GEOG 507 Final Project Report- Edward Tang

My project consists of a visualization map and data analysis from the data of US gun violence from Jan 1 2014 to September 2 2021. The original data can be found here https://www.kaggle.com/konivat/us-gun-violence-archive-2014.

	incident_id	incident_date	state	city_or_county	address	killed	injured
0	2128682	September 29 2021	Illinois	Chicago	800 block of N Milwaukee Ave	0	5.0
1	2127374	September 28 2021	Illinois	Chicago	65th St and Calumet Ave	1	3.0
2	2128023	September 28 2021	Minnesota	Minneapolis	2700 block of Bloomington Ave	1	3.0
3	2126747	September 27 2021	New York	New York (Manhattan)	3964 10th Ave	0	5.0
4	2126450	September 27 2021	Illinois	Chicago	826 N Ridgeway Ave	1	3.0

The data consists of 7 columns: incident_id (which is not relevant to this project), incident date, state, city_or_county, address, killed and injured.

The first step is to create a choropleth map of killed victims in gun violence by states. This is made using plotly.express library. The map is colored by counts of killed victims. As shown, the interactive map created using plotly.express can be used to visualize the victim data of each state.
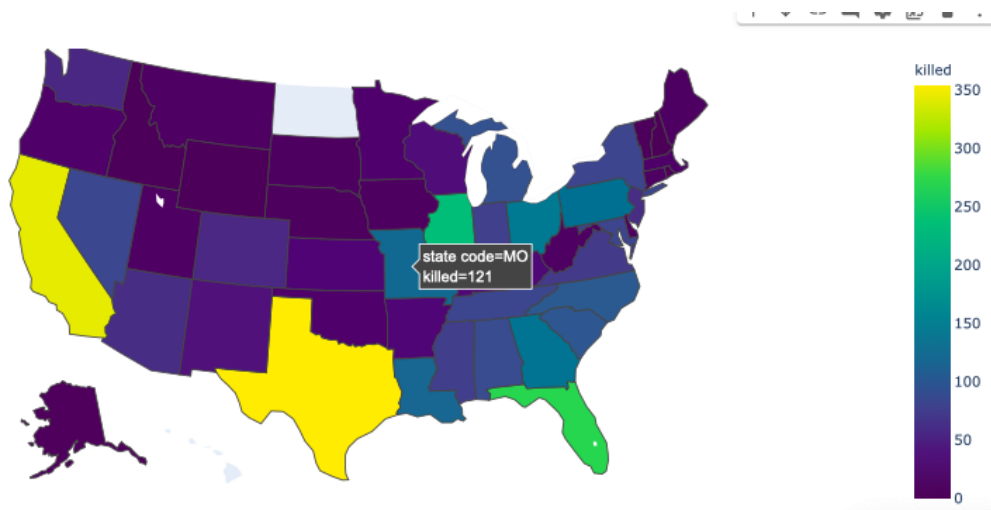
Similarly, we can do the heatmap analysis for injured victims:
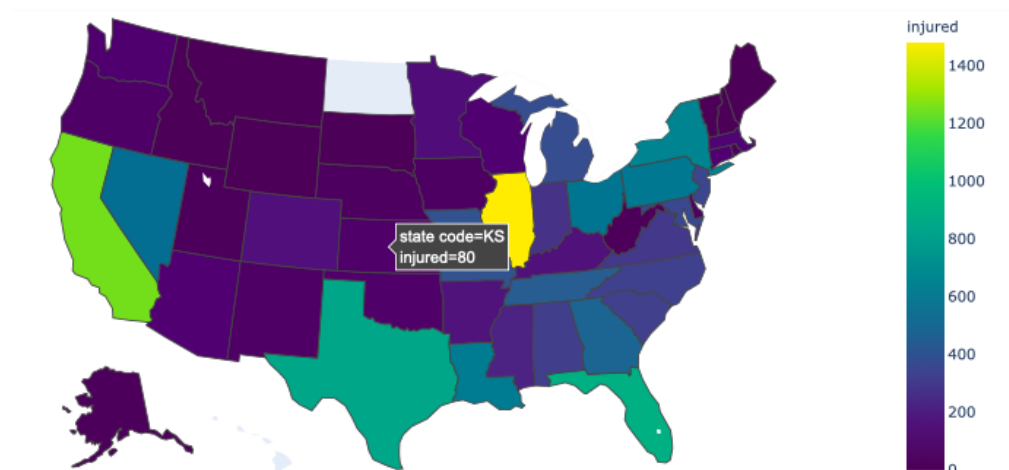
With raw data, we can map a color-coded histogram of victims by cities/counties
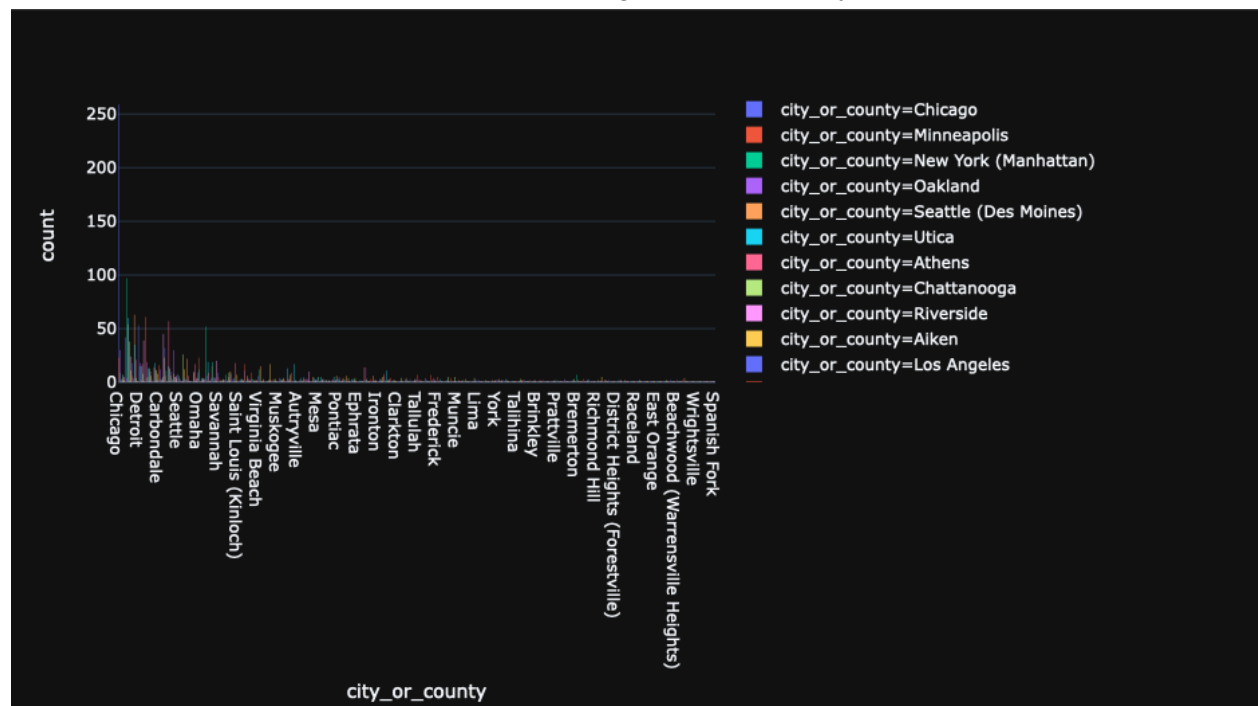

