

Assessing 2019 Novel Coronavirus (COVID-19) Related Sentiment: Insights from Twitter Posts

WonSeok Kim, Weizhe Weng

Department of Economics, State University of New York at Geneseo, 1 College Circle, Geneseo 14454 NY, USA



GENESEO
THE STATE UNIVERSITY OF NEW YORK

ABSTRACT

The outbreak of Coronavirus Disease 2019 (COVID-19) has spread and affected many countries, causing global attention and concern. Understanding the underlying sentiment of a disease outbreak can help to keep track of spreading epidemics and provide a potential explanation for associated human behaviors. Social media, i.e., Twitter can serve as an important source to provide real-time information. Utilizing sentiment analysis, analysis of opinions can be gathered through Twitter. Therefore, performing sentiment analysis on the tweets related to the disease gives a better insight on the impact of the COVID-19 in our society. This study sheds light on how partisan preference affect individuals' sentiments. AFINN lexicon analysis has been used to rate sentiment score of each tweet, which shows that Democrats tend to obtain higher positive sentiment compared to Republicans. To support this claim, a word analysis was also conducted to identify that Republicans express more negative sentiment over words pertaining to social distancing rules than Democrats. These findings may provide a better understanding of the development of public discourse on social media and the difference in pattern of behaviors in compliance to physical distancing rules according to an individual's partisan identity.

INTRODUCTION

An ongoing epidemic of Coronavirus Disease 2019 (COVID-19), caused by SARS-CoV-2, started in December 2019. COVID-19 was first identified in Wuhan, the capital of Hubei, China. As of Jul. 30, 2020, more people have died from this disease than from SARS — which killed about a million people worldwide, according to the World Health Organization. Social media today evolved as a basic medium of communication among Internet users worldwide. Twitter is one of the popular social media platforms where people interact by creating tweets. This serves as a means for individuals to express emotion and opinion about different topics. The outbreak of COVID-19 has resulted in attracting public interest worldwide and led to millions of Coronavirus-related tweets. Since January 21, 2020, there have been over 700 million tweets worldwide and more than 450 million tweets tracked in U.S. about the COVID-19. These tweets portray Twitter users' personal attitudes and emotions during the pandemic.

This study provides a wider overview of public sentiment by using longer term of tweet data while most of previous research focus on the span of early stages of the COVID-19 pandemic. Using tweets data related to COVID-19 between February 1, 2020 and June 30, 2020, I evaluated an AFINN lexicon score of Twitter data to analyze a list of English words and their associations with the sentiment score. In addition, this research measures partisanship in conversation through Twitter and explores the impact of partisan preference on an individual's sentiment. This research contributes to the literature in a way that this study shed light on the interaction between county partisanship in U.S. and public sentiment.

Keyword Tracked	Starting Date
Coronavirus, Koronavirus, Corona, CDC, Wuhan, coronavirus, Wuhan, lockdown, Ncov, Wuhan, N95, Kungflu, Epidemic, outbreak, Sinophobia, China	01/28/2020
covid-19	02/16/2020
corona virus	03/02/2020
Covid, covid19, sars-cov-2	03/06/2020
COVID-19	03/08/2020
COVID, pandemic	03/12/2020
coronapocalypse, cancel everything, Coronials, SocialDistancingNow, Social Distancing, SocialDistancing	03/13/2020
panicbuy, panic buy, panicbuying, panic buying, 14DayQuarantine, DuringMy14DayQuarantine, panic shop, panic shopping, panicshop, InMyQuarantineSurvivalKit, panic-buy, panic-shop	03/14/2020
coronakindness	03/15/2020
quarantinelife, chinese virus, chinesevirus, stayhomechallenge, stay home challenge, sflockdown, DontBeASpreader, lockdown, lock down	03/16/2020
sheltering in place, sheltering in place, staysafestayhome, stay safe stay home, trump pandemic, trump pandemic, flattenthecurve, flatten the curve, china virus, chinavirus	03/18/2020
quarantinelife, PPE shortage, saferathome, stayathome, stay at home, stay home, stayhome	03/19/2020
GetMePPE	03/21/2020
covidiot	03/26/2020
epitwitter	03/28/2020
pandemie	03/31/2020
wear a mask, wearamas, kung flu, covidiot	06/28/2020

