Visualizing Sentiment of COVID-related Tweets

Gwen Rathgeber and Eu Jin Lee
DSI-12 The Matt

Instructor: Mahdi Shadkam-Farrokhi

Repo (GA only):
git.generalassemb.ly/gwenrathgeber/project_5

linkedin.com/in/gwenrathgeber/
linkedin.com/in/eeujinlee/

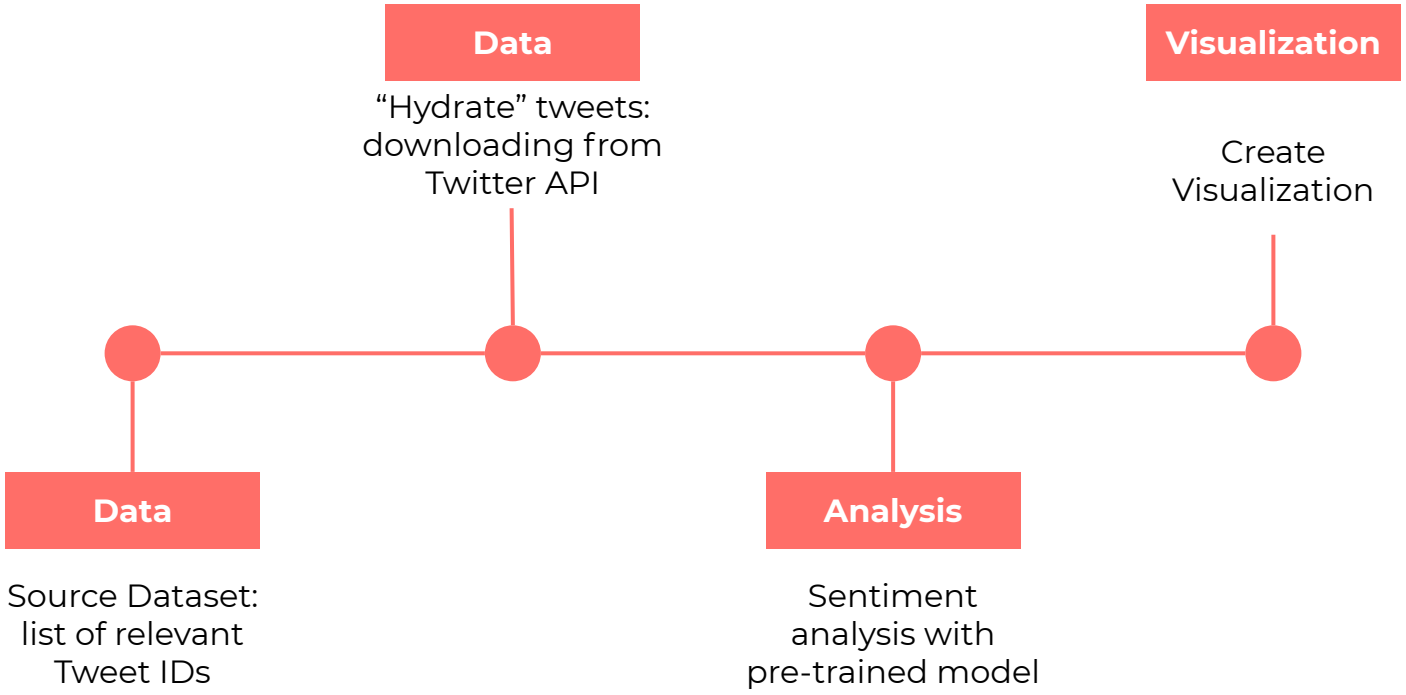
PROJECT SUMMARY

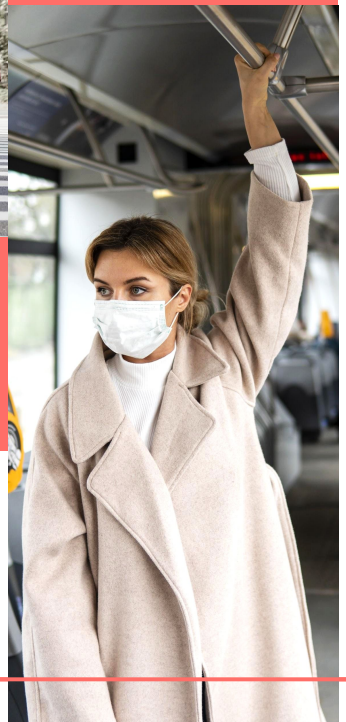
- **Goal:** shed light on the United States' reaction to the COVID-19 pandemic
  - Anyone can go on Twitter and see how the country is talking about the coronavirus pandemic, but we'd like to visualize that information in an interpretable way.