The second part of this project is to do a folium map on the US map for the specific cities/counties of the most recent 200 incidents of gun violence. Initially I tried using the exact address and using geopandas Nominatum to get their coordinates. Unfortunately, a lot of those addresses are incomplete and Nominatum has a service time limit so I cannot get through all the 3000+ data.

The map generated is shown here:
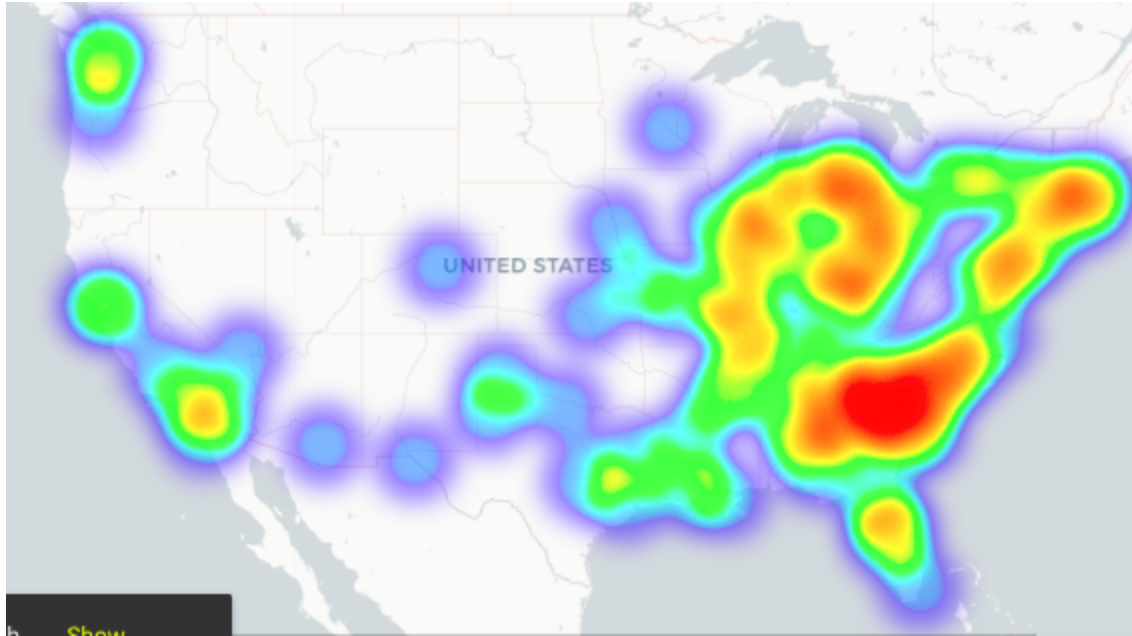
Folium map by date

Also, we can generate a folium map by date on the past 200 incidents in the list:
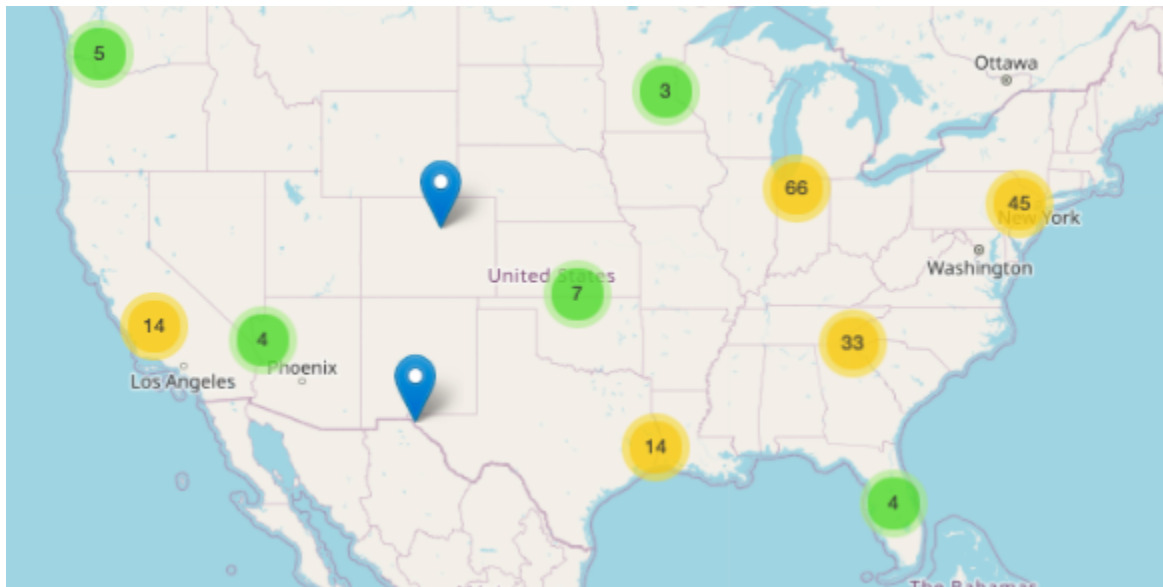
Analysis:

First it is (depressingly) unexpected that processing the amount of gun violence data could drain the service timeout limit and I had to truncate it to the most recent 200 cases. Second, while dealing with the locations on the map, it was not easy to convert the physical location to coordinates due to either 1) the incompleteness or inaccuracy of geographical address and 2) the difficulties to store them in the dataframe without extra data mining. Eventually I stored the

latitude and longitude into numpy arrays and concatenated them together in order to construct the folium heatmap.