Phase 3

For phase 3, we try to use a dashboard toll—Tableau for virtualization. It is a tool that you can load your data in local or database and then use dashboard to view the data.

First we try to view the virtualize the top hashtags using a bar chart dashboard. The dashboard has two parameters we should give definition. One is for column and the other one is for the rows. We put first column on of our sheet which contains the hashtag names to the columns on the plots shown in the dashboard. And we use a sum(F) function to get the second column which contains the data for count in the sheet. The result shows as below:

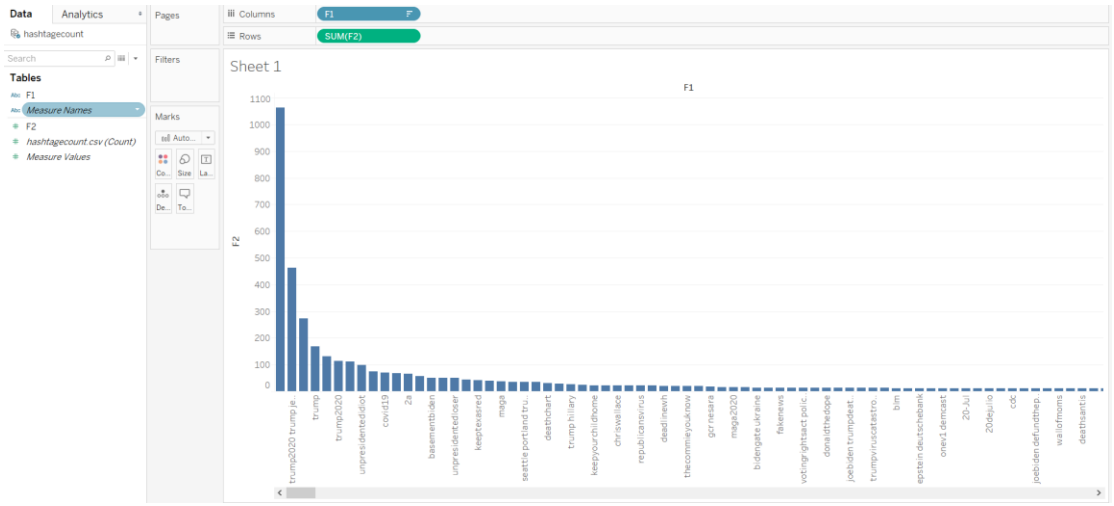

Figure 1

The x axis shows the text of the hashtags and the y axis shows the counts. The plot clearly displays the popularity of the top hashtags.

Second, we want to virtualize the change of the popularity over time. We select two type of charts. Bar chart to show the popularity count over timeline and the bubbles to show the most popular using average count at one point.

Figure 2

The bubble chart use the average of the counts of the followers as the column and the time as the row, which show the most popular with large bubbles. Also it will show the data for each point which u can check the exact data at that point.

Then we can use the bar plot below to check the real time popularity change over time line.

