

Measuring the changes in sentiment and emotion towards COVID-19 in tweets posted from within the United States

The 2021 Aspiring Scientists' Summer Internship Program

James Stassinis, Zachary Richardson, Kevin Cevasco, Taylor Anderson (GGS), Andreas Züfle (GGS), Hamdi Kavak (CDS)
 Geography and Geoinformation Sciences (GGS) and Computational and Data Sciences (CDS)

Introduction and Purpose

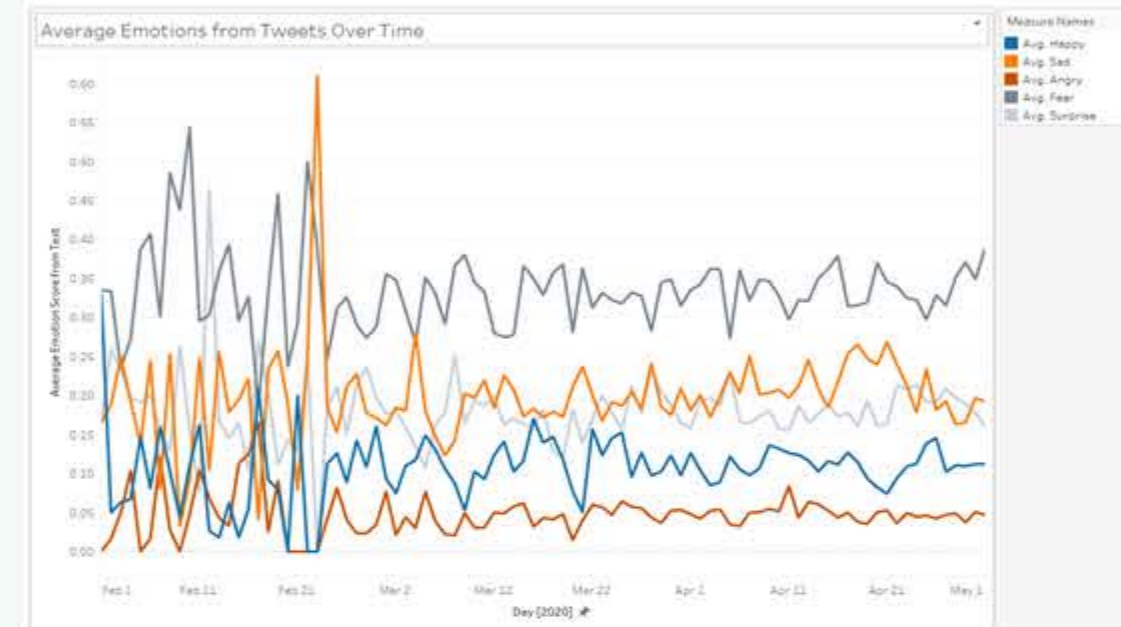
The COVID-19 pandemic was a difficult time for the world and many people took to Twitter to discuss their thoughts and feelings towards COVID-19. Understanding the sentiment and emotions felt by a population can give insight on how the general public's opinions are changing as the pandemic progresses. We can see trends in sentiment and discover if online behavior is an indicator of offline behavior in response to COVID-19. To begin our research, we used a dataset created by the Qatar Computing Research Institute (QCRI) that contains millions of tweets with enriched geolocation information attributed to each tweet.