Table 1: Keywords Tracked and Starting Date

METHODS

Accessing Twitter Data:

This research utilizes publicly available data repository of COVID-19 tweets collected between February 1, 2020 and June 30, 2020 (Chen, Lerman, and Ferrara, 2020). I exclude the non-English Tweets and retweets, but only include geo-tagging Tweets of which the locations are identified in the United States because the objects of this research are Twitter users in the United States. After filtering the data, the final data set contains **462,274** geo-tagging Tweets and consists of the following information: (1) tweet creation date, (2) full text, (3) Id, and (4) geo-coordinate (latitude/longitude of the location).

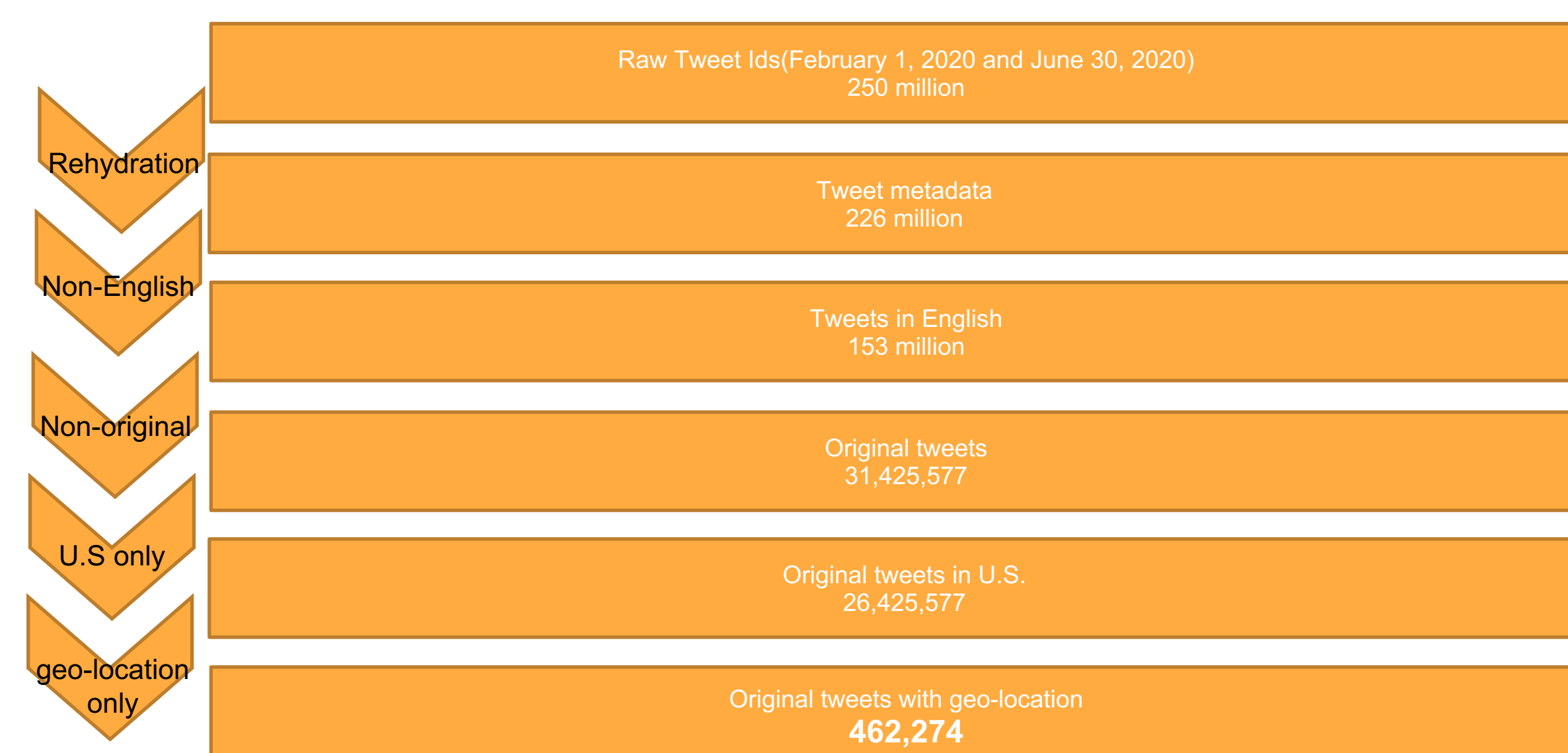


Table 2: Pre-processing Twitter Data and corresponding Tweets number

Measuring expressed sentiment :

