Team O Group Project Report: POIMAGIC: Early Warning System

### The goal of the project is to create an Early Warning System called POIMAGIC for streaming spatial events based on Earthquake hotspots and then to study if the developed system can be reused to detect Earthquakes in a different region, with different amounts of available data. We also assume a hotspot is a contiguous polygon in a 2D longitude-latitude space for which the event density of points inside the polygon is above a user-defined density threshold. The system should create two kinds of hotspots:

### a. Small, very hot spots whose density is above a “high” density threshold D1

### b. Large, more regional hotspots whose density is above a “medium high” density threshold D2; D1>D2.

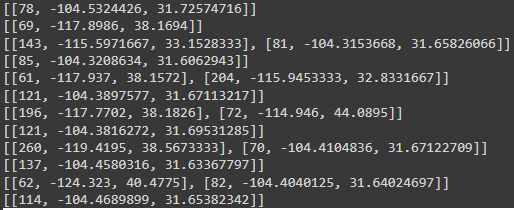
### How we plotted the data and determined what is D1 and D2 data:

( **United States of America Dataset** )

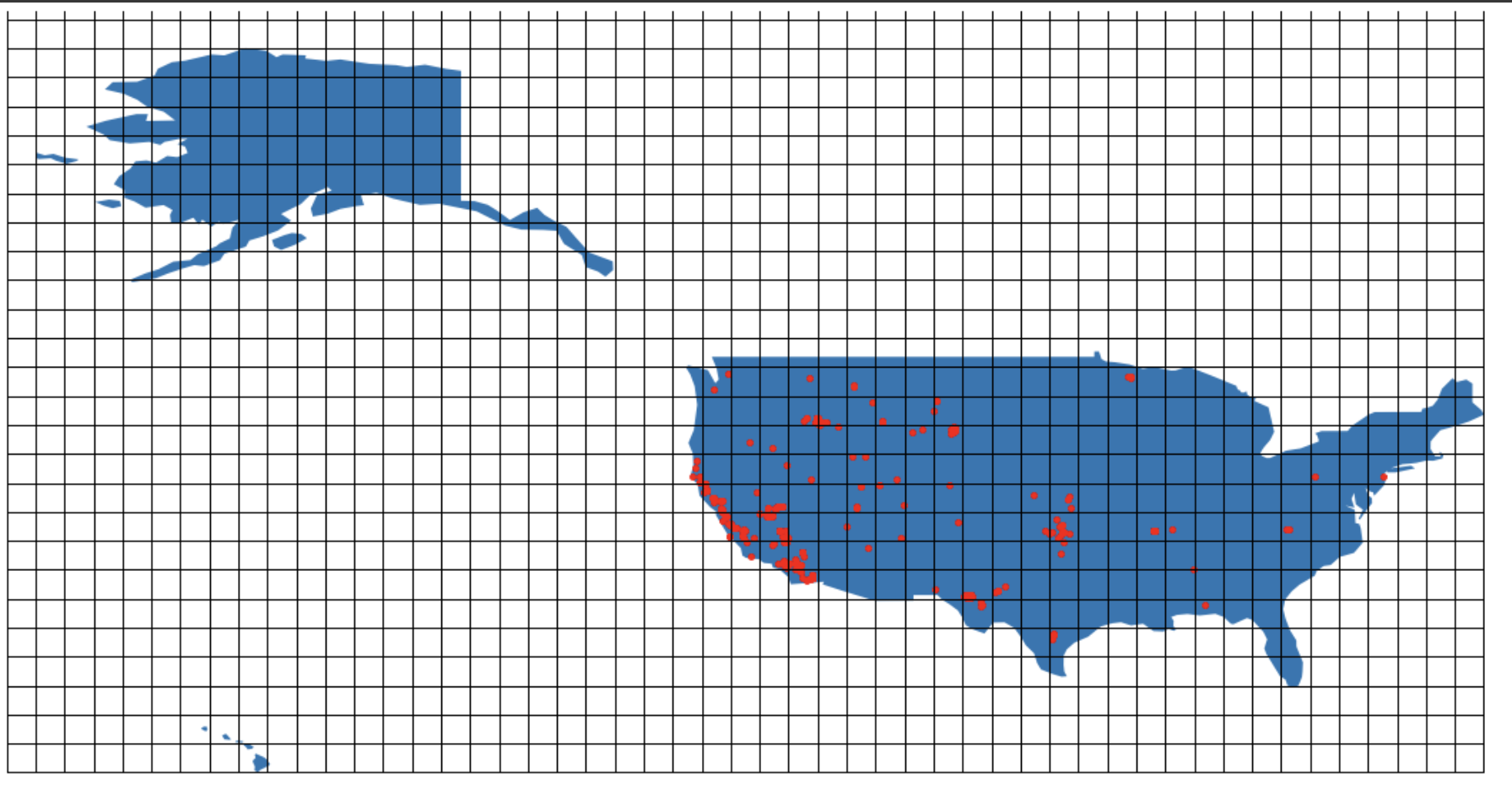
* The simplified explanation of how we plotted out data is that we first read in all of the CSVs and convert them into dataframes, then using geopandas we plot the x and y coordinates located in the data frames onto a map of the US. We do this process for each and every CSV file.
* Now that for each month we have a map with all the coordinates plotted onto it, for each map we overlaid a grid (for D1 a cell size of 50, and for D2 a cell size of 25). We know the coordinates of where these edges of each cell on the grid is, as well as the coordinates of each data point on the map.
* From here to count how many data points are within each cell, we loop through the map comparing the edges of the cells with the x/y coordinates and have a counter which keeps track of how many points are within each cell. We do this process across all 12 maps.
* At this point we now know for each of the 12 maps how many data points are located in each cell, using this information we can now move onto determining which cells are D1 hotspots and which cells are D2 hotspots.

