

The Tide of Anti-China Attitude

Term Paper for SURVMETH 727

Yuyan Han

2020-12-16

Contents

Introduction	1
Data	2
Results	2
Discussion	12
References	13

Introduction

The US-China relation has slipped into a downward spiral since the trade war in 2018, and the strike of COVID-19 pandemic pushed the already-intense US-China relation to a boiling point in 2020. President Trump publicly condemned China for the spread of the COVID-19 in United States through multiple platforms on countless occasions and insisted on calling COVID-19 “China Virus”. Furthermore, the Trump Administration took multiple aggressive anti-China actions in 2020, including but not limited to ordering China to close Huston Consulate (BBC 2020), scrutinizing Chinese students and scholars at airports under the premise of them being spies (Feng 2020), issuing bans on two Chinese social media apps TikTok and WeChat (Carvajal and Kelly 2020). As BBC News pointed out, such anti-China policies and campaign-speak were aim to resonate with voters, and meanwhile emphasized that China, rather than Trump, should be responsible for huge mess of COVID-19 situations in the United State (Usher 2020). In the large anti-China environment, the public’s view of China seems to be negatively impacted too. Incidents of anti-Chinese or more generally anti-Asian prejudice were frequently reported during the pandemic (Cheung, Feng, and Deng 2020). A news article coming out on September 16th in the Diplomat magazine frankly pointed out that “this was the year that the tide of public opinion in the United States turned against China” (Conrad 2020). As one may imagine, if such negative attitude towards China is indeed a widespread trend among the population rather than isolated incidents, there will be huge and long-term consequences in both macro and micro dimensions. On the large scale, as the Diplomat article commented “if American public opinion remains at such historically low levels, we are likely to see more American politicians abandoning the pretense of a “mostly cooperative” relationship with China”. Such change undoubtedly will influence the US-China relation, and further the global politics and economy. On the small scale, the public opinion may directly impact the life of Chinese and Asians (since Chinese are almost indistinguishable from other East Asians from appearance) in the United States.

Given the large background, the current research aims to investigate the public attitude towards China in the year of 2020. First, we would like to figure out whether there is actually a negative tide of attitude towards China among Americans in the year of 2020 as the news articles demonstrated? Then, if the year of 2020 is a critical turning point for the public opinion about China, we shall expect a downward trajectory of public attitude over time, either steeply or gradually. Furthermore, considering the aggressive anti-China policies and campaign-speak adopted by the Trump administration, the Trump supporters are more likely to be the audience of such anti-China messages compared to non-supporters. This naturally leads to the question of whether there are any differences between Trump supporters and non-supporters in their attitude towards China? Lastly, is there any concurrence in terms of time series between Trump’s attitude and the his

supporters' attitude towards China? If Trump's attitude change is leading his supporters' attitude change or vice versa?

This research hopes to get insight into the above questions through the lens of social media with the aid of sentiment analysis. In the following section, we will specify the data sources and data gathering process.

Data

Reddit Data

To access the public opinions, we resorted to a subreddit `r/AskTrumpSupporters`. Reddit is one the most popular websites in United States over years. At the time of writing, Reddit is the 7th most popular (Alexa 2020). The popularity implies the abundant data available and the possibility for representative opinions. Second, Reddit provides an anonymous environment such that users may express their genuine opinions without the worry about social desirability. Third, compared to other popular social media, like facebook and twitter, Reddit is more research-friendly, by providing free access to most of its data. Specifically regarding the subreddit `r/AskTrumpSupporters`, it has around 926000 participants and stays active since 2016. One specific advantage of this subreddit is that users are required to explicitly tag themselves as Trump supporters or not before making submissions or comments. This is the exact information that fulfills the purpose of the current research.

We collected relevant data using Pushshift API on Python. Briefly, we searched through all the comments within the subreddit `r/AskTrumpSupporters` during the period of 1/1/2020 to 11/2/2020 (right before the Presidential Election) with four key words "China", "Chinese", "Wuhan", "CCP", and stored them in four separate csv files. Note that we store comments about different key words in separate csv files in case that we may want to analyze them separately in future studies. The csv files as well as the Python codes are provided in a separate txt file. Together, we collected 12010 comments containing the key word "china", 3811 comments with the word "chinese", 997 comments with the word "wuhan", and 596 comments with the word "ccp". The four csv files may have some overlapping comments since some comments contain multiple key words. We dealt with the repetition in the data exploration section to obtain an aggregated dataset.

In addition to the china-relevant comments, we also collected a set of random comments from the subreddit over the same period as the control group. The Pushshift API provides no direct arguments to get random comments. Therefore, to achieve this we first generated 200 random hours between 1/1/2020 to 11/2/2020, and collected all comments created during these random hours, with no key words specified. The raw csv file collected a total of 18369 comments. Again, we dealt with possible repetitions later in the data exploration.

Twitter Data

As it is known, Twitter is the major social media that Trump constantly posts his opinions and communicates with the public over the past four years. Currently he has about 88.6 million followers on Twitter, including a substantial portion of his supporters. His tweets can have up to millions of comments, likes and retweets, which implies a big impact on the public. Therefore, Trump's tweets might be a good candidate to study the concurrence of china-relevant attitudes between Trump and his supporters.

Consistent with what we've collected on Reddit, we searched through Trump's tweets (including original replies but excluding retweets) over the same period with the same set of key words using the advanced search function on the Twitter web page. Here is the link to the search results: [https://twitter.com/search?pf=on&q=\(china%20OR%20chinese%20OR%20wuhan%20OR%20ccp\)%20\(from%3ArealDonaldTrump\)%20until%3A2020-11-02%20since%3A2020-01-01%20-filter%3Alinks&src=typed_query](https://twitter.com/search?pf=on&q=(china%20OR%20chinese%20OR%20wuhan%20OR%20ccp)%20(from%3ArealDonaldTrump)%20until%3A2020-11-02%20since%3A2020-01-01%20-filter%3Alinks&src=typed_query). In total we collected 121 tweets. Twitter only allows its official API to scrape its data with multiple limitations for a high price. Considering the small number of tweets we need to collect, instead of web scraping we simply manually copied these tweets and stored in a csv file.

