Final Project Report
Visualization of Tweets about COVID-19

Leonardo Craven | Allison Oborn | Timothy Stander
Group 1

Overview and Goals

Our goal was to create a visualization for tweets about COVID-19 around the globe. We wanted to be able to see specific areas that mentioned COVID-19 and compare the number of tweets to other areas over a period of several months. We were also hoping to visualize how certain hashtags spread from one location to another, and find some interesting trends about spreading information globally. We chose this visualization because we thought the idea of a global visualization would be very interesting visually, and we are very pleased with the visual appeal of the final result.

Background

We used https://cesium.com/ for a lot of our research and to create the globe that we used for our visualization. Cesium is a platform for 3D Geospatial visualizations, and we found it had some very useful packages for our project. We got the data from https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset#files and we used some of the information on that site as well to learn about and adapt the data. We used the Twitter Developer Platform to learn about hydrating the data: https://developer.twitter.com/en/docs/twitter-api, which was very helpful in creating the dataset that we worked with.

Description

Our project focused on the visualization of tweets globally. The dataset included tweets that mentioned COVID-19 from October 2019 through February 2020, which was around the beginning of the outbreak of COVID-19, so it was interesting to see which areas

around the world reacted first, and how the information spread over those months on the Twitter platform. After completing our project, the main question we still have is how these trends continued after this time period. We didn't want to overwhelm our computers with data, so we only visualized those first few months, but it would be interesting to see the trends over a longer period of time. It would also be interesting to know how the data would change if everyone had their geolocation enabled. Our dataset was limited to those who had location services enabled, which appears to be a pretty good sample size, but it would be interesting to see how that data might change if we didn't have that limitation.

Learning Outcomes

While working on this project, we learned that the collection and manipulation of data can take a lot of time. We knew that hydrating the tweets would take a while, but it also took a lot of time and effort to manipulate the data so that we could actually use it for our visualization. We also got really worried when we initially visualized our data and there were hundreds of tweets coming from Antarctica and nothing coming from the US, but the fix was as simple as switching the latitude and longitude values, and then the visualization looked a lot more believable. We also learned that it's best to start with something simple and then build on the visualization from there. At first it was a little overwhelming to think about all the possibilities for visualizing our data, but we just started with a small amount of specific data, and then expanded from there.

Changes

We accomplished the overall goal of the project, which was to visualize the tweets globally, but there were a few smaller goals that we had to modify along the way. We changed our original tweet topic from Ukraine to COVID-19 because the data for tweets about Ukraine wasn't available to us, but we were still able to visualize a common twitter topic, we just needed to make sure we could actually access all the data we needed. We also adapted our goal of visualizing the spread of certain hashtags because that turned out

to be very complicated, so we decided to simplify our implementation of that objective. We still have a visualization of the spread, it just involves less data.

Evaluation

Overall, our project was successful because we accomplished our main goals, and were able to adapt to the changes that we had to make, and overcome the obstacles that we encountered in our visualizations. The strengths of our project are the dataset and the global visualization. The dataset is even bigger than we originally planned for, and it is very organized, so we were able to use it for several visualizations. The global visualization turned out to look very exciting, and visually appealing, so we felt that was a strength as well, however the weaknesses of the project would probably be our visualization of the spread of certain hashtags because we didn't quite have enough data to see a really impressive visualization of how those hashtags spread globally over time. The dataset simply didn't have enough repeated hashtags during the time frame, but we still did our best to visualize some spreading of data.

Implementation

Our first step in implementation was collecting and hydrating the data, and then we had to manipulate the data so it would work with the library that we were using. Then we had to host a web server using nodejs, so that we could implement the cesium package and visualize our global data. Then we used cesium to visualize the spreading of hashtags during the timeframe, and implement our final visualization. We split the responsibilities as follows: Leonardo was in charge of researching data collection techniques and implementing the hashtag spread visualization. Allison set up the account to collect the data, wrote the python script to manipulate the data into a usable format, and was in charge of organizing the project reports. Timothy hydrated the tweets, set up the nodejs server, and created a time slider for our global visualization. Everyone helped out with several other tasks, but these were our main responsibilities.

Additional Information

The software we used in the project includes Nodejs, Cesium, Matplotlib Pyplot for the charts, and we used github to share the code. The link to our github repository is: https://github.com/leonardo-craven/FinalProject

Our project can be reproduced by going to the FinalProject directory and typing 'npx serve' in the command line. The site will then be accessible on localhost:3000 if available, and the html code for visualizing the data is in the gallery folder. Or you can simply visit https://timothystander.com/twitter/gallery/twitter.html to view the visualization.