Hispanic immigrants and population, and the more recent period involving the discussion of Coronavirus where he displayed negative viewpoints on Chinese people. We hypothesized that an increasing frequency of Trump's negative tweets towards the Hispanic and Chinese people would lead to: (1) Increased numbers of tweets about Hispanic and Chinese people; (2) increasingly negative attitudes towards Hispanic and Chinese people; and (3) higher incidents of hate crimes against the Hispanic population. The findings in our analyses suggested that the frequency of Trump's negative tweets led to increases in numbers of hate crime incidents against the Hispanic population in the states with higher Twitter usage. Nevertheless, the frequency of Trump's negative tweets did not affect the number of tweets by Twitter users that mentioned the keywords in Trump's tweets. Moreover, how often Trump displayed his negative sentiment towards the Hispanic and Chinese people did not influence the negative sentiment of the Twitter users towards the two groups as well. In conclusion, while Trump's Twitter usage had led to more hate crimes against the targeted minorities, the president did not really influence the sentiment of Twitter users on the Hispanic and Chinese people.

Limitation

1. For our data analysis we only used data since Trump officially was in office on 20th of January 2017. We decided on restricting the time frame because officially holding the position as president might have changed people's perception of Trump and thereby his influence. His status might have changed from being a well-known public figure to being an important politician who holds judicial power. However, people were already exposed to anti-Hispanic rhetoric during his presidential campaign. Our starting point of data collection for tweets about Hispanics therefore isn't in line with the actual starting point of him using negative Hispanic rhetoric. For the negative rhetoric addressed towards Chinese and Corona on the other hand, we observe the initial reactions when phrases like "China Virus" and "Chinese Virus" have first been used. Therefore, the underlying processes might have been different at the different time points, since anti-Hispanic attitudes already had time to develop.

2. Another thing that might influence the comparability of the two minority groups is, that the population of Chinese were only in the center of attention for a couple of weeks due to other incidences like the murder of Georg Floyd and Trump's conflict between Twitter and Trump. Long term permanent change towards more negative attitude which would then transform into hate crime towards Chinese is therefore unlikely.

3. One limitation of our study is the representativeness of the results. In total relatively few adult Americans are using Twitter, namely only 22 %. Out of them only one-in-five is following Trump's twitter account (Pew Research Center 2019b). We argued that tweets sent by Trump have a larger reach and also influence non-followers. Due to homophily preferences in network formation there might still be a bias towards more republicans seeing his posts. Our results concerning H1 and H2 are therefore likely to be overestimated for American twitter users overall.

4. Another point to consider is the fact that official statistics cannot account for every single crime incident. This disadvantage is even more pronounced for hate crimes where official reports hardly reflect their real prevalence. The number of reported hate crimes can be highly inaccurate due the challenges in identifying criminal incidents that are premeditated by the intention to inflict hate on the groups in which the victims belong to.

Appendix

Table A1: Variables

Variables	Description	Source
Trump Tweets	Frequency of negative tweets per week · ·	Trump Twitter Archive Brown, Brendan. The Trump Twitter Archive. http://www.trumptwitterarchive.com/
Hate Speech	Number of tweets using the same keywords as Trump (can either be positive or negative) Average sentiment of these tweets	Twitter
Hate Crime – Hispanics	Number of weekly incidents (by State)	FBI

Control Variables	Description	Source
Election Results 2016	Dummy0 = Majority of votes for Democrats1 = Majority of votes for Republicans	Wikipedia 2016 United States presidential election, https://en.wikipedia.org/wiki/2016_United_States_presidential_election#Statistical_analysis
Total Population	Total number of citizens	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP05
Hispanic Population	Number of Hispanics Population (by State)	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP05
Chinese Population	Number of Chinese Population (by State)	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP05
Age	Under 2525 to 4445 to 6465 and over	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP05
Education	Less than High School degreeHigh School DegreeSome college or Associate's degree Bachelor's degreeGraduate or professional degree	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP02
Household Income	Mean Household Income	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP03
Unemployment Rate	Percentage of unemployed	US Census Bureau 2018: ACS 5-Year Estimates Data ProfilesTable: DP03
State Poverty Rate	Percentage of population below poverty rate	US Census Bureau 2018: ACS 5-Year Estimates Subject TablesTable: S1701
Twitter Use	Total number of hashtags (by State)	Gesis Datorium https://data.gesis.org/sharing/#!Detail/10.7802/1166