All the data, analyses and associated files are available at: <https://github.com/luluknowsnothing/SURVME-TH727-final-project>

Results

Data exploration

Clean up raw data To start, we cleaned up the raw csv files for further sentiment analysis. We only selected useful variables from each of the four original data sets of china-relevant comments, respectively, “author” (user name), “author_flair_text” (trump supporter or not), “body” (the body of comments), “id” (unique id for each comment), and “created_utc”(the exact time when each comment was created in the format of unix time stamp), and then combined four data sets into an aggregated one excluding repetitions with `dplyr::full_join()`.

```
# select useful variables and combine the four data sets by full_join
keyvar <- c("author", "author_flair_text", "body", "id", "created_utc")
comments_all <- comments_china[keyvar] %>%
  full_join(comments_chinese[keyvar]) %>%
  full_join(comments_wuhan[keyvar]) %>%
  full_join(comments_ccp[keyvar])
```

```
## Joining, by = c("author", "author_flair_text", "body", "id", "created_utc")
## Joining, by = c("author", "author_flair_text", "body", "id", "created_utc")
## Joining, by = c("author", "author_flair_text", "body", "id", "created_utc")
```

The aggregated dataset had a total of 14843 unique comments. Next, we converted the unix time stamps into readable dates and times of `Date` objects to prepare for time series analysis.

```
# convert unix epoch to human readable time and date by Greenwich Mean Time
# create_time: counts to seconds
# create_date: counts to dates
comments_all2 <- comments_all %>%
  mutate(create_time = as.POSIXct(created_utc, origin="1970-01-01", tz = "GMT")) %>%
  mutate(create_date = as.Date(create_time)) %>% # store only the date
  arrange(created_utc) # sort by created_utc from 2020/1/1
```

We also unified the tags of `author_flair_text` and stored the new tags in a variable `TrumpSupport`, because users tagged their support status with different formats. Together we have 6350 comments from non-supporters, 7983 comments from supporters, and 474 comments from undecided users, and 36 comments with missing tags. These comments were from 884 supporters, 1838 non-supporters, 173 undecided authors.

We repeated the same process to clean up the control comments. Together we got 18369 unique comments. Among them, 7813 comments were from non-supporters, 7637 comments were from supporters, and 577 comments were from undecided users. The rest comments were either regulating messages from auto moderators or with missing tags. Among authors, 2141 were supporters, 2162 were non-supporters, and 373 were undecided.

For Trumps’ tweets, we just transformed the dates into `Date` objects to prepare for time series analysis.

Analysis

Get Sentiment Polarity There are lots of packages for sentiment analysis using `r` language, but there are no official standards or rankings that tell us which one is the most superior. To avoid artificial effects simply due to the choice of packages or lexicons, we thus employed two packages and three lexicons for sentiment analysis. In this section, we would describe how we obtained sentiment scores for each comment.

The first approach relied on the package “tidytext”, which simply treats a text as a combination of individual words, and the overall sentiment is just the sum of sentiment of each individual words (Silge and Robinson 2017). This approach requires tidy data before sentiment analysis. Therefore, we transformed the original comments into tidy text, in which one row only contained one word, and then obtained the sentiment for each word by `inner_join` with two lexicons, namely “Afinn” and “Bing”, both providing a sentiment polarity score for each word included. Specifically, “Afinn” includes 2477 words, where the most positive word has a

score of +5, and the most negative word has a score of -5; “Bing” includes 6786 words, where positive and negative words are simply coded in a binary method. Additionally, considering the huge variations regarding the length of the comments, I normalized the sentiment scores by dividing the number of sentences of a comment, thus getting a mean sentence-wise sentiment for each comment. In this way, we reduced possible noises elicited by variations in comment length, and also kept consistent with the default normalization by our second approach with “sentimentr” package. The following r chunks exemplify the process of getting sentiment scores with the “tidytext” approach

```
# Example codes: getting afinn sentiment polarity score for each Reddit comment
# separate sentences of comments
sentences <- get_sentences(comments_all2$body)
# get the number of sentences for normalization of bing/afinn sentiment scores.
sentencecount<- sapply(1:N_all, function(x) length(sentences[[x]]))
# attach sentencecount back to the data frame
comments_all2$sentence_count <- sentencecount
```

```
#1. tidytext approach:
# tokenization of comments
# remove stop words:
comments_clean <- comments_all2 %>%
  unnest_tokens("word", body) %>%
  anti_join(get_stopwords())
```

```
## Joining, by = "word"
```

```
#1.1 with dictionary afinn (get net sentiment/or called polarity)
comments_afinn <- comments_clean %>%
  inner_join(get_sentiments("afinn")) %>%
  group_by(id) %>%
  summarise(afinn = sum(value))
```

```
## Joining, by = "word"
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# Note that not all words have an afinn sentiment score,
# and thus, not all comments has a sentiment score.

# join the normalized sentiment back to the data frame comments_all2
# join by the unique comment id.
comments_all2 <- comments_all2 %>%
  left_join(comments_afinn, by = "id") %>%
  mutate(afinn = afinn/sentencecount)
```