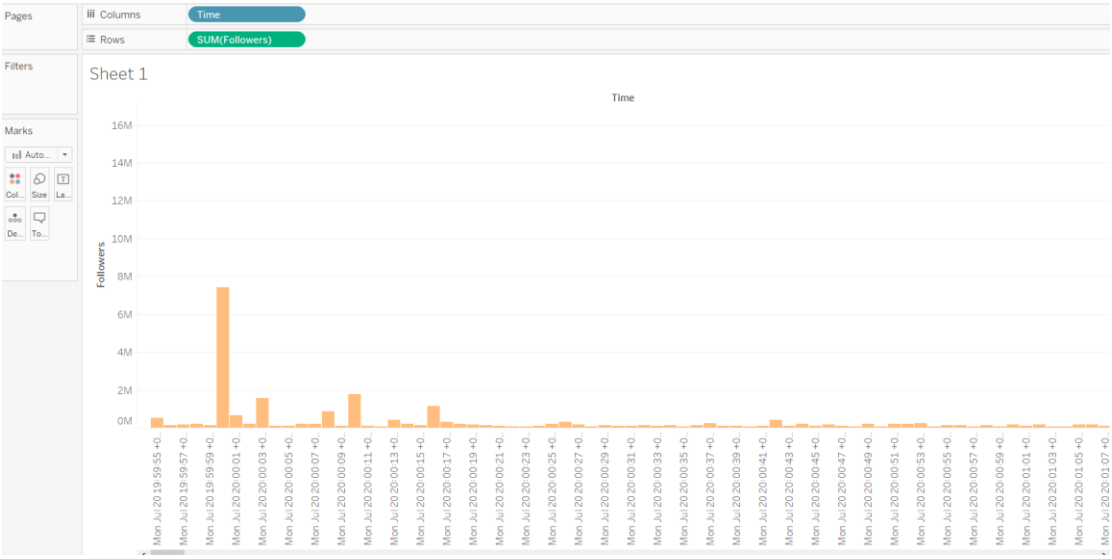

Figure 3

For language distribution, we use a new dashboard to display. We load data source languageCount.csv. And then use the sum of count for column and name of location for row. The size of the box and how deep the color is represent the data size for this location.

The mouse point in the below picture to show the data on the this part.

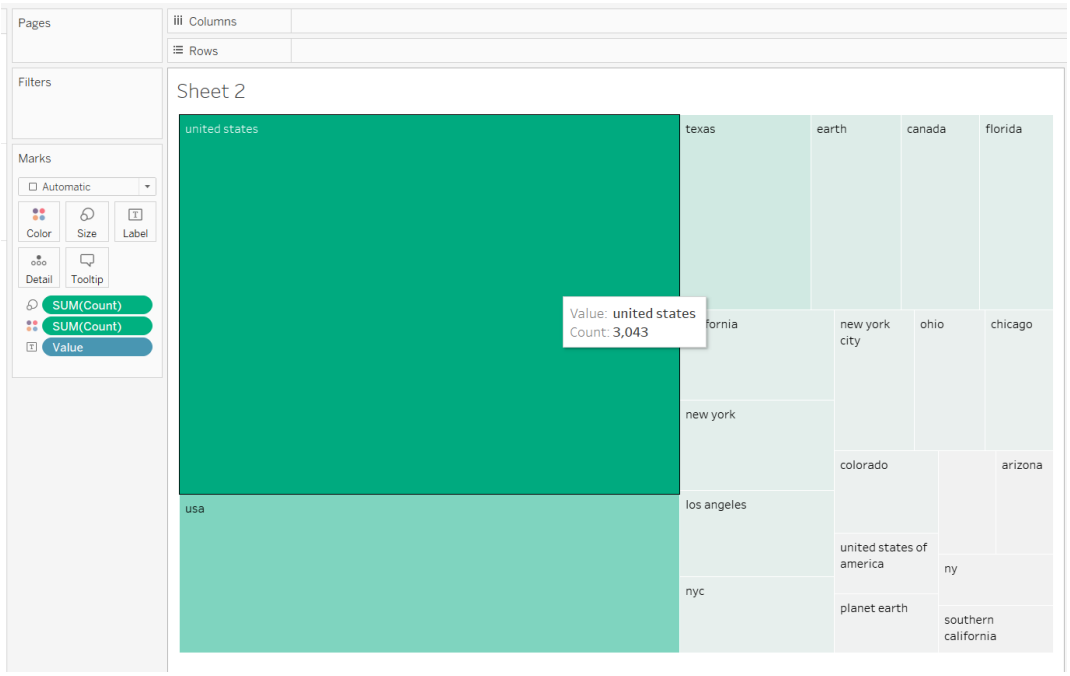

Figure 4

We also do the join table in the tableau. We choose to use languagecount.csv and languageFollowerCount.csv to join the table together based the equal of language name in the first csv file and the also the language name in the second file. We choose to use left join for the two tables.