Table A2: Keyword table

Keywords for tweets about Mexicans	Keywords for tweets about Chinese
Border	China virus
criminal	China
immigration	Chinese virus
drug	Coronavirus
gang	Covid
Make America	Pandemic
Mexico	W.H.O
Wall	World Health Organization
Rapist	
Human Trafficking	

References

2019a, Pew Research Center. n.d. "10 Facts About Americans and Twitter." https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/.

2019b, Pew Research Center. n.d. "About One-in-Five Adult Twitter Users in the U.s. Follow Trump." https://www.pewresearch.org/fact-tank/2019/07/15/about-one-in-five-adult-twitter-users-in-the-u-s-follow-trump/.

Asch, Solomon E. 1975. "Effects of Group Pressure Upon the Modification and Distortion of Judgments." Edited by H Guetzknow. *Groups, Leadership, and Men*, 177–90.

Álvarez-Benjumea, Amalia, and Fabian Winter. 2018. "Normative Change and Culture of Hate: An Experiment in Online Environments." *European Sociological Review* 34 (3): 223–37.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only If Others Do so." *Journal of Behavioral Decision Making* 22 (2): 191–208.

Blumer, Herbert. 1958. "Race Prejudice as a Sense of Group Position." *Pacific Sociological Review* 1 (1): 3–7.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2017. "From Extreme to Mainstream: How Social Norms Unravel." National Bureau of Economic Research.

Chong, Dennis, and James N Druckman. 2007. "Framing Theory." *Annu. Rev. Polit. Sci.* 10: 103–26.

Cialdini, Robert B, and Noah J Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annu. Rev. Psychol.* 55: 591–621.

Conger, Kate, and Mike Isaac. 2020. n.d. "Defying Trump, Twitter Doubles down on Labeling Tweets." https://www.nytimes.com/2020/05/28/technology/trump-twitter-fact-check.html.

Crandall, Christian S, Amy Eshleman, and Laurie O'brien. 2002. "Social Norms and the Expression and Suppression of Prejudice: The Struggle for Internalization." *Journal of Personality and Social Psychology* 82 (3): 359.

Gervais, Bryan T. 2013. "Incivility in Online Political Discourse and Anti-Deliberative Attitudes: An Experimental Analysis."

Karakostas, Alexandros, and Daniel John Zizzo. 2016. "Compliance and the Power of Authority." *Journal of Economic Behavior & Organization* 124: 67–80.

Krupka, Erin, and Roberto Weber. 2008. "Why Does Dictator Game Sharing Vary? Identifying Social Norms in the Laboratory." Working paper.

Müller, Karsten, and Carlo Schwarz. 2019. "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment." *Available Here: Https://Ssrn. Com/Abstract* 3149103.

Rogan, Joe. 2019. "Joe Rogan Experience #1236 – Jack Dorsey" 1236.

Schaffner, BF. 2018. "Follow the Racist: The Consequences of Trump's Expressions of Prejudice for Mass Rhetoric. Semantic Scholar."

Schaffner, Brian F, and Patrick J Sellers. 2009. *Winning with Words: The Origins and Impact of Political Framing.* Routledge.

Titley, Gavan, Ellie Keen, and László Földi. 2014. "Starting Points for Combating Hate Speech Online." *Council of Europe, October 2014.*

Wahlström, Mattias, and Anton Törnberg. 2019. "Social Media Mechanisms for Right-Wing Political Violence in the 21st Century: Discursive Opportunities, Group Dynamics, and Co-Ordination." *Terrorism and Political Violence*, 1–22.

Zitek, Emily M, and Michelle R Hebl. 2007. “The Role of Social Norm Clarity in the Influenced Expression of Prejudice over Time.” *Journal of Experimental Social Psychology* 43 (6): 867–76.