



**Trinity College Dublin**

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

**School of Computer Science and Statistics**

## **CS7DS4 / CSU44065 Data Visualization 2019-20 Assignment 3**

Student Name: Tanvi Bagla

Student ID Number: 19300699

Declaration: "I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>."

Signed: Tanvi Bagla

Date: 19-04-2020

## 1 Introduction

The visualization “Tweet Analysis in the states of US during COVID-19” is aimed to analyze the happenings/incidents taking place in US during this pandemic. Since there is a situation of lockdown everywhere, it becomes of great importance for the country officers at work and the general public sitting at home to be informed about the activities going in their city. By analyzing the volume and type of tweets an idea can be made about the type and intensity of situation in a particular area at particular time of the day.

## 2 Description

### Graphs plotted:

- Map showing cities with most tweets
- Horizontal bars to show most frequent hashtags used
- Heat map showing volume of tweets w.r.t states/cities at particular time
- Bar graph to show number of tweets in US w.r.t time
- Bar graph to show number of tweets in US w.r.t state

### Dashboard created:

The dashboard contains all the above mentioned graphs. The key points that the dashboard displays is that how the cities are correlated with different hashtags at a particular point of time. The novelty displayed over here is how social media can help in analyzing the status of the country when the work and finance status is currently at a pause. There are several insights or results that can be highlighted from this dashboard like:

- Hashtag #shutdownflorida was used by the users on 22<sup>nd</sup> night and 23<sup>rd</sup> morning which shows Florida lockdown took place on that day
- Houston shows the least number of tweets with the hashtag #covid19, depicting the least number of cases of covid19 detected
- There was an event named Prayers for corona free world that was triggered in the entire US on 22<sup>nd</sup> March at 7AM.
- In the city of El Paso in Texas, most of the people tweeted about ‘Remote Learning’ showing the practise of working remotely

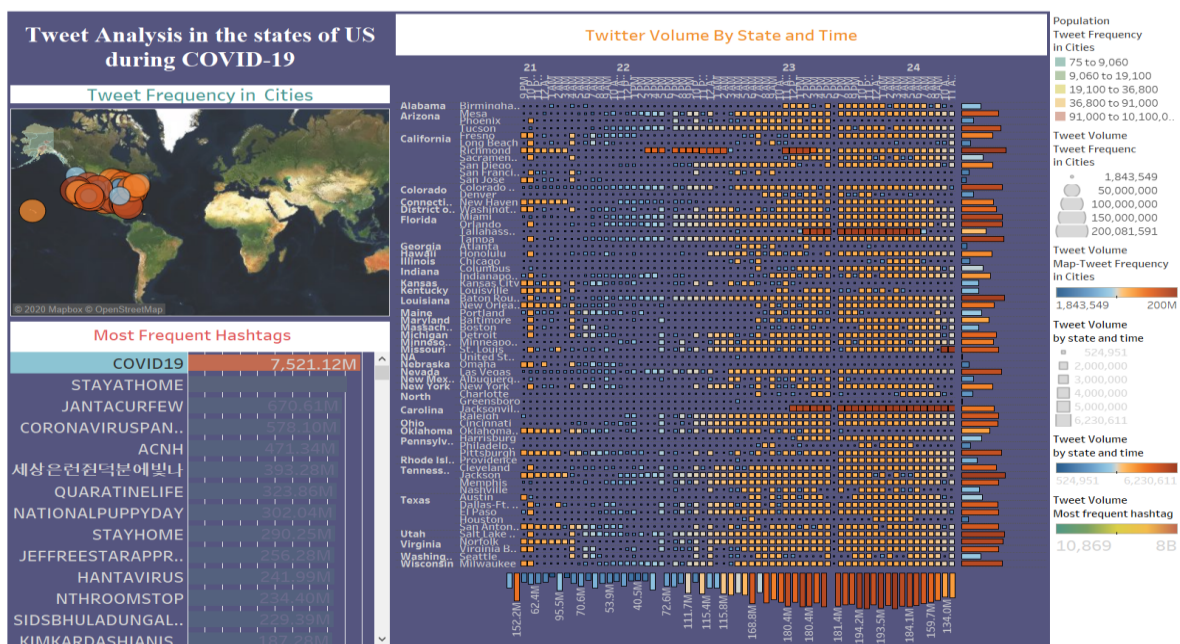


Fig 1. Dashboard of visualization

**Tool used:**

The visualizations are structured on Tableau Desktop 2019.4.5

**Dataset used:**

Data is taken from twitter API that contains number of features containing both quantitative and categorical data like hashtags, datetime, city, state, trend, tweet volume. Nominal fields like hashtag, city, trend and state and ordinal fields like datetime and tweet volume and few more are considered. The possible dataset types in the dataset are tables and fields. Tables because the dataset is in the form of rows and columns where each row stores the information about place and frequency of a particular tweet. Key attributes used in the visualization is tweet volume and state/city which is the base for the graphs made. Fields like latitude and longitude are generated in tableau that comes under field dataset type.