Our second approach relies on the “sentimentr” package (Rinker 2019). A recent review on the r sentiment computation packages recommended the “sentimentr” package over other packages (“syuzhet”, “rsentiment”, and “sentimentanalysis”) due to its proper treatment for negators (Naldi 2019). Compared to the unigram “tidytext” approach, the “sentimentr” package incorporates valence shifters, including negators, amplifiers, and de-amplifiers into weighted sentiment scores. We went with the default lexicon choice, `lexicon::hash_sentiment_jockers_rinker`, which contains 11709 words, each may take one of 19 values in the range of [-2,1], where positive polarity scores indicate positive sentiment and negative scores indicate negative sentiment. As mentioned above, the function `sentiment_by()` by default returns a mean sentiment score across sentences of a comment. One additional difference we shall notice between “sentimentr” and “tidytext” approaches is that “sentimentr” processes the whole text together rather without the removal of stop words. The chunk below exemplifies the process of the “sentimentr” approach,

Table 1: Sentiment Polarity Across Three Methods

method	mean	sd	df	t	p
Afinn	-0.216	1.647	13744	-15.372	<0.0001
Bing	-0.071	0.839	14842	-10.062	<0.0001
SentimentR	-0.028	0.163	14842	-21.134	<0.0001

```

# Get sentiments of each Reddit comments
#2 using sentimentr package
Sentiment_r <- data.frame(matrix(data=NA,nrow=N_all,ncol=2))
names(Sentiment_r) <- c("word_count", "sentimentr")
for (i in seq_along(comments_all2$body)){
  # get an aggregated sentiment score for each comment
  output <- sentiment_by(comments_all2$body[i])
  # return values
  Sentiment_r$word_count[i] <- output$word_count
  Sentiment_r$sentimentr[i] <- output$ave_sentiment
}

# attach the values back to the comments_all2
comments_all2 <- cbind(comments_all2,Sentiment_r)

```

We repeated the same process for control comments and Trump’s tweets. For convenience, we denoted these three methods as “afinn”, “bing”, and “sentimentr” in the present article.

Sentiment Across Groups First, we would like to figure out whether there is actually a negative tide of attitude towards China among Americans in the year of 2020 as the news articles demonstrated? We thus examined the overall attitude towards China as reflected in the subreddit comments. As shown in the table below, we found that the mean sentiment scores are significantly negative for all three methods (Table 1).

Second, is there any difference between Trump supporters and non-supporters? As we expected, Trump supporters showed significantly more negative sentiments in their comments about China compared to non-supporters regardless of methods. As shown by one-way ANOVA, there was a significant main effect of **TrumpSupport** on sentiment scores for all three methods (Afinn: $F(2,13708)=4.91$, $p<0.001$; Bing: $F(2,13929)=17.26$, $p<0.001$; SentimentR: $F(2,14804)=9.52$, $p<0.001$). In addition, post hoc tests using Turkey’s HSD (see Table 2) revealed that the mean sentiment were always significantly more negative for the supporters than the non-supporters across three methods. To demonstrate the difference visually, below is a bar plot with standard error bars comparing the sentiment scores of subreddit comments about China. We also plotted the sentiment scores for Trump’s tweets by side as a reference. Due to the large variations of sentiments scores for Trump’s tweets, the issue of different platforms, and the mixed comparative results between Trump’s sentiments and reddit users’ sentiments, here we avoided over-interpreting the relative negativity between Trump’s sentiments and his supporters’ sentiments.

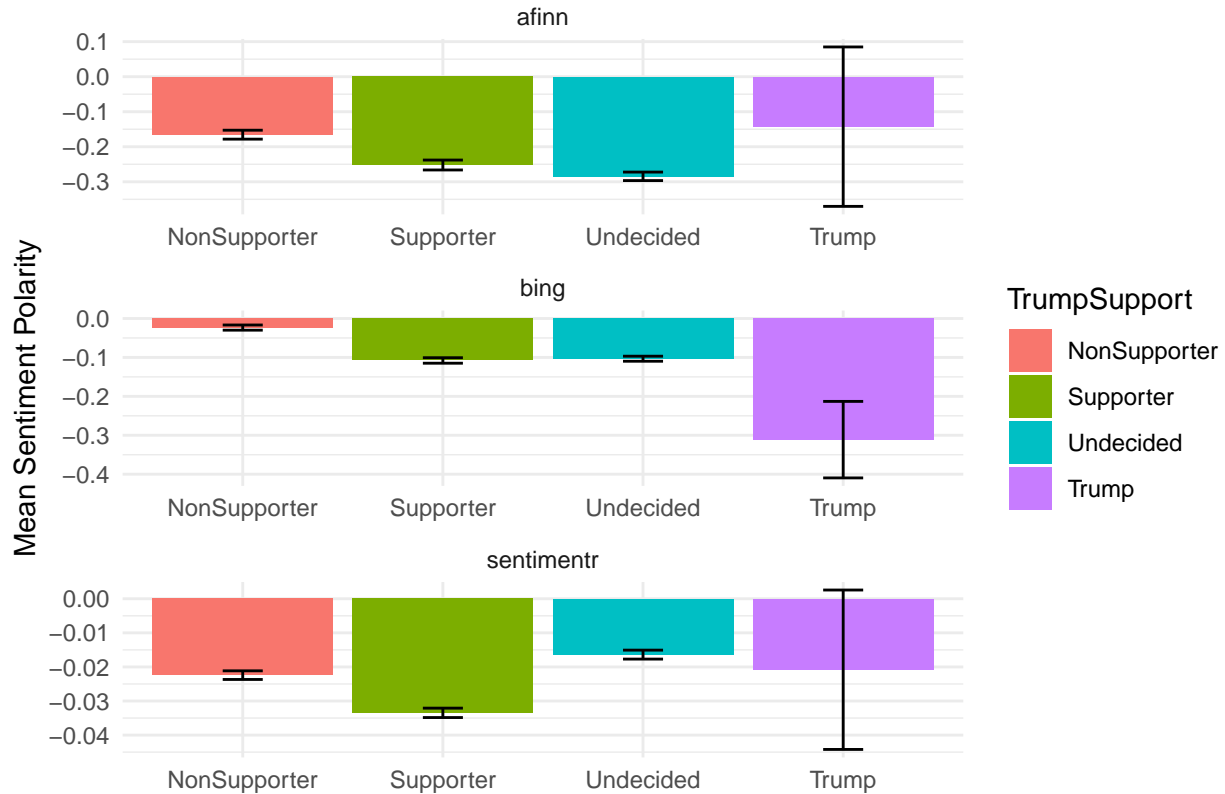
```

## `summarise()` regrouping output by 'TrumpSupport' (override with `.groups` argument)
## `summarise()` ungrouping output (override with `.groups` argument)
## Joining, by = c("method", "grp.mean", "grp.se", "TrumpSupport")

