Visualization for Data Science Project Proposal

2021 in Tweets

By Ashley Lan, Hibban Butt, Kelsey Nicholson

Email Addresses u1261009@utah.edu

hibbanbutt@gmail.com

u1315045@utah.edu

UIDs u1261009, u1053351, u1315045

Repository
https://github.com/nicholke/dataviscourse-pr-2021-in-Tweets

Background and Motivation

The prevalence of social media in the past decade has changed the way people receive and exchange information. Social media has become one of the largest and most popular tools for trading information between friends, students, organizations, and businesses. A large network of public information is useful and important because it spreads new ideas, events, and broadens awareness of existing ideas.

Twitter is an especially powerful platform because of its ability to display trends in real-time. Trending data provides insight into not only what topics are important or popular, but specific sentiments towards news, issues, and interests. Twitter's algorithms use trending hashtags to prioritize displays of the most popular trending topics. Trending topics can be used to spread breaking news globally and because of Twitter's public API, trending data can be used for analysis. An analysis of trending topics could provide an extensive snapshot of the most important current events at a specific time on a global scale. This information is specifically interesting to our group because of our extensive use of social media for communication and news. A number of impactful and life altering current events have occurred in the past few years and we are curious about the specific events that defined 2021 across the globe.

Project Objectives

Twitter is one of the most popular social media platforms available today. It helps by providing insights into popular trends and important cultural and political moments. There are many things that have happened during the past year and we are interested in discovering the twitter trends, especially for hashtag trends based on different countries, then making visualizations on that. Here are some benefits:

- Inform people of most popular issues of 2021 on Twitter
- Be able to understand different geographic breakdown of popular issues
- Rhetoric around specific popular topics over time

Data

We are considering using twitter API to get our 2021 trends data, but twitter has a limit, it only allow us to retrieve 500000 tweets per month, so if we cannot get enough data, we will use an existing dataset, which is from Kaggle, called Twitter: Trending Tweets per week.

A link to the dataset is below: https://www.kaggle.com/datasets/rsrishav/twitter-trending-tweets. We might use the hashtag_trends_data.csv as our primary dataset, since it contains data for all trends only with the hashtag and also contains country information.

Data Processing

We are not expected to do substantial data cleanup, but a small amount of data wrangling needed. We have two options to get our data, if we are using twitter APT, we will use python, probably a python wrapper library for twitter API called tweepy, or whichever can make data processing easier, to handle data. Then we will filter applicable information. For now, we are going to get tweets date, tweets country, tweets content, then filter on hashtag. And finally we will have a csv file available for our visualization. If we are using the existing Kaggle dataset, we will need to handle the missing value for tweet_volumn column and combine those weekly csv files to a single one.

Visualization Design

For our final design we plan to have two distinct sections that will showcase different parts of the 2021 trending Twitter topics. For the first page of the website, we plan to have two different visualizations. Our project aims to look at 2021 trending topics from a global perspective, to showcase this we have decided to use an interactive world map that will update an accompanying area/bubble chart. The default state of the area/bubble chart will showcase the top 20-50 trending topics of 2021 globally. Differences in size and potentially color will be used to group trending topics by popularity. The larger the area, the more popular the topic. This method will be used to

show trends from a broad perspective and grab the attention of the user. This may prompt the user to further examine trends on the second page. The name of the trending topic will be displayed in the center of the circle. The area chart will also update based on the countries that are clicked from the world map. Once a country is clicked the default state will clear and the user can select one or several countries and then examine the most popular trends for the selected countries. Depending on the data available, we may also include a dropdown on this page that can filter data for a specific period in 2021 and then the user can click on countries to see trends at those specific times.

The second part of the project website will visualize the most popular trends of 2021 throughout the year. The second page will display approximately the top 50 trends of 2021 in text on the right side of the page. The different trends will be represented by a color scale and will update a line graph on the left side of the page, making it easy for the user to see all of the trends at a glance. The user can then click on any number of the popular trends and visualize them in a time-series chart. The lines will be represented by different colors to allow the user to visually track the rise and fall of the trends topics over 2021. The y-axis will contain over popularity which will be measured by the total number of occurrences of each trending topic. This method can be used to track trend popularity over time and see when current events became popular and for how long.

