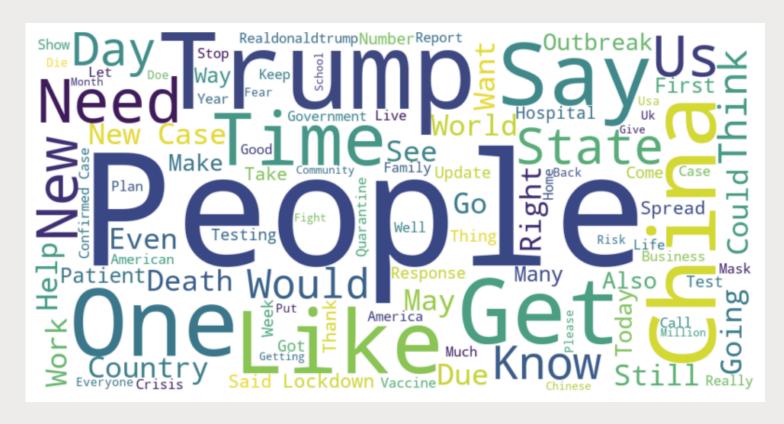
What we talk about when we talk about COVID-19?

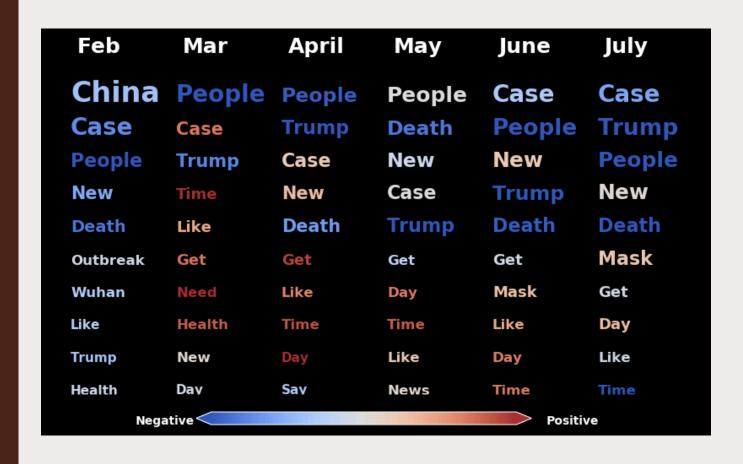
Information

- Data:
 https://twitter.com/?lang=e
 n (16 mb)
 https://ourworldindata.org/c
 ovid-cases (6.5 mb)
- NLP package: nltk, vaderSentiment, spaCy
- Machine learning model:
 - K-means clustering
 - Classification model, e.g., random forest



Word cloud of keywords related to COVID-19 on twitter.com from Feb. to Jul. 2020.

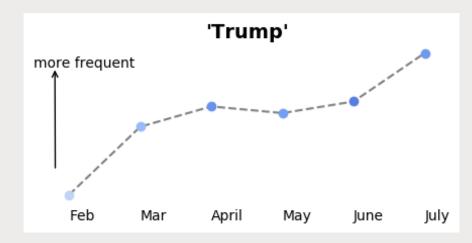
How does the sentiment change?



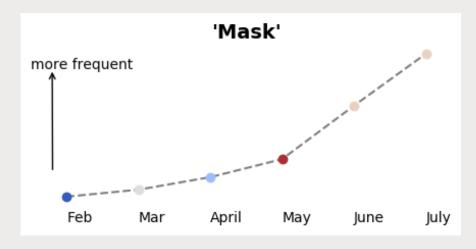
Top 10 frequent keywords each month from Feb to Jul.

Font size: frequency

Font color: sentiment of each month



Frequency (height) and sentiment (color) for tweets containing keywords



Takeaways:

-By scarping and tokenize tweets related to COVID-19, I find several keywords that people frequently mention, e.g., people, trump, etc.

-People are increasingly pessimistic while mentioning COVID-19 from February to July.

More information:

Heroku app: https://covid19-twitter-study.herokuapp.com/

GitHub: https://github.com/wandiyu/covid19-twitter/

Personal website: wandiyu.github.io