```

Table 2: Summary of Tukey's HSD

compare	diff	lwr	upr	p adj
Afinn				
Supporter-NonSupporter	-0.087	-0.154	-0.019	0.007
Undecided-NonSupporter	-0.119	-0.310	0.072	0.311
Undecided-Supporter	-0.032	-0.222	0.157	0.916
Bing				
Supporter-NonSupporter	-0.084	-0.119	-0.050	0.000
Undecided-NonSupporter	-0.080	-0.177	0.017	0.127
Undecided-Supporter	0.004	-0.091	0.100	0.993
Sentimentr				
Supporter-NonSupporter	-0.011	-0.017	-0.005	0.000
Undecided-NonSupporter	0.006	-0.012	0.024	0.715
Undecided-Supporter	0.017	-0.001	0.035	0.067



To eliminate the possibility that the negative attitude towards China was merely a result of COVID-relevant discussions, we further coded each comment as COVID-relevant or irrelevant. Specifically, if a comment contained any of the following six strings: “corona”, “covid”, “sarscov”, “ncov”, “virus”, “pandemic”, this comment was considered as COVID-relevant comment, and marked as 1 for the variable `covid_relevant`. Otherwise, the comment was considered as irrelevant and marked as 0. We selected the six key strings based on the key words Lamsal (2020) used to collected the COVID-19 Tweets dataset.

```
# encode covid-19 comments with sjmisc package
covid_keywords <- c("corona", "covid", "sarscov", "ncov", "virus", "pandemic")
covid_relevant <- rep(NA, N_all)
for (i in 1: N_all) {
```

```

  if (str_contains(comments_all2$body[i], covid_keywords, logic = "or")){
    covid_relevant[i] <- 1
  }else{
    covid_relevant[i] <- 0
  }
}

table(covid_relevant)

## covid_relevant
##      0      1
## 10192 4651

# attach back
comments_all2$covid_relevant <- factor(covid_relevant)

# setting it as a control in analysis
# Any differences between trump supporters and non-supporters?
grp.sum_covid <- comments_all2 %>%
  select(TrumpSupport, afinn, bing, sentimentr, covid_relevant) %>%
  filter(!is.na(TrumpSupport)) %>%
  pivot_longer(-c(TrumpSupport, covid_relevant), names_to="method", values_to="polarity") %>%
  group_by(TrumpSupport, method, covid_relevant) %>%
  summarise(grp.mean = mean(polarity, na.rm=TRUE),
            grp.se = sd(polarity, na.rm=TRUE)/sqrt(N_all))

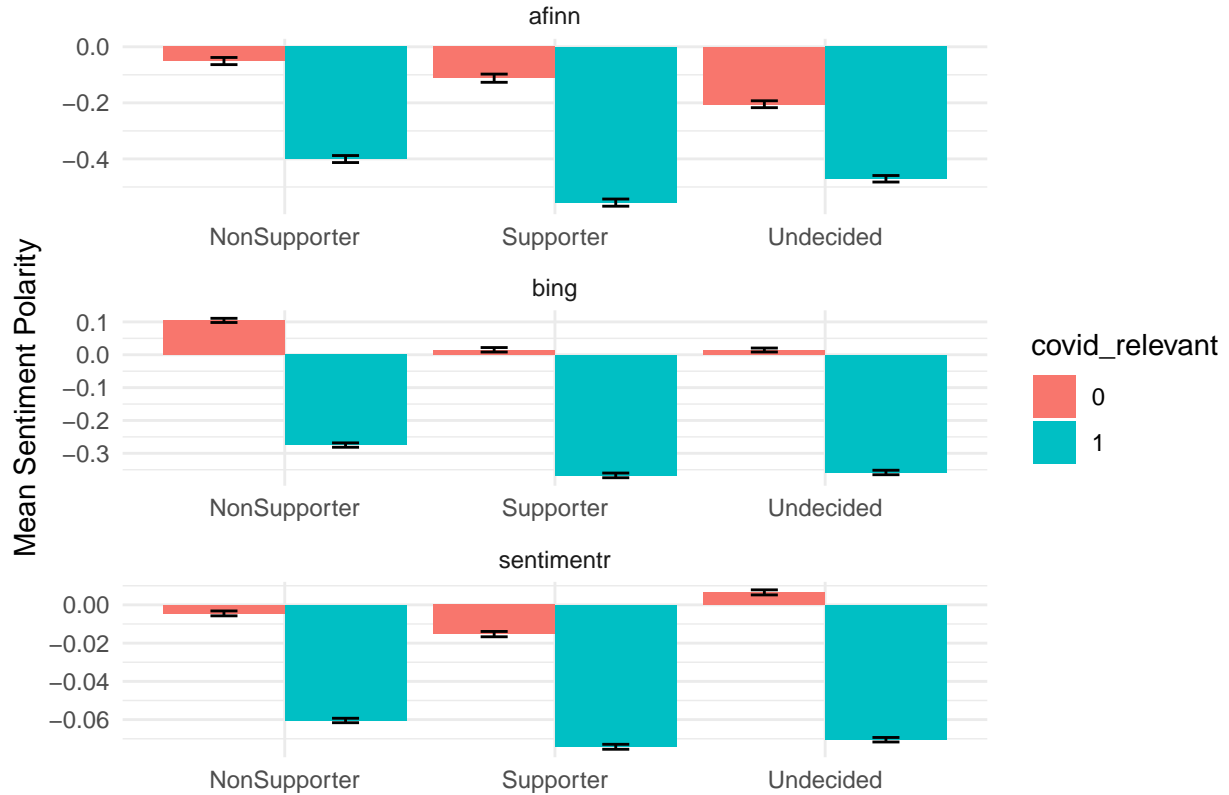
## `summarise()` regrouping output by 'TrumpSupport', 'method' (override with `.groups` argument)

## barplot
grp.sum_covid %>%
  ggplot(aes(x=TrumpSupport, y=grp.mean, fill=covid_relevant)) +
  geom_bar(stat="identity", position=position_dodge()) +
  geom_errorbar(aes(ymin=grp.mean-grp.se, ymax=grp.mean+grp.se), width=.2,
               position=position_dodge(.9)) +
  facet_wrap(~ method, scales = "free", nrow = 3) +
  theme_minimal() +
  xlab("") + ylab("Mean Sentiment Polarity")

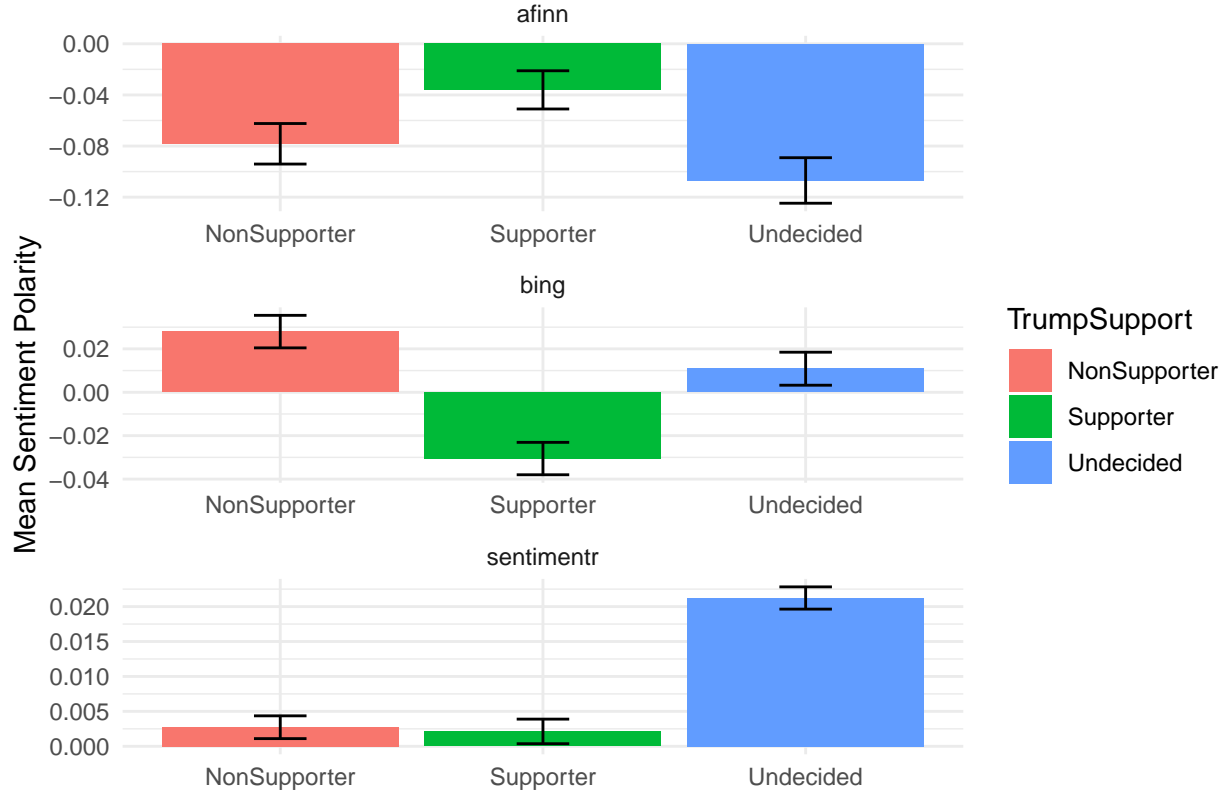
```

