

Analysis and Comments for Part 3

1. In this part we use rtweet library and retrieve tweets that contain certain key words from Jan 1st 2019. Then clean the tweets to remove retweets and duplicate tweets. Then we consider tweets only those whose creators have locations enabled. Then get the latitudes and longitudes from google maps api and get the longitudes and latitudes for the locations of the tweets. We use these locations to get the frequency of the tweets with respect to regions or states and plot a heat map based on these results for the country of USA. It helps study the distribution of tweets over the country based on the season.
2. We perform the same algorithm for fewer keywords and again study the same pattern based on the results. Then we compare these heatmap with that of the CDC website and mark observations.