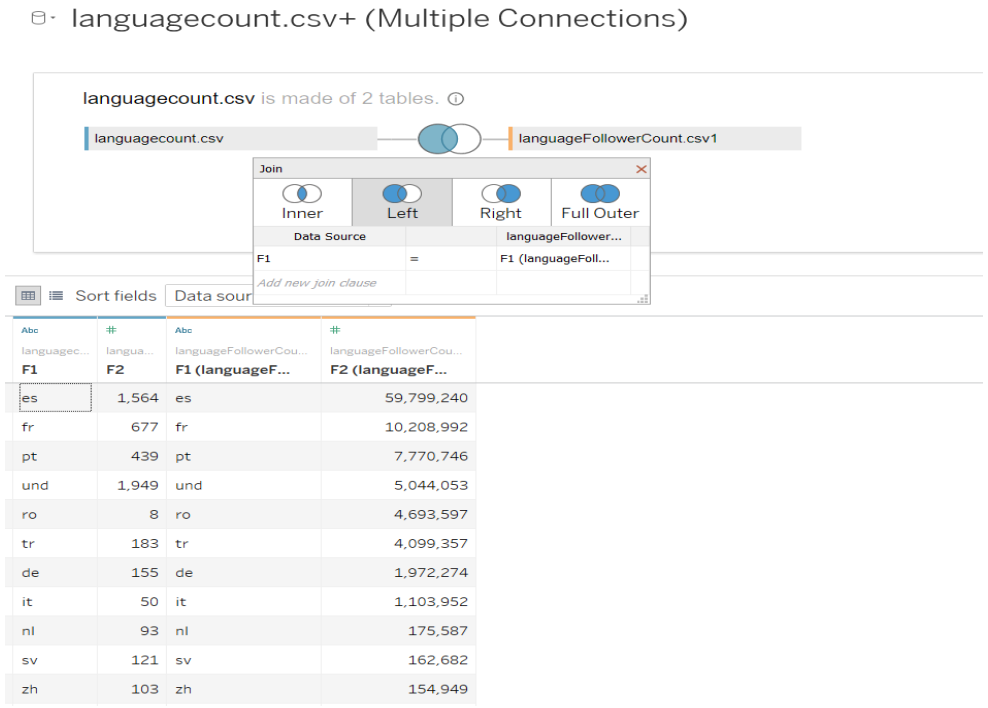

Figure 5

After join the table, we use the dashboard to compare the columns in the new table in one plots. Then we can see that the difference between the count of the tweets for each language and also the count for followers for each language.

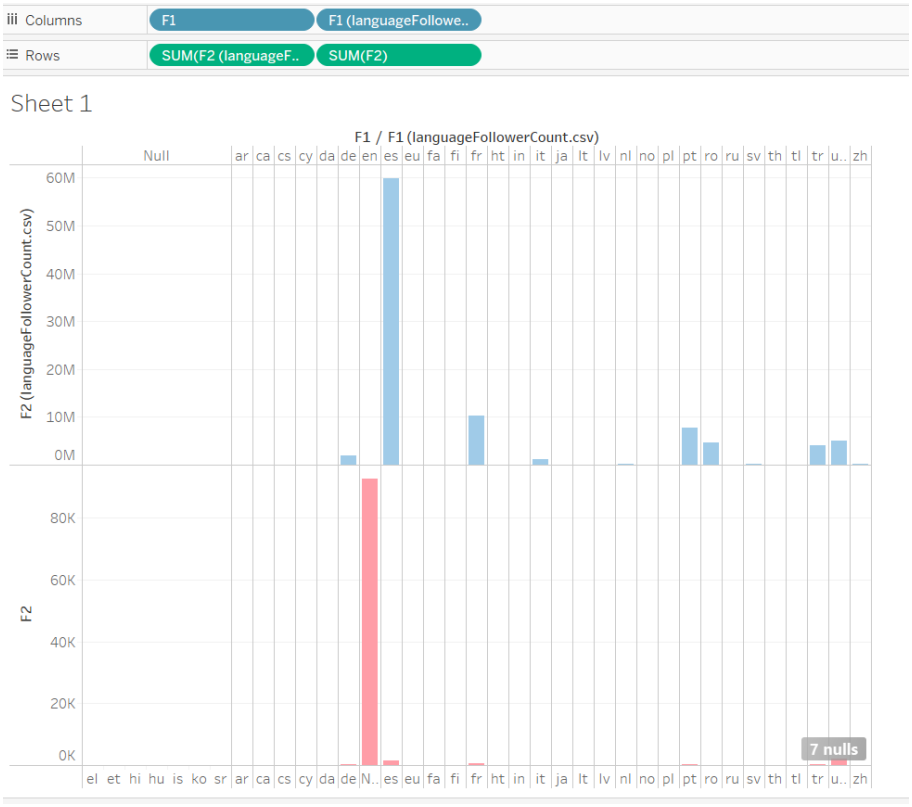

Figure 6

In phase 2, we try to use python to do the virtualization, the advantage for that one is we can edit the content and change the data or the design flexible. But the disadvantage is the design cost too much time and it is not efficienct. Also it costs much to make the plot easy for readers to read. But the tableau, which contains multile models, provide an efficient way to finish the virtualization. Though it is lack of flexibility in some aspects, it can be tolerant.