- **Limitations:** time, scope, resources
  - We needed a relatively prepared dataset and pre-trained sentiment analysis model. Twitter API rate limiting makes a custom tweet search prohibitive, and labeling a training set to train our own sentiment classifier would likewise take too long for this project.
  - The actual classification process was extremely time-intensive, taking 1.5 days to classify our final set of ~65,000 tweets.

- **Product:** a time-lapse visualization of tweet sentiment throughout the US
  - The final .gif is a 2 minute, 20 second summary of almost 5 months of the pandemic

WORKFLOW


Data
“Hydrate” tweets: downloading from Twitter API
Visualization
Create Visualization
Data
Source Dataset: list of relevant Tweet IDs
Analysis
Sentiment analysis with pre-trained model



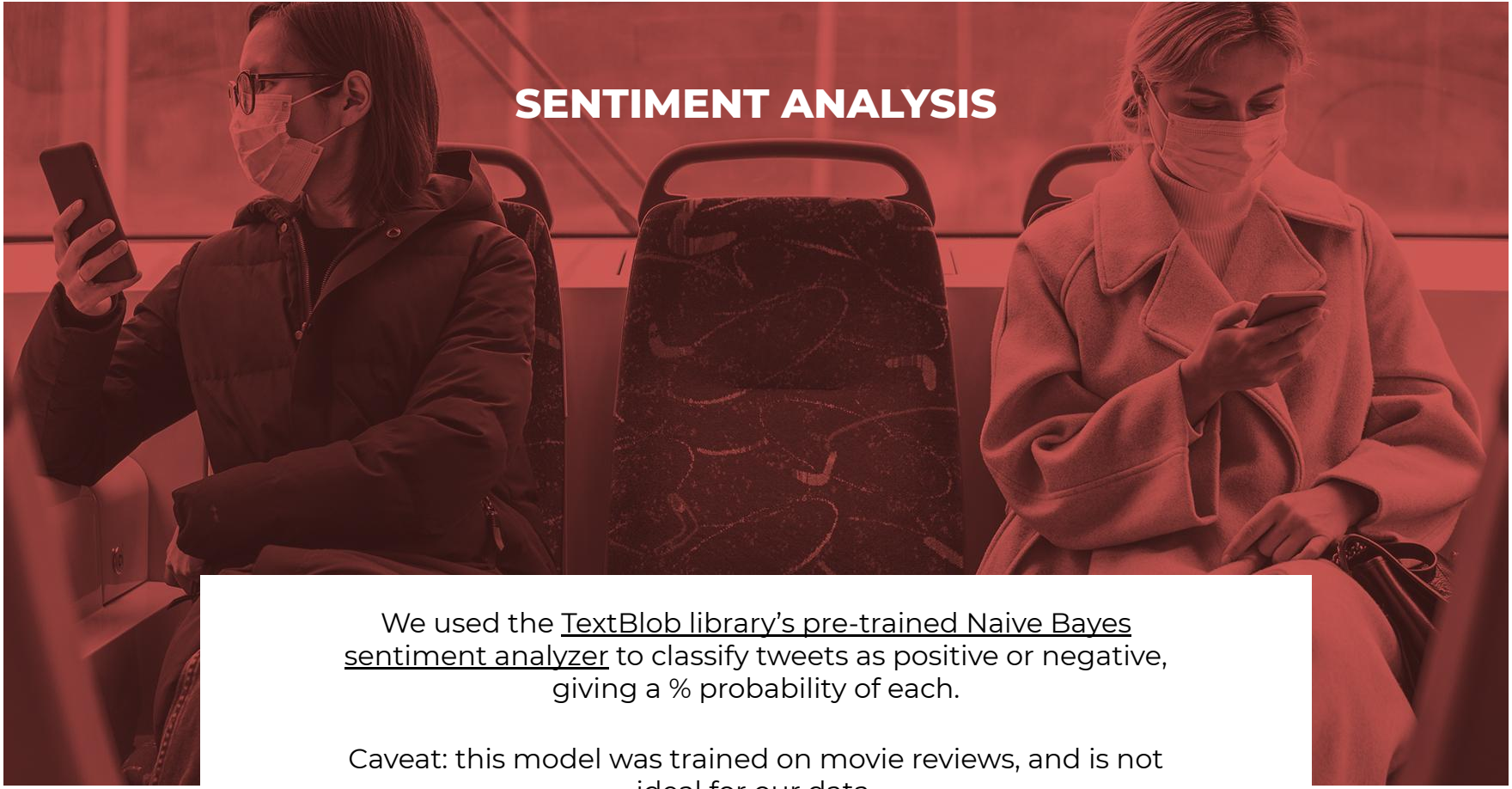
DATA SOURCE

~165,000 geotagged tweet IDs containing coronavirus-related words and hashtags

Downloaded and merged list of IDs from the IEEE Dataport

TWEET HYDRATION

Used the twarc command-line tool to request all information about each tweet from the Twitter API



SENTIMENT ANALYSIS

We used the TextBlob library's pre-trained Naive Bayes sentiment analyzer to classify tweets as positive or negative, giving a % probability of each.

