



BY HUONG PHAM

GEOCODING & CHOROPLETH MAP FINAL PROJECT GEOG-582

DATA SOURCE

coordinates	created_at	hashtags	media	urls	favorite_count	id	in_reply_to_user_id	in_reply_to_status_id	in_reply_to_user_id_str	lang	place	possibly_sensitive	retweet_count	retweet_id	retweet_src	source	text	t
	Sun Sep 20 04:05:05 +0000 2020				1	1.3075E+18	BueschelJohn	1.3075E+18	1.2109E+18	en			0		< a href="http://@BueschelJoh			
	Sun Sep 20 04:04:45 +0000 2020			https://twitt	0	1.3075E+18				en		FALSE	0		< a href="http://Vote Amy M			
	Sun Sep 20 04:04:02 +0000 2020			https://metr	3	1.3075E+18				en		FALSE	5		< a href="http://Hundreds of			
	Sun Sep 20 0 USA bernies	https://twitt	https://politi		0	1.3075E+18				en		FALSE	0		In Town Hall,			
	Sun Sep 20 0 Pune PimpriChinchwad			https://www	3	1.3075E+18				en		FALSE	1		< a href="http://Coronavirus			
	Sun Sep 20 0 coronavirus covid19			https://pape	0	1.3075E+18				en		FALSE	0		< a href="http://The latest Fi			
	Sun Sep 20 04:04:07 +0000 https://twitt	https://www			2	1.3075E+18	JohnCornyn	1.3075E+18	13218102	en		FALSE	1		< a href="http://@JohnCornyn h			
	Sun Sep 20 04:04:12 +0000 2020			https://youtu	1	1.3075E+18				en		FALSE	1		< a href="http://Virologist Dr			
	Sun Sep 20 04:04:13 +0000 2020			https://bit.ly	0	1.3075E+18				en		FALSE	0		< a href="http://via Domain:			
	Sun Sep 20 04:04:10 +0000 2020				1	1.3075E+18				en			0		< a href="http://Delhi records			
	Sun Sep 20 04:04:23 +0000 https://twitter.com/Ferde				41	1.3075E+18				en		FALSE	34		< a href="http://COVID19 ht			
	Sun Sep 20 04:04:56 +0000 2020			https://nypo	0	1.3075E+18	justinputney	1.3075E+18	21277152	en		FALSE	0		< a href="http://@justinputn			
	Sun Sep 20 04:04:22 +0000 2020			https://twitt	1	1.3075E+18				en		FALSE	0		< a href="http://Rest In Peac			
	Sun Sep 20 04:04:11 +0000 2020				2	1.3075E+18	tarblondie	1.3075E+18	165119652	en			0		< a href="http://@tarblondie			
	Sun Sep 20 0 COVID19 Coronavirus				0	1.3075E+18				en			0		< a href="http://Due to			
	Sun Sep 20 0 AWS AZURE https://twitt	https://buff.			0	1.3075E+18				en		FALSE	1		< a href="http://Is Your			
	Sun Sep 20 0 coronavirus				3	1.3075E+18	tedcruz	1.3075E+18	23022687	en			0		< a href="http://@tedcruz			
	Sun Sep 20 04:25 +0000 2020			https://twitt	0	1.3075E+18				en		FALSE	0		< a href="http://As in https://			
	Sun Sep 20 04:18 +0000 2020			https://twitt	0	1.3075E+18				en		FALSE	0		< a href="http://All the peop			
	Sun Sep 20 04:51 +0000 2020				0	1.3075E+18	MysterySolve	1.3075E+18	1.0947E+18	en			0		< a href="http://@MysterySo			
	Sun Sep 20 04:16 +0000 2020			https://twitt	0	1.3075E+18				en		FALSE	0		< a href="http://Another Sup			
	Sun Sep 20 0 covid19				0	1.3075E+18				en			0		< a href="http://The			
	Sun Sep 20 04:46 +0000 2020			https://twitt	91	1.3075E+18				en		FALSE	38		< a href="http://Good news e			
	Sun Sep 20 04:26 +0000 2020				0	1.3075E+18	BallouxFranc	1.3075E+18	1.1137E+18	en			0		< a href="http://@BallouxFra			
	Sun Sep 20 04:33 +0000 2020			https://airwa	2	1.3075E+18				en		FALSE	0		< a href="http://Air Canada t			
	Sun Sep 20 04:21 +0000 2020			https://www	0	1.3075E+18	Thoriq52188	1.3075E+18	1.2656E+18	en		FALSE	0		< a href="http://@Thoriq521			
	Sun Sep 20 05:02 +0000 2020			https://www	0	1.3075E+18				en		FALSE	0		< a href="http://Three			
	Sun Sep 20 0 COVID19	https://twitter.com/kathr			1	1.3075E+18				en		FALSE	0		< a href="http://Be kind ,Ä			
	Sun Sep 20 04:44 +0000 2020			https://www	0	1.3075E+18				en		FALSE	0		< a href="http://U.S. Covid-19			