To assess the change in users' emotions over time, this study performs a sentiment analysis using the AFINN lexicon score (Nielsen 2011). The AFINN-111 dictionary contains human annotations of emotion associations for approximately 2477 words and phrases which are evaluated for valence with an integer between negative five (negative sentiment) and positive five (positive sentiment).

Word	Sentiment Score
outstanding	5
win	4
successful	3
strong	2
solve	1
some kind	0
cautious	-1
careless	-2
catastrophe	-3
bullshit	-4
bastard	-5

Table 3: Examples of AFINN-111 dictionary and the corresponding sentiment score

Constructing Covid word lists :

I conduct word analysis on Twitter data classified by partisan identity to better understand the sentiment gap between Democrat and Republican party as well as the content of public discourse on social media.

RESULTS

Figure 1: The Wordclouds of Wordlist: Democrat (Left) Republican (Right)

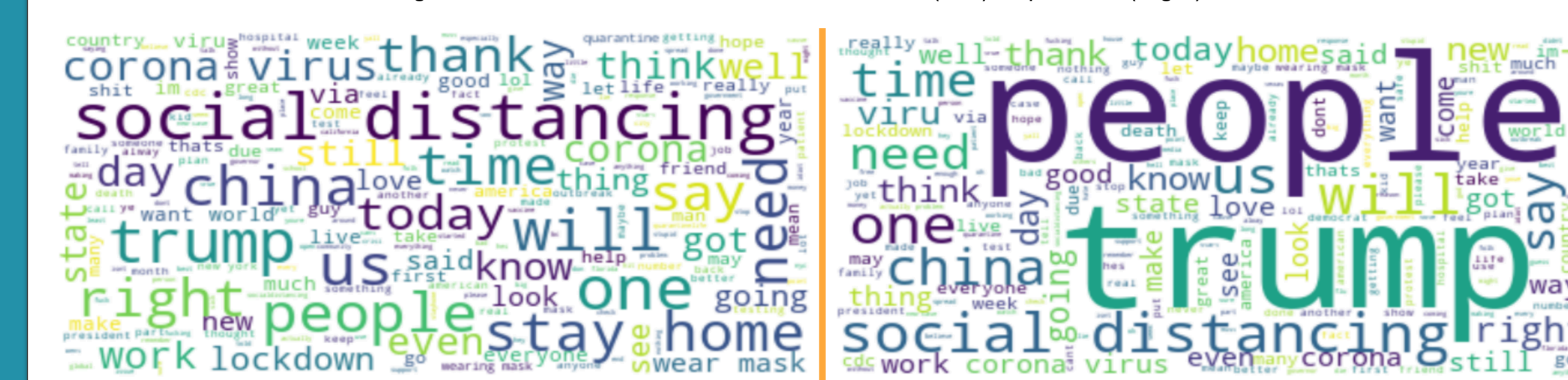
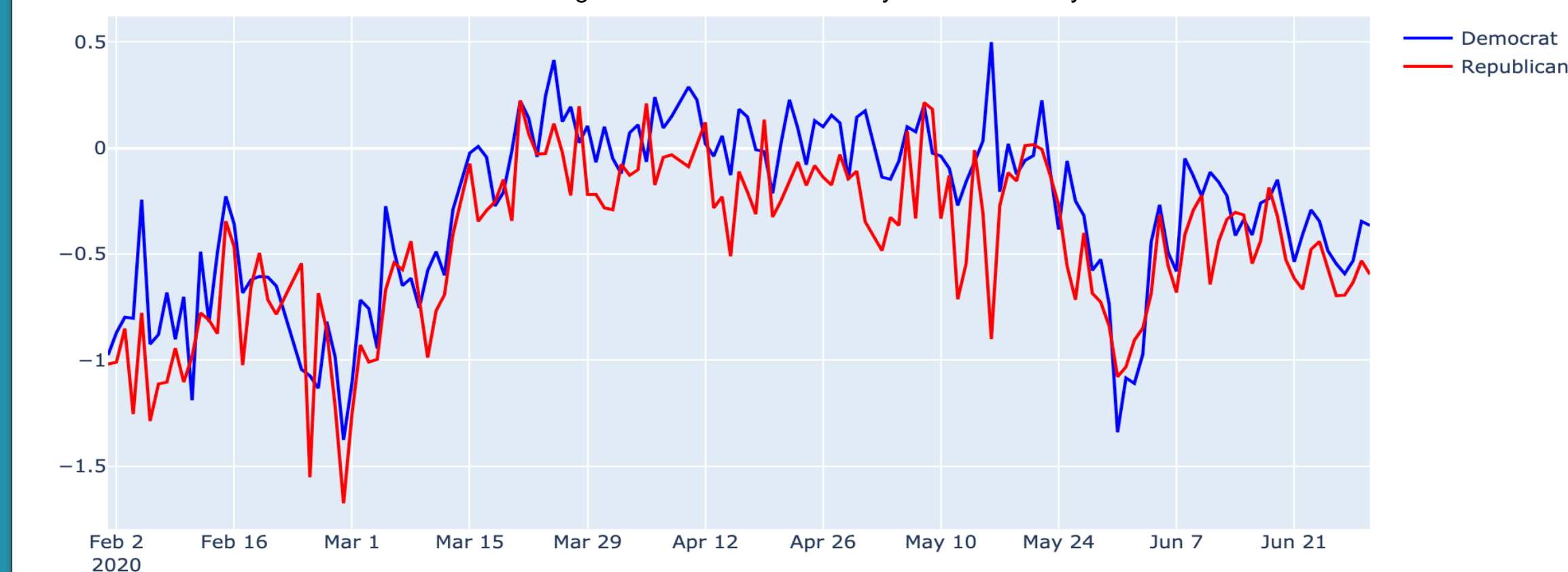