### **D1 threshold selection:**

* We selected 60 to be the minimum value for how many points must be in a single cell on the grid in order for that cell to be considered a D1 hotspot. We went for this because we believed that each of the 12 maps should have at least a single D1 hotspot, at first we used values such as 120 and 100, but this left us with only 1 or 2 maps which fulfilled that criteria. After some consideration we landed on 60 as it was low enough to have at least a single D1 hotspot on each map, but not more than 2.



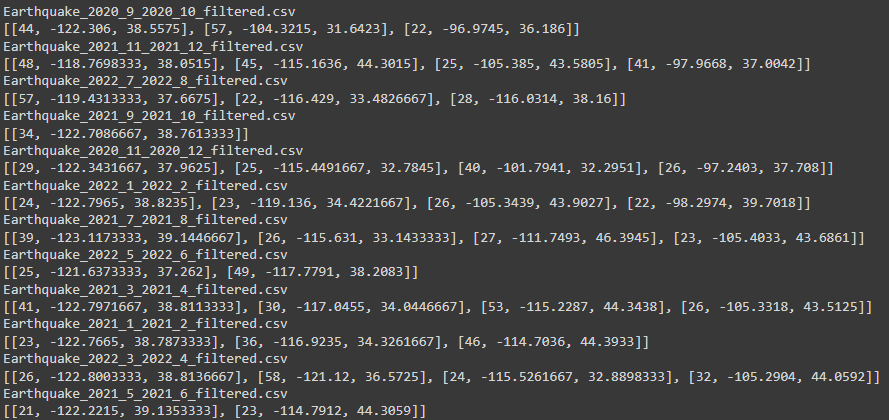
* Each row on the array above represents each of the 12 maps, and within each of the 12 maps we got the coordinates of where the D1 points were (first column shows the number of earthquakes in a cell, and second column were the longitudes and latitudes where D1 located. Using these coordinates we plotted them for each month on a map in order to get our final D1 hotspot locations.
* We chose a cell size of 50 because it created small cells on the map (as portrayed on the map), we did this because D1 points are “small and high density”, so in order to be as precise as possible we chose the smallest reasonable cell size. On the other hand for D2 points, since they are larger and more regional we chose a larger cell size so it is not as precise.