Prototype 1-Interactive world map with updating bar chart:
Prototype 1 will focus on the design of section 1 for the project. This prototype will use an interactive map and instead of an updating area chart, this design will utilize a bar chart to display varying trends between countries. The bar chart will use one country at a time and will update with the top trends and number of total counts for each trend. This design will focus on looking at the trends in 2021 for each individual country. The bar chart will have a drop-down feature that allows the user to select how many trending topics they want to examine. As this parameter is updated, the bar chart will display animation that updates that width of the bar charts depending on how many trending elements are displayed. If there are only a few trends displayed, the bars will get larger and the more trends selected, the smaller the bars will appear. For this prototype the second section of the visualization will be the same as described in the final section above. A line chart with top 50 trends of 2021 will update based on the specific trend selected by the user, this provides specific analysis of trends over time.

Prototype 2-Time-series scatter plot:
Prototype 2 will focus on designs for section 2 of the final visualization. The first section of this design will incorporate the interactive world map and updating area/bubble chart thoroughly described in the final design section. This first section focuses on the top trends for each country in 2021. The alternative design for the second section will look at the top trending topics globally and provide an in-depth view of the trends at specified time. This design will utilize a scatterplot that will allow the user to filter the scatter based on a specific time provided in the drop down. This will populate the visualization with the 50-100 top trends at the given time. The dots will have a hover feature that displays text with the trending topic and number of instances in the dataset. The user will be able to examine data on a day, month, or yearly basis. Duplicate dots will be displayed with the same colors so it is easier to see trending topics that span several days, weeks, or months

Prototype 3- Interactive World Map with popouts:
The third prototype utilizes a world map with interactive hovers and pop outs. The first section of the design will only display an interactive world map. Each country will have a specific circle that can be clicked on to examine overall 2021 trending topics per country. When a country is clicked, the small colored circle will expand and cover up part of the world map. The expanded circle will display information about the top 5 most trending topics for each country globally. This provides an interesting interactive experience for the user. This allows them to click on any country of interest and see each country's 2021 trending topics in a quick snapshot.

Must-Have Features

The essence of our project is to be able to understand Twitter trends in that we look at the most popular twitter trends by geography and time being the main parameters. To our team, this necessitates that we have the ability to visually select a country on a map and then see trending topics related to it. We then want to be able to filter based on date. We cannot set a certain expectation for granularity of date and time because that is highly dependent on the dataset, but at the very least we should be able to look

at the top trends for a year in a given country and then be able to zoom in from year to some lower unit, like month or week.

We also have to be able to visualize the trends in relation to one another. This could take the form of a pie chart, bubble chart, bar chart, or anything of that nature–what it is will be subject to the design process–but the requirement we set for ourselves is that we *must* be able to visualize the trends. We will also display a list of these trends as a requirement just as a way for the user to be able to see them in a hard-and-fast ranked fashion as well.

Optional Features

Some extra features would be for us to be able to expand upon the core visualization. We would want the user to be able to click a bar/bubble/pie-slice in the chart and have a pop-up of an enrichment of that given trend. For example, some description about the trend from Wikipedia or Google.

For filtering data by date, we may want to make it more than dropdown and add a cool calendar visualization/menu.

Finally, we would also consider optional some more advanced transition animations/techniques, for example more than bars just sliding up, maybe we have some kind of parallax effect with scrolling, etc.

Project Schedule

Week	Task	Priority Dates
Oct 17-23	-Draft Proposal -Access Datasets -Sketch project graph designs	-Oct 21-Proposal Due

	-Finalize project ideas	
Oct 24-30	-Work on obtaining twitter API access -data cleaning/wrangling	-Oct 25-Peer Feedback
Oct 31-Nov 6	-Making first visualizations -skeleton of web interface	
Nov 7- 13	-Meet with TA -First section with world map and area chart working	-Nov 11- project milestone -TA feedback by app
Nov 14-20	-incorporate TA feedback -Begin working on second design section	
Nov 21- 28	-Write up -Adding animations/transitions	
Nov 28 - Dec 2	-Polish and make final updates	-Dec 2- Project Due

Sketches-Hibban