Table 3: Summary of ANOVA

Effect	Df	Sum Sq	Mean Sq	F value
Afinn				
TrumpSupport	2	26.606	13.303	4.973
covid_relevant	1	471.941	471.941	176.415
TrumpSupport:covid_relevant	2	7.986	3.993	1.493
Residuals	13705	36663.365	2.675	NA
Bing				
TrumpSupport	2	24.195	12.098	18.082
covid_relevant	1	445.412	445.412	665.745
TrumpSupport:covid_relevant	2	0.013	0.006	0.009
Residuals	13926	9317.107	0.669	NA
Sentimentr				
TrumpSupport	2	0.501	0.251	9.788
covid_relevant	1	10.794	10.794	421.744
TrumpSupport:covid_relevant	2	0.043	0.021	0.836
Residuals	14801	378.815	0.026	NA



Again, we visually presented the results in the format of bar plots with standard error bars. As we can see, COVID-relevant comments were associated with significantly more negative sentiments than irrelevant comments. But Trump supporters' comments about China were significantly more negative than non-supporters for both COVID-relevant and irrelevant contents. It indicates that Trump supporters' extra negative sentiments about China is not only concentrated on the topic of COVID-19, but spills over to topics about China in general. We also confirmed the above observations with two-way ANOVA (Table 3), where we found significant main effects of `covid_relevant` and `TrumpSupport` on sentiments for all three methods.



