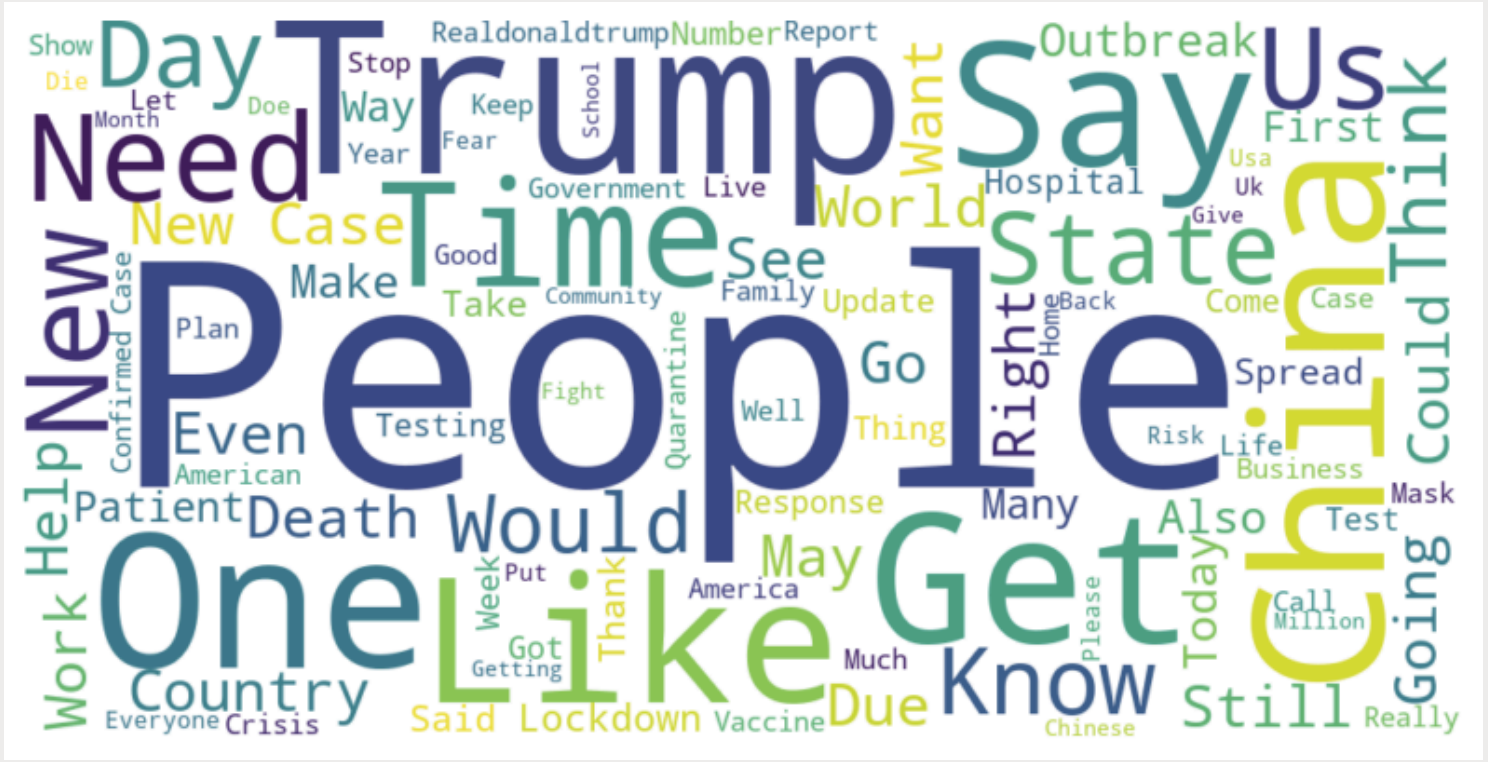
What we talk about when we talk about COVID-19?

Information

- Data: https://twitter.com/?lang=en (16 mb) https://ourworldindata.org/covid-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.

Wandi Yu
Texas A&M University

How does people's sentiment change?

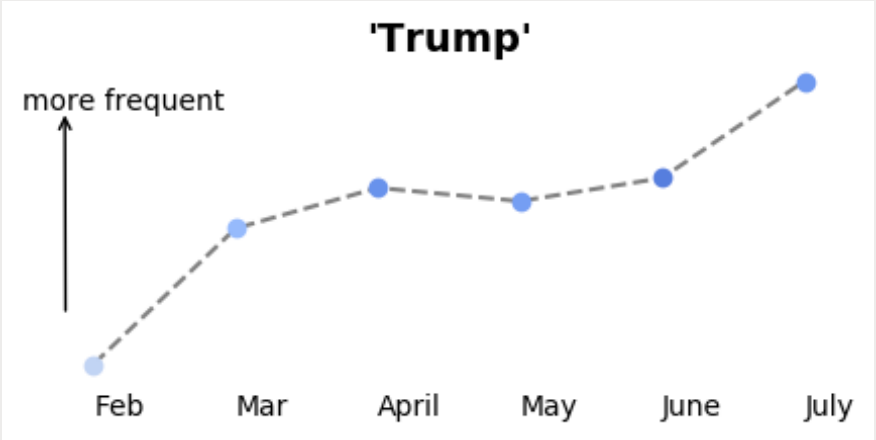
Feb	Mar	April	May	June	July
China	People	People	People	Case	Case
Case	Case	Trump	Death	People	Trump
People	Trump	Case	New	New	People
New	Time	New	Case	Trump	New
Death	Like	Death	Trump	Death	Death
Outbreak	Get	Get	Get	Get	Mask
Wuhan	Need	Like	Day	Mask	Get
Like	Health	Time	Time	Like	Day
Trump	New	Day	Like	Day	Like
Health	Dav	Sav	News	Time	Time

Negative — Positive

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.

Future work:

- Use K-means clustering model to cluster tweets into different topics, compare the result with keywords.

- Build up a classification model application to predict whether each keyword/topic will be hotter given the trend of positive cases.

More information:

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

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

Personal website: wandiyu.github.io