Visualizing the global spread and effect of COVID-19

Group 1: Micaela Flores, Sheetal Laad, Ziv Schwartz, Kelly Sooch, Orion Taylor

Data

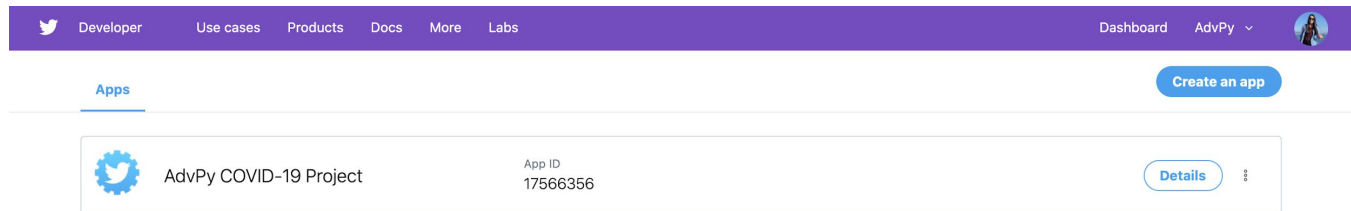 Twitter data hashtagged with #coronavirus, #covid19, #pandemic, #cdc, #n95, etc.

 Vantage Alpha (ex. DJI, Lyft, Clorox)

 Our World in Data

Twitter Data

- Create Twitter Developer API application
- Three attempts to obtain tweets:
  - Manually pulling tweets → Premium free-tiered API offered only a limited number of API calls per month (180 API calls/month or 5k tweets/month)
  - Using Python libraries like *tweepy* → issues with Twitter API rate limits/date limits
  - **GitHub repository of collected tweets since January 21, 2020**
- Only tweet IDs are to be publicly published for non-commercial research purposes
- Hydrating Tweets
  - Taking each tweet ID and using the Twitter Search API to retrieve its metadata
  - Hydrating process extremely long to complete due to volume of tweets and tweet fetching rate limits (all 29 days of February 2020 took one week)
- Created script to convert raw hydrated JSON files into a dataframe with tweet creation date, text, location, country code, latitude/longitude, etc.

Twitter Data (cont.)

- Uploaded hydrated per-day Tweet pickle files to AWS S3
- Conducted preliminary location cleaning manually, using country and state coding
- Used *geopy* API calls to standardize remaining locations, compiling a master dictionary for mapping
- Accumulated daily Tweet counts by country, and mapped resulting data using *plotly* scatter_geo
- *Employed multithreading at point of upload to S3 and in cleaning process with individual DataFrames

Financial Data

- Integrated Vantage Alpha's daily stock API

- Filled in missing records (weekend) with their most recent value

- Got data for 41 different companies from different industries

- Experimented with different ways to analyze stock prices

- Created a visual that normalizes each company's stock price based on its

  Nov 2019 through Jan 2020 stock prices

Coronavirus Data

- Found a csv consisting of count of new cases per country per day from Our World in Data
- Retrieved latitude/longitude coordinates for each country using geopy
- Created heatmap using *plotly* to show new cases over time on a world map

This graph shows the normalized stock indices for different companies between Jan. 1, 2020 and March 31, 2020. The dropdown menu allows to subset the companies by industry and the range slider allows for smaller window analysis.

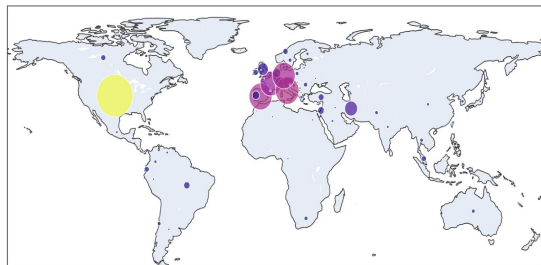## Hypothesized Stock Index

This graph shows the normalized stock indices for different companies between Jan. 1, 2020 and March 31, 2020. The difference in this graph is that it shows our teams hypoth[...]ow different companies would fare duing the pandemic: increase or decrease stock. The dropdown menu can show explicitly which companies we hypothesized to increase or decre[...]n a smaller window analysis.

Dashboard

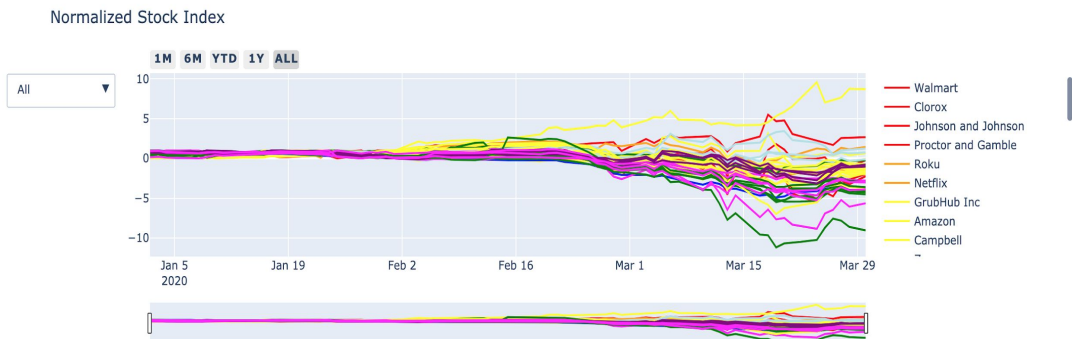New COVID-19 Cases Reported per Day

This graph shows the spread of new COVID-19 cases by day across the world from Jan. 1, 2020 and March 31, 2020.

Daily COVID-19 Tweets

This graph shows the quantity of users on Twitter tweeting about COVID-19 around the world between Jan. 1, 2020 and March 31, 2020.

Dashboard

Normalized Stock Index

This graph shows the normalized stock indices for different companies between Jan. 1, 2020 and March 31, 2020. The dropdown menu allows to subset the companies by industry and the range slider allows for smaller window analysis.

Hypothesized Stock Index

This graph shows the normalized stock indices for different companies between Jan. 1, 2020 and March 31, 2020. The difference in this graph is that it shows our teams hypoth[...] how different companies would fare duing the pandemic: increase or decrease stock. The dropdown menu can show explicitly which companies we hypothesized to increase or decre[...] a smaller window analysis.