Methods

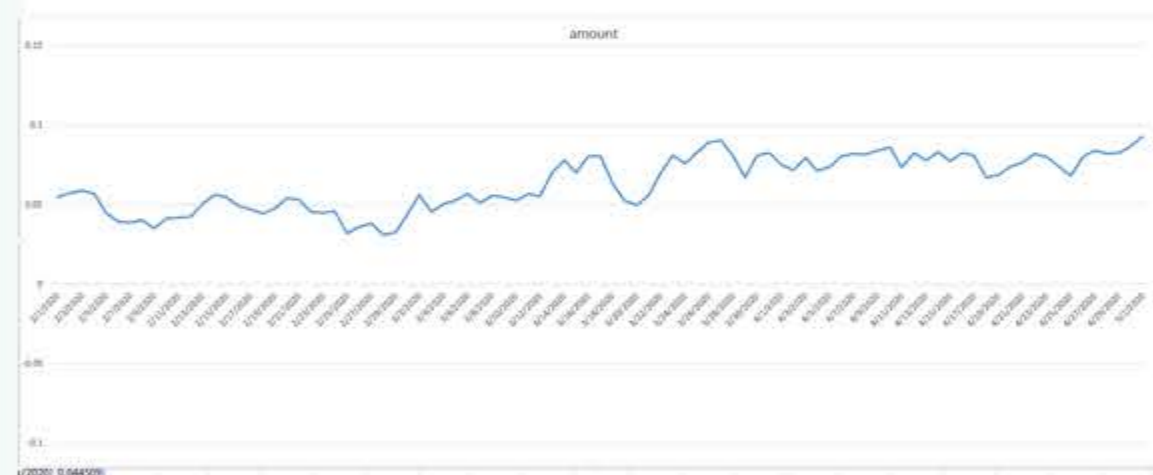
We first had to extract tweets from the Qatar dataset, then perform our own sentiment and emotion analysis. To calculate sentiment, we used the python library TextBlob to read the text within tweets and give it a score within the range -1 to 1 based on how negative or positive the attitude of text was. To measure emotion, we used the python library Text2Emotion to give tweets a percentage score based on the five emotions: happy, sad, angry, surprise, and fear. We aggregated the sentiment scores and emotion percentages by day from February 1st, 2020, to May 1st, 2020.

Results

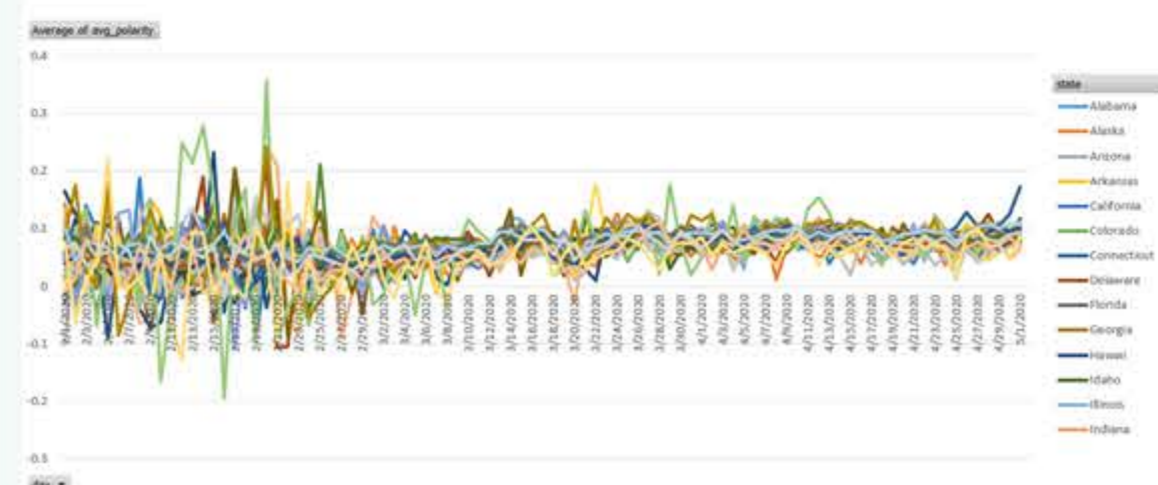
Emotions in Fairfax County



Average sentiment for entire US.



Average sentiment for each US state.



Discussion and Conclusions

We first started with tweets only within Fairfax County. We found that on average fear was the top emotion felt by Twitter users within Fairfax County, and the sentiment was neutral.

Viewing data from the entire county shows there is an upwards trend in sentiment over time, with several drops in sentiment that may be explainable by various policy changes and news about COVID-19 as it reached the general public.

Extracting and cleaning the data collected by the Qatar group took much of our time this summer. We are working with 100+ GB of data to extract, clean, and place into a PostgreSQL database for analysis.

Major Citation

Umair Qazi, Muhammad Imran, and Ferda Ofli. 2020. *GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information*. *SIGSPATIAL Special 12, 1 (March 2020)*, 6–15. DOI:<https://doi.org/10.1145/3404820.3404823>

Acknowledgements

All projects NSF Acknowledgement: This work is supported by National Science Foundation Grant #2109647. NSF Acknowledge Alternative: This work is supported by National Science Foundation Grant #2109647 titled "Data-Driven Modeling to Improve Understanding of Human Behavior, Mobility, and Disease Spread" This work is supported by a 2021 Summer Team Impact Grant of the George Mason University Office of the Provost and Executive Vice President.