City1	Contains?	Datetime	Hashtag	Is english	State1	Trend	F1	Number of Records	Tweet Volume	Woied
Baton Rouge	True	24-03-2020 01:26:30	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	141	1	17,473.00	2,359,991
Baton Rouge	True	24-03-2020 02:11:36	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	129	1	21,340.00	2,359,991
Baton Rouge	True	24-03-2020 02:26:39	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	130	1	22,516.00	2,359,991
Baton Rouge	True	24-03-2020 03:11:45	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	136	1	26,362.00	2,359,991
Baton Rouge	True	24-03-2020 03:26:47	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	140	1	27,888.00	2,359,991
Baton Rouge	True	24-03-2020 04:11:54	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	140	1	30,968.00	2,359,991
Baton Rouge	True	24-03-2020 04:26:56	21DAYSLOCKDOWN	True	Louisiana	#21daysLockdown	151	1	32,174.00	2,359,991
Baton Rouge	True	23-03-2020 18:10:05	911ONFOX	True	Louisiana	#911onFOX	132	1	Null	2,359,991
Baton Rouge	True	23-03-2020 18:25:03	911ONFOX	True	Louisiana	#911onFOX	136	1	Null	2,359,991
Baton Rouge	True	23-03-2020 19:10:17	911ONFOX	True	Louisiana	#911onFOX	125	1	10,410.00	2,359,991
Baton Rouge	True	23-03-2020 19:25:13	911ONFOX	True	Louisiana	#911onFOX	132	1	10,547.00	2,359,991
Baton Rouge	True	23-03-2020 20:10:28	911ONFOX	True	Louisiana	#911onFOX	134	1	10,817.00	2,359,991
Baton Rouge	True	23-03-2020 20:25:40	911ONFOX	True	Louisiana	#911onFOX	135	1	11,108.00	2,359,991

Fig 2. Sample of the dataset used

Dataset is used for both exploratory and explanatory visualization. Data is explored by drilling down at the grass root level to find out the association and correlation between tweet trend of the users and the place from where it is highlighted more prominently. The visualizations are explanatory that present, inform and recommend the users about the latest trend going in their city and country.

**Visualization tasks used:**

- Identify if there is any hash tag that is more prevalent in a particular city. For example 'Is there a place where there is no or minimum cases of Covid-19?'
- Locate the place (latitudinal and longitudinal) in US which is following sincere social distancing.
- Cluster the hashtags according the day on which it is used. For example 'Which event is likely to have occurred on which date?'
- Compare the quality and quantity of tweets during day with those during night.
- Correlate Is there any correlation between covid19 hashtags and the number of cases in a particular city.

**Encoding channels used:**

Colour is used here to show the frequency of different tweets. Colour shades moving from orange to green where tweet volume is moving from highest to lowest.

Size encoding is used in two graphs. In map and the heatmap to show the tweet volume, where size of a bubble is bigger in case of more tweets.

Text labels are used to show the details like tweet volume, city and state

### 3 Conclusion

Concluding on a note that this visualization makes novel attempt to analyze the sector of social medial and come to some fruitful insights. I used the exploratory approach to get to know how corona is spreading in US and explained it. The strengths of this visualization are that it is easy to access and use. The general trends in US can easily be understandable. The display of map and bar graphs with colour and size encodings enhances the readability of dashboard. There is a scope of improvement in the field of data. The data related to deaths and confirmed cases can be included and linked to these graphs that can also be identified as the weakness of this visualization.

Deliverable 1 (Visualization Artefact):

<https://public.tableau.com/profile/tanvi.bagla#!/vizhome/TweetAnalysisduringCOVID-19inUS/StoryHashtagtrends>

Deliverable 3 (Presentation Video): [https://youtu.be/\\_ySS1sjbg8g](https://youtu.be/_ySS1sjbg8g)

### References

- [1] Vibhuti G., Harnessing the Power of Hashtags in Tweet Analytics, 2017.
- [2] <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>