INTRODUCTION

- What inspired me
- What data – one day of tweets relating to COVID-19, no coordinates
- What's the goal – Geocoding + plotting Choropleth maps
- What's the tool – Nominatim, GeoPandas, Matplotlib, Regular Expression, TextBlob library

PART 1: GEOCODING

- Use Nominatim, Time.sleep(1) to slow down the geocoding as required by the OpenStreetMap Foundation
- Iterate each index of the dataframe and geocode
- Get latitude and longitude and put them in a new column “coordinates”
- Delete rows that have empty strings in the coordinates column
- Output the results to a csv file for creating maps in part 2

America, you're not just poorer - you're dumber.", "Oakland, CA", "37.8044557, -122.2713563"
187802, Mon Sep 21 03:52:10 +0000 2020, Do not take Trump's vaccines. He does not care if you die <https://t.co/oL58WromXz>, "San Francisco, CA", "37.7790262, -122.4199061"
187881, Mon Sep 21 03:53:18 +0000 2020, "Emmys: Covid-19 'Test;' Jimmy Kimmel Mocks MAGA Rallies, Cracks Jokes About Russian Interference", "Los Angeles, CA", "34.0536909, -118.242766"
187921, Mon Sep 21 03:52:52 +0000 2020, "Not good, not good. <https://t.co/XboYoX12Au>", "Los Angeles, CA", "34.0536909, -118.242766"
187976, Mon Sep 21 03:54:02 +0000 2020, "CDC says coronavirus spreads mainly in the air, through respiratory aerosols and droplets <https://t.co/2DfJyQHkW>", "West Hollywood, CA", "34.0923014, -118.3692894"
187996, Mon Sep 21 03:53:48 +0000 2020, "NINE months in... CNN: Updated CDC guidance acknowledges coronavirus can spread through the air via @GoogleNews", "West Hollywood, CA", "34.0923014, -118.3692894"
<https://t.co/zfNDib7LRq>

via @GoogleNews", "West Hollywood, CA", "34.0923014, -118.3692894"
188078, Mon Sep 21 03:54:37 +0000 2020, "Now Reading: CNN: Updated CDC guidance acknowledges coronavirus can spread through the air. <https://t.co/XhfDqfjVX6>"

via @GoogleNews", "San Jose, CA", "37.3361905, -121.890583"
188081, Mon Sep 21 03:55:04 +0000 2020, "20 new cases of COVID-19 recorded in Saskatchewan push provincial total past 1,800 <https://t.co/8cDQsHbXkW>", "Ontario, CA", "50.000678, -86.000977"
188085, Mon Sep 21 03:54:22 +0000 2020, "@RachelShelley It's really hard for me to believe that there are so many that think COVID-19 is a hoax."
188115, Mon Sep 21 03:55:13 +0000 2020, "We're giving away our BRAND NEW clementine-scented hand sanitizer 🍊, along with a Hint face mask! RT this for a chance to win."