Table 4: The Most Common Words Used by Democrats (Left) Republican (Right)

Word/Sentiment	Positive	Negative	Neutral	Word/Sentiment	Positive	Negative	Neutral
covid	36.16	40.48	23.36	covid	34.74	42.80	22.46
stay	44.05	34.26	21.69	stay	39.49	36.10	24.42
us	35.86	44.92	19.22	us	35.67	46.03	18.30
home	37.97	37.56	24.47	home	34.11	41.10	24.80
corona	28.55	42.64	28.81	corona	26.68	45.81	27.51
new	33.93	34.80	31.27	new	34.42	37.56	28.02
trump	24.43	58.31	17.27	trump	24.39	58.64	16.97
china	27.03	49.18	23.79	china	26.15	50.29	23.56
virus	28.40	50.97	20.62	virus	29.37	53.50	17.13

Figure 2: The AFINN Score by Partisan Identity



CONCLUSION

Using 462,275 tweets tagged with "COVID-19" between February 1, 2020 and June 30, 2020, this paper displays the differences in sentiment of individuals by their partisan preference. I apply AFINN lexicon analysis on each tweet to classify it into the three classes of sentiments. Then, this research claims that Democrats tend to express positive sentiment toward topics pertaining to COVID-19. To support this argument, I conduct word analysis to identify the pattern of word usage across sentiments. The result of word analysis is consistent with the argument by confirming that Democrats achieve higher positive scores with terms related to social distancing rules, while Republicans tend to obtain negative sentiments over most of words that they used.

ACKNOWLEDGMENTS

This project supported by the Geneseo Foundation Undergraduate Full-time Fellowships