Next, we compared the sentiments of China-related comments to the control comments, which serves as a baseline. One-sample t-test shows that the overall sentiments of these random comments were significantly positive for all three methods (Afinn: $t(13232)=5.46$, $p<0.001$; Bing: $t(13474)=2.77$, $p<0.01$; Sentimentr: $t(18368)=6.26$, $p<0.001$). Compared to what we have showed previously that the overall sentiments were significantly negative for china-related comments, it implies that in general there seemed to a tide of anti-China attitude. We also performed an one-way ANOVA of sentiments on comments contents (china-related vs. control) to confirmed the above findings. In deed, the mean sentiments for random comments in this subreddit were significantly more positive than those for china-related comments for all three political groups as revealed by the main effect of comments contents (afinn: $F(1,26976)=186$, $p<0.001$; bing: $F(1,27438)=76.9$, $p<0.001$; sentimentr: $F(1,33210)=340.2$, $p<0.001$).

We then further examined that difference between Trump supporters and non-supporters in terms of their sentiments on random comments. As shown in the bar plots below, there is no difference between Trump supporters and non-supporters in terms of baseline sentiments. The results indicated that Trump supporters held certain exclusive negative attitude about China.

Sentiment Trend Over Time In this section, we examined the trend of sentiment change over time. To achieve this, we first transformed the sentiments into time series objects and ordered them by date with the aid of “xts” package. Below is a sample chunk of transforming to Date object with “xts” package.

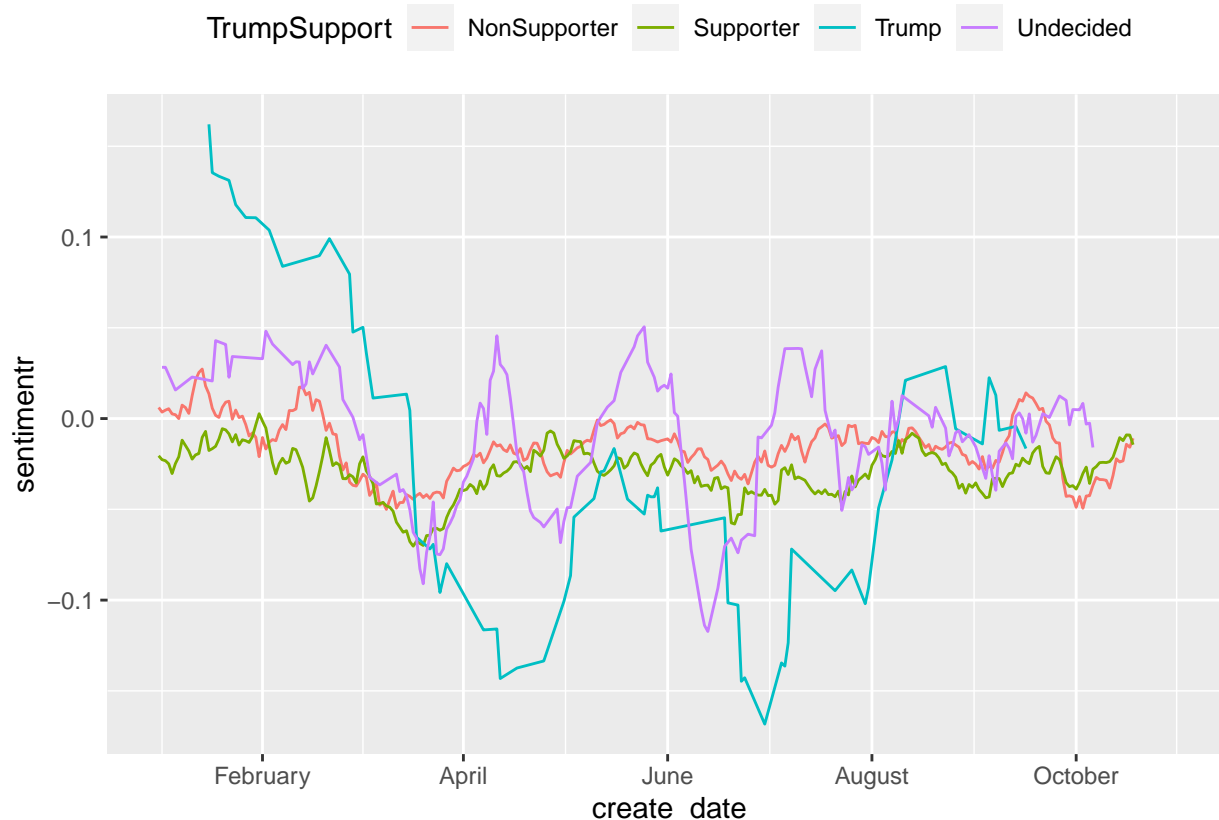
```
# Sample codes for transforming to Date objects
ts_sentimentr_all <- senti_all %>%
  select(create_date, TrumpSupport, sentimentr) %>%
  pivot_wider(names_from = TrumpSupport, values_from = sentimentr)
ts_sentimentr_all <- xts(ts_sentimentr_all[-1], order.by=ts_sentimentr_all$create_date)
```

Next we visualized the time series over the whole period. We did it in two ways. The first way is a static time series plot with the aid of “ggplot2”. In particular, we took the rolling average across a period of 15 days to smooth the time series curve, which shall ease the readers’ eyes to spot the trends. Specifically, each point on the time series plot represents the mean sentiments over 15 days staring at the date of selection. The plot

below is an example of time series plot, with sentiment polarity scores obtained by “sentimentr” package.

```
# sample codes for time series plot with ggplot2
ggplot(senti_all, aes(x = create_date, y = sentimentr, color = TrumpSupport)) +
  scale_x_date(date_breaks = "2 month", date_labels = "%B") +
  geom_line(data = senti_all %>%
    group_by(TrumpSupport) %>%
    mutate(sentimentr = rollmean(sentimentr, 15, align = "left", fill = NA))) +
  theme(legend.position="top")
```

Warning: Removed 56 row(s) containing missing values (geom_path).