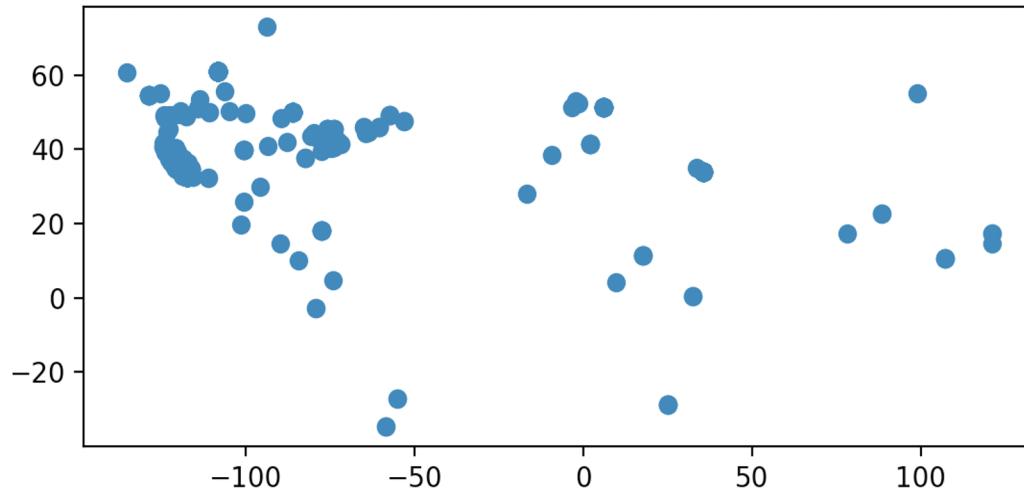
Can't wait? Get it here: <https://t.co/gd9N4aq1vG> <https://t.co/S24djDrccdk>, "San Francisco, CA", "37.7790262, -122.4199061"
188116, Mon Sep 21 03:55:18 +0000 2020, "@CrazySthrnSmurf @WattoPhotos @JesseOSheaMD Covid19 causes sepsis, cardiomyopathy, autoimmune disease, and more." <https://t.co/2DfJyQHkW>

PART 2: PLOT MAPS

1. POINT MAP

- put locations for all tweet users on a same map
- use Point geometry

Figure 1

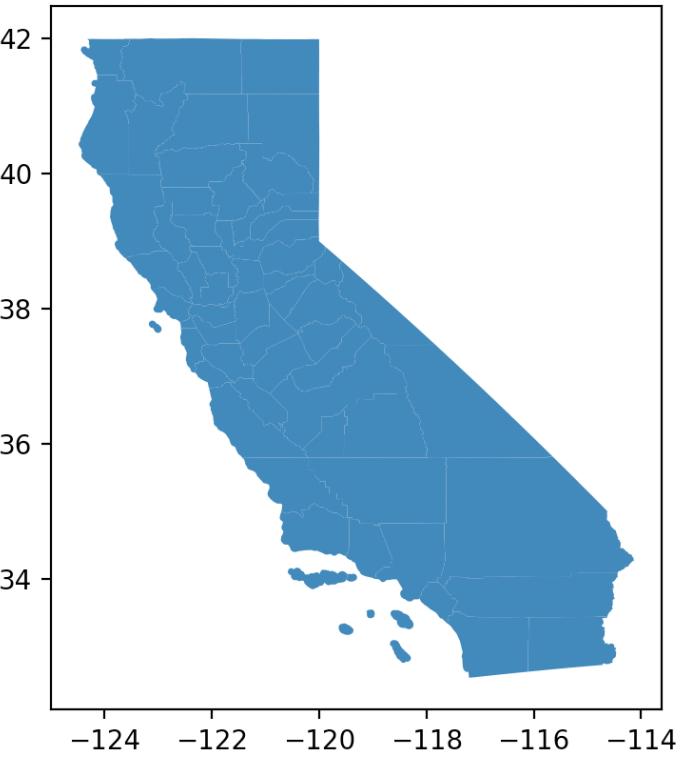


x=-107.531 y=64.2311

PART 2: (CONTINUED)