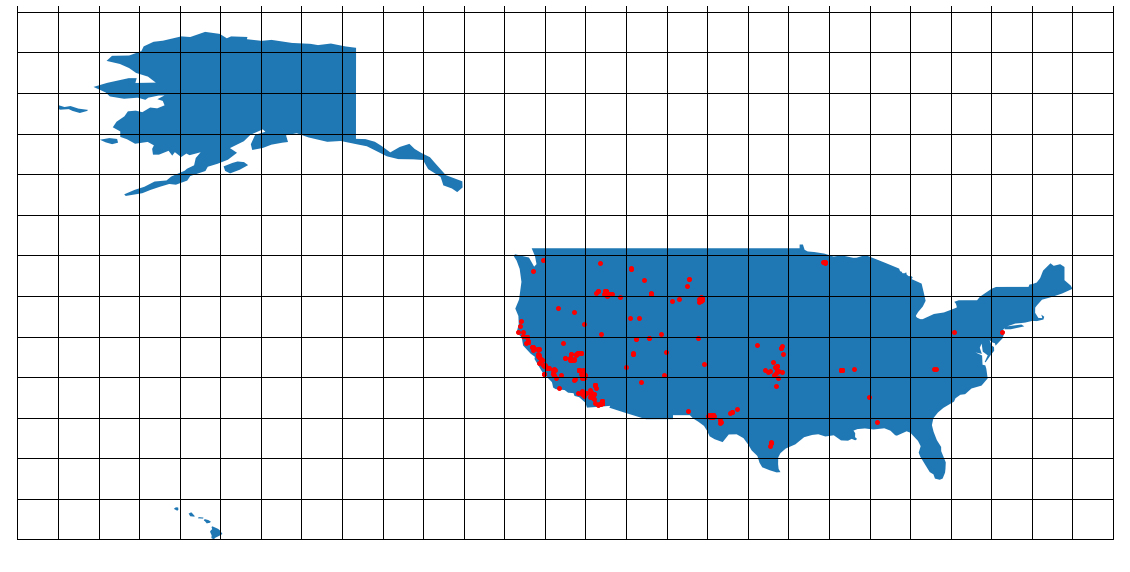
* Above we can see the map where Earthquake data has been plotted, in this graph since we are looking for D1 points we made the grid size 50 so it is much smaller (compared to 25 which we did for D2 points). The reason we chose a smaller size for D1 is because we are only looking for very dense areas so by decreasing the size of each cell we only will pick up on very dense cells.

### **D2 threshold selection:**

* We selected a smaller minimum, where in D1 we only considered cells which had a density of at least 60, for D2 we considered cells that had densities of between 20 and 60 for the POIMAGIC to consider it a hotspot. We also increased the grid size in order to allow D2 to cover more area and require less density.



* The above picture shows all the individual locations of each hotspot for each month. We can observe when compared to the D1 points, there are more hotspot locations. For D1 there was usually 1 hotspot, and sometimes 2. In the case of D2 points, each map has about 3 - 4 hotspots, and occasionally 2. This makes sense as D2 points require much less density of earthquakes, as well as look in larger cells (more regional/widespread).

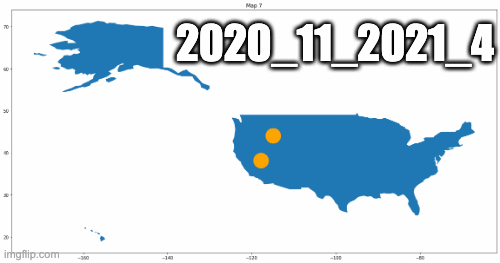
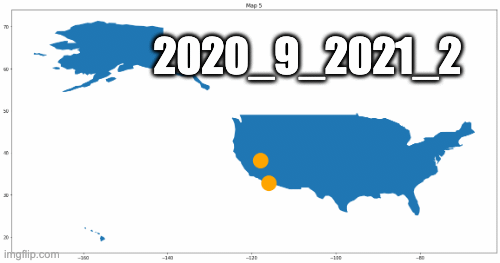


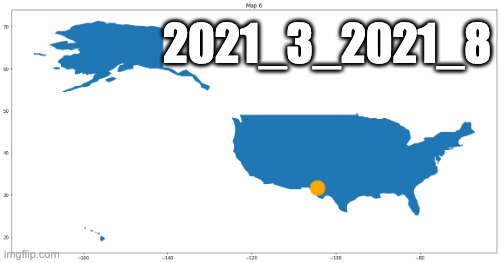
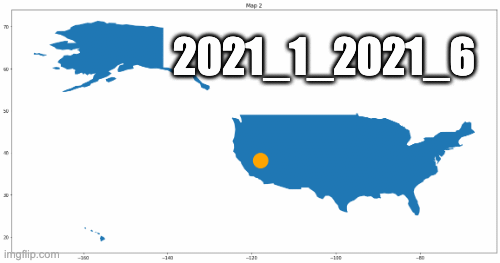
* This map above depicts all the individual locations of the earthquakes, and in this case the cell size is 25 which is much larger than what we had for the D1 points which were a cell size of 50 as we look for the regional hotspots.

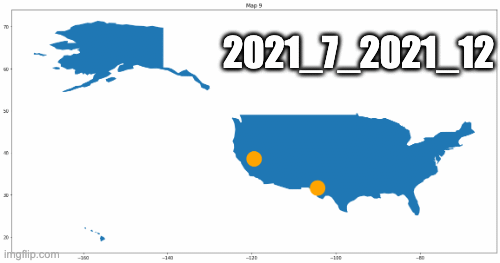