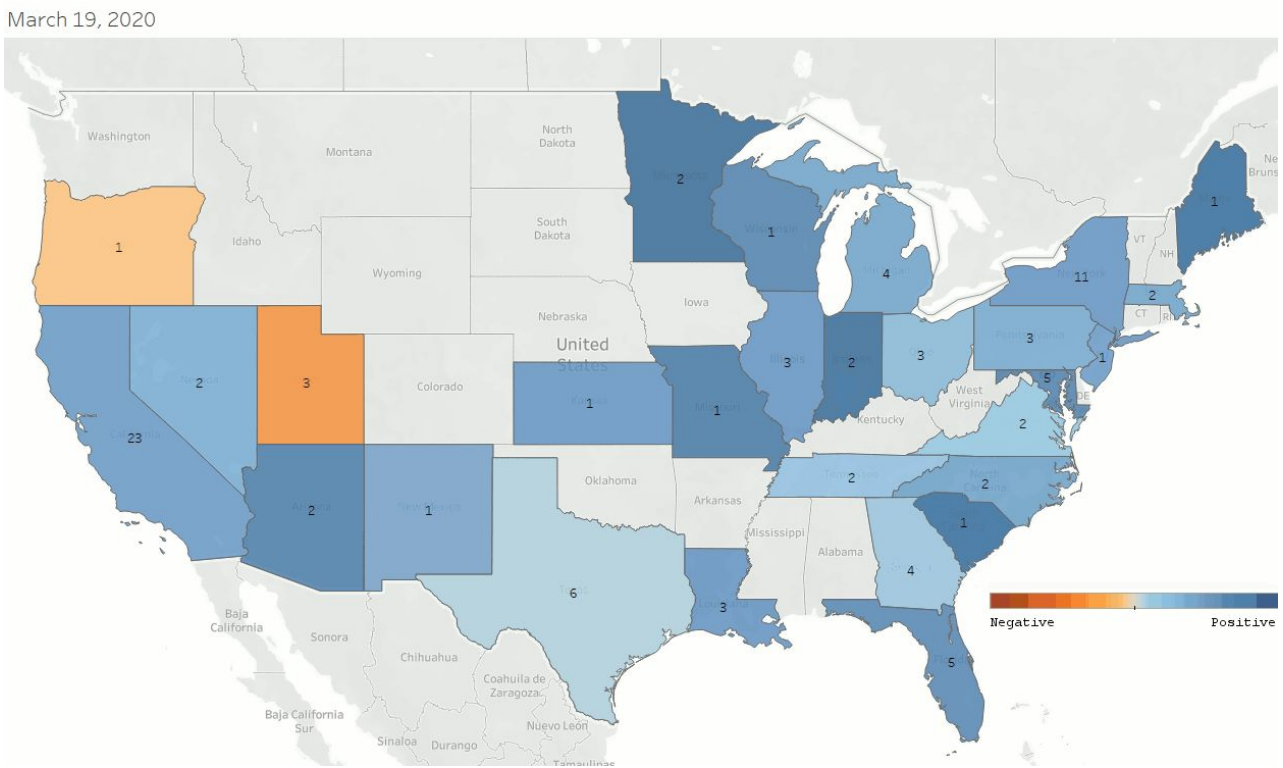
Caveat: this model was trained on movie reviews, and is not ideal for our data.

RESULTS

Scale:
Orange - most negative
Blue - most positive

Labels: Number of tweets considered in average

Transformations:
Rolling average of the previous 4 days



FURTHER WORK

REFINE SENTIMENT ANALYSIS

Try a wider variety of pre-trained models, or use an API.

FIND MORE DATA

Twitter API limitations restricted our options, but getting a higher-tier API key or finding a larger dataset would improve the results.

CREATE ADDITIONAL VISUALIZATIONS

Add annotations of major events, display time series, visualize more features such as retweets and likes.

REFERENCES

- Data sourced from IEEE Data Point
- Twitter API Wrapper
- TweetAnalyzer class
- Everything You Need to Know About Sentiment Analysis
- Making a request to download csv

THANKS!

Do you have any questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

Gwen Rathgeber & Eu Jin Lee

Repo (GA only):
git.generalassemb.ly/gwenrathgeber/project_5

linkedin.com/in/gwenrathgeber/
linkedin.com/in/eeujinlee/