In particular, we also plot interactive time series plots with “dygraph” package in case readers would love to investigate into the details on each day for each group. Again, we smoothed the day-by-day jitters by plotting the rolling means across a period of 15 days to make the plot more readable. The interactive plots were save as html files (“dy_afinn.html”, “dy_bing.html”, “dy_sentimentr.html”) within the same file.

As we may notice, visually there seemed to be a sharp decrease of sentiment at the beginning of the year and then stayed relatively stable for reddit comments about China (the trend is more clear with dygraph plots than ggplot2). Trump’s tweets about China, showed greater variations in sentiments—a sharp decrease at the beginning of the year and a big dip around August. We shall note that the large variations might be due to the small sample size (121 tweets in in total).

```
# sample codes for interactive time series plot with dygraphs
dy_sentimentr <- dygraph(ts_sentimentr_all, main = "Sentiment by SentimentR") %>%
  dyRangeSelector() %>%
  dyRoller(rollPeriod = 15) %>%
  dyOptions(colors = RColorBrewer::brewer.pal(4, "Set2")) %>%
  dyLegend(show = "always", hideOnMouseOut = FALSE, width = 150)
```

Table 4: Mann-Kendall Trend Test Over 1/1-11/2

	Afinn_tau	Afinn_p	Bing_tau	Bing_p	SentiR_tau	SentiR_p
Trump	-0.106	0.170	-0.193	0.013	-0.167	0.026
Supporter	0.069	0.074	0.012	0.763	0.013	0.731
Non-Supporter	0.023	0.546	0.059	0.125	-0.027	0.489
Undecided	-0.060	0.226	0.005	0.926	-0.004	0.935

Table 5: Mann-Kendall Trend Test Over 1/1-4/1

	Afinn_tau	Afinn_p	Bing_tau	Bing_p	SentiR_tau	SentiR_p
Trump	-0.197	0.214	-0.296	0.056	-0.261	0.078
Supporter	-0.238	0.001	-0.316	0.000	-0.218	0.002
Non-Supporter	-0.165	0.020	-0.239	0.001	-0.201	0.005
Undecided	-0.236	0.020	-0.209	0.040	-0.239	0.017

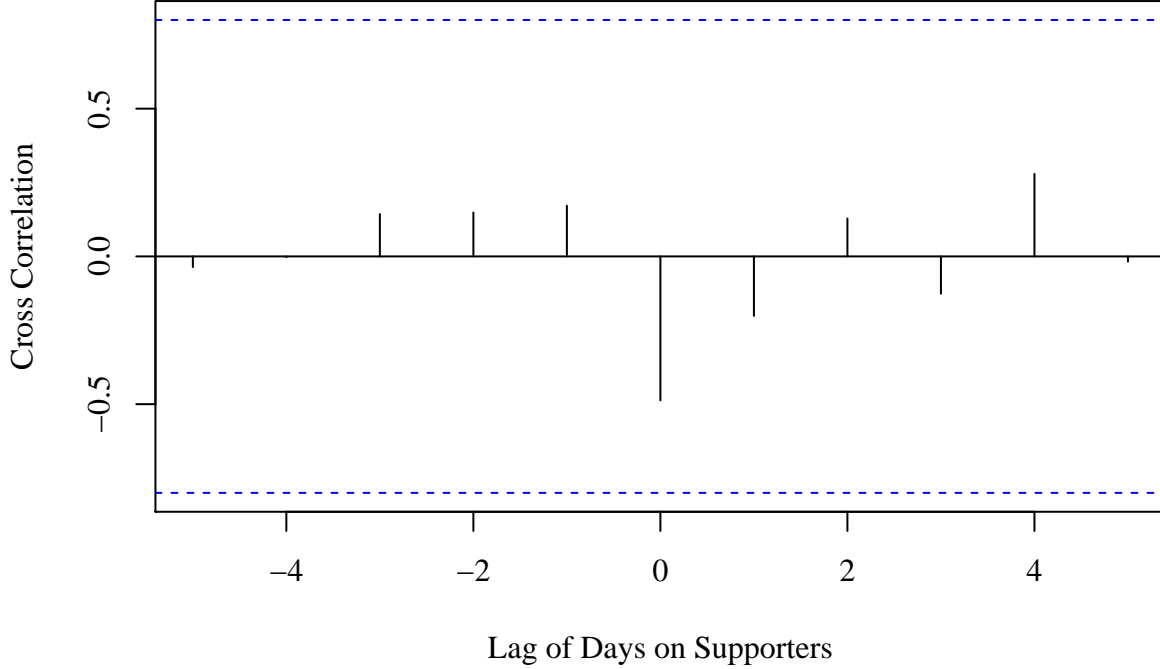
```
# save as html file
# htmlwidgets::saveWidget(widget = dy_sentimentr, file = "dy_sentimentr.html")
```

Statistically, we employed the Mann-Kendall trend test to examine whether there is a monotonic trend in the sentiment change over time. As shown in the Table 4, the results seemed to be mixed across methods. It is hard to conclude whether there was a significant monotonic trend over the whole ten months. However, if we focused only the first three months, the trend was quite clear. Across all methods and groups, the Mann-Kendall trend tests revealed a significantly downward trajectory of sentiments (Table 5). It indicates that the possible critical turning point mentioned by the Diplomat article appeared at the beginning of this year. It might be triggered by the outbreak of COVID. But the current research did not examine the causality and thus we had no exact answer.

Concurrent Trends Lastly, we examined whether there is concurrence between Trumps’ sentiment trends and his supporters sentiment trends. To achieve this, we used the “forecast” package. We computed the cross-correlations and lagged cross-correlations between Trump’s sentiment trends and his supporter’s sentiment trends over the whole period with the function `Ccf()`. In both the plot and Table 6, the lag of days means the lags on Trump’s supporters. For example, when lag of days= k , the cross-correlation was computed by correlating Trump supporters’ sentiment at Day $N+k$ and Trump’s sentiment at Day N . As shown in the plot (using `afinn` as an example) and Table 6, no significant cross-correlations or lagged cross-correlations were found up to a lag of 5 days in either direction (i.e. all lower than the blue-dashed line, which indicates significance).