2. CHOROPLETH MAPS

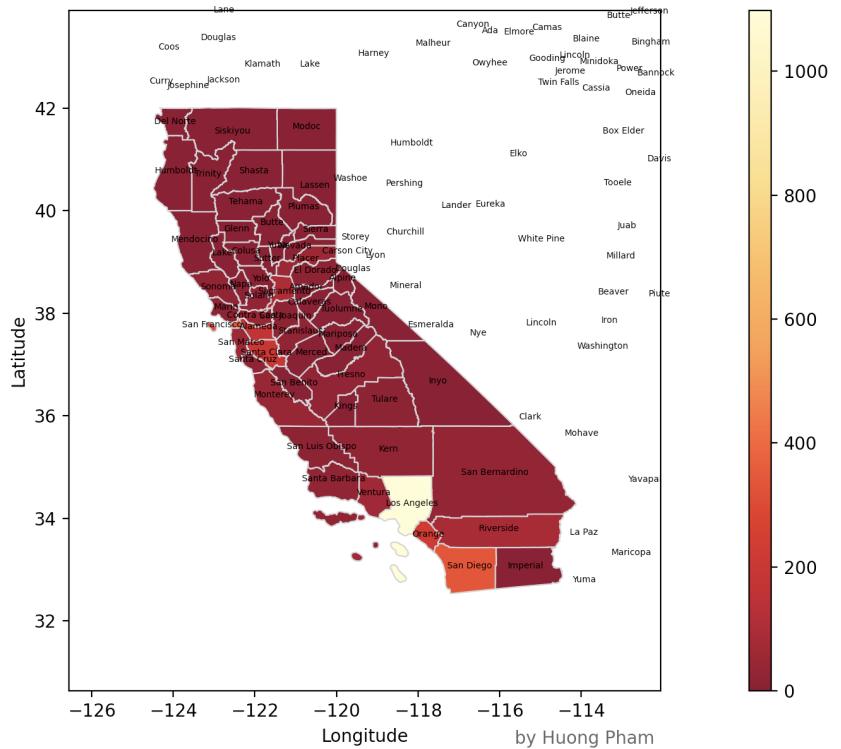
- To present what counties in California have higher number of tweets, more positive/negative/neutral tweets compared to others.
- Use polygon Geometry
- Split latitude and longitude in coordinates column
- Read a CSV file to create GeoPandas DataFrame 'geo_df_textblob'
- Get US county map from TIGER/Line Shapefiles (us_county)
<https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>



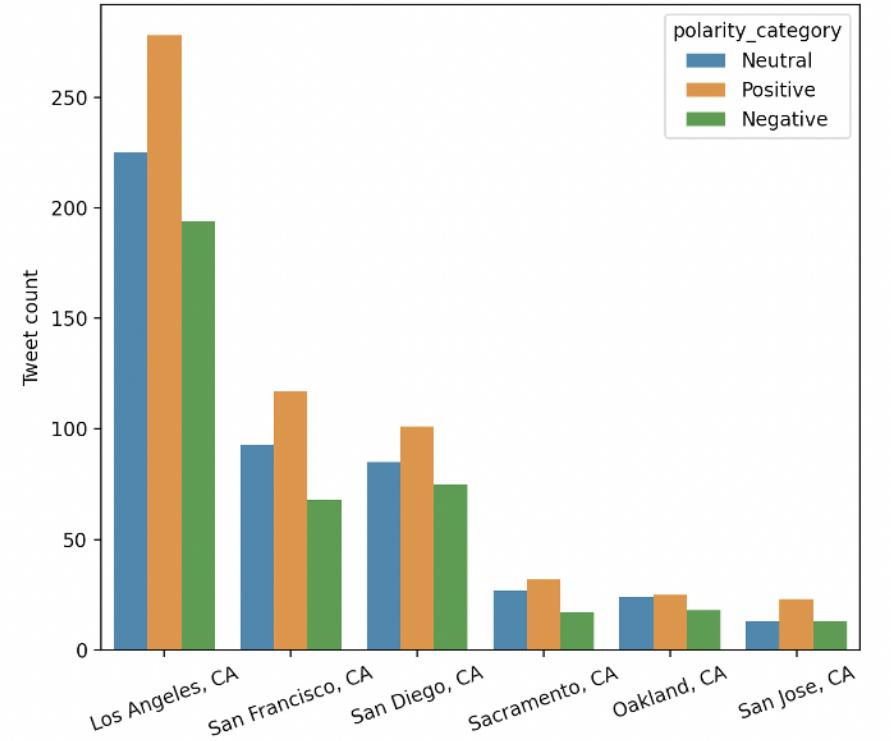
PART 2 (CONTINUED)

- Join geo_df_textblob + us_county => county_with_textblob
- Group by county and polarity category counts
- Create new columns for positive tweet counts, negative tweet counts, neutral tweet counts, and total counts of all tweets for each county
- Plot choropleth maps

Choropleth Map - total of tweets, CA counties



Top 06 Cities
data/tweet_09_20_2020_English_204K.csv



total



positive



negative



neutral



Huong Pham, Final Project GEOG 582, Fall 2020