Table 6: (Lagged) Cross-Correlations

	-5	-4	-3	-2	-1	0	1	2	3	4	5
Afinn	-0.04	0.00	0.14	0.15	0.17	-0.49	-0.20	0.13	-0.13	0.28	-0.02
Bing	0.00	0.12	-0.40	0.65	-0.46	-0.28	0.80	-0.50	-0.06	0.21	-0.07
SentimentR	-0.08	-0.16	0.33	0.41	0.19	-0.56	-0.54	0.21	-0.11	0.10	0.21



Discussion

In conclusion, we did observed an anti-China tide in the year of 2020 within the subreddit r/AskTrumpSupporters. The comments about China were significantly more negative than an average comment from the same subreddit, regardless of the authors' political partisanship. In particular, Trump supporters showed significantly greater negativity than non-supporters in their comments about China. Such difference was not found in their sentiments about random comments, which indicates that Trump supporters' negative attitude towards China was specialized, rather than a general venting of negative emotions. Furthermore, we found that the Trump supporters' negativity about China was not limited to topics related to COVID-19. Not only did they show more negative sentiments in their comments about COVID-19, but also more negativity in their other comments about China that were irrelevant to COVID-19. In summary, there seemed to be an anti-China tide of attitude, and particularly Trump supporters were more negative about China than non-supporters.

Regarding the trend of attitude over time, we observed a sharp drop at the beginning of the of year in general regardless of the authors' partisanship. The trend then stayed negative over the rest of the period. It imply that the beginning of year 2020 might be a critical turning point in terms of the public attitudes towards China as the news articles claimed. Intuitively, the hit of COVID-19 might be the trigger. But without comparing with data from previous years, as well as a large and representative sample of U.S. population, the exact trigger of such drop is still open to many other possibilities.

Additionally, we did not find significant cross-correlations between Trumps' attitude change and his followers' attitude change in the current research. But we shall be careful in interpreting the results. It by no means concludes that there is no mutual influence on Trump's and his supporters' attitudes towards China. The current research only glimpsed at Trump's attitudes through a very limited number of tweets, which brought

much noise into the time series analysis. The comparison of sentiment trends from two different social media further noised the effect if there was one. To better study the dynamics between Trump and his supporters, future studies may also look into the interactions between Trump and his supporters, such as likes, comments, and retweets on Twitter.

There are many limitations of the current study. As mentioned above, we do not know how representative the subreddit sample were of the whole U.S. population. Therefore, we could not blindly generalize the findings to the whole United States. Nevertheless, it is good start. In the future studies, we may collect data from other social media, or spread surveys on different survey platforms to compare with the current findings. Additionally, we may expand the time period under examination, including the post-Trump-presidency period. On the one hand, the historical data would help us to reduce some noises and give us a more precise sense of how the public attitude about China changed over the year of 2020. On the other hand, if there was actually a large-scale anti-China tide, how long it would linger after the end of Trump’s presidency is undoubtedly a critical question for individuals as well as the whole society. Finally, there are many other variables we might look at in order to get a nuanced understanding of the anti-China tide. For example, is there any difference in their attitudes about China between Trump supporter and non-supporters within Republicans? How about the public’s attitude about other countries this year—is it just about China, or a general Xenophobia? Is there any regional differences? As we can imagine, the impact of the negative attitude might differ enormously between regions where Chinese clustered and where Chinese distribute sparsely.

In the end, the current study is just a small step in understanding the tide of anti-China attitude. We believe that understanding the phenomenon is the first step to identify the key factors that elicit the tide of anti-China attitudes, which would eventually help us to figure out approaches to reduce misunderstandings, prejudices, discriminations, and conflicts between groups, and bring harmony back to the society.

References

- Alexa. 2020. “The Top 500 Sites on the Web.” *Alexa.com*. <https://www.alexa.com/topsites>.
- BBC. 2020. “Chinese Consulate in Houston Ordered to Close by Us.” *BBC.com*. <https://www.bbc.com/news/world-us-canada-53497193>.
- Carvajal, Nikki and Caroline Kelly. 2020. “Trump Issues Orders Banning Tiktok and Wechat from Operating in 45 Days If They Are Not Sold by Chinese Parent Companies.” *CNN.com*. <https://www.cnn.com/2020/08/06/Politics/Trump-Executive-Order-Tiktok/Index.html>.
- Cheung, Helier, Zhaoyin Feng, and Boer Deng. 2020. “Coronavirus: What Attacks on Asians Reveal About American Identity.” *BBC.com*. <https://www.bbc.com/news/world-us-canada-52714804>.
- Conrad, Justin. 2020. “US Public Opinion Is Changing the Relationship with China.” *Thediplomat.com*. <https://thediplomat.com/2020/09/Us-Public-Opinion-Is-Changing-the-Relationship-with-China/>.
- Feng, Zhaoyin. 2020. “Chinese Students Face Increased Scrutiny at Us Airports.” *BBC.com*. <https://www.bbc.com/news/world-us-canada-54016278>.
- Lamsal, R. 2020. “Corona Virus (Covid-19) Tweets Dataset.” *IEEE Dataport*.
- Naldi, Maurizio. 2019. “A Review of Sentiment Computation Methods with R Packages.” *arXiv Preprint arXiv:1901.08319*.
- Rinker, T. 2019. “Sentimentr: Calculate Text Polarity Sentiment (R Package Version 2.7. 1).”
- Silge, Julia and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. " O’Reilly Media, Inc."
- Usher, Barbara P. 2020. “Why Us-China Relations Are at Their Lowest Point in Decades.” *BBC.com*. <https://www.bbc.com/news/world-us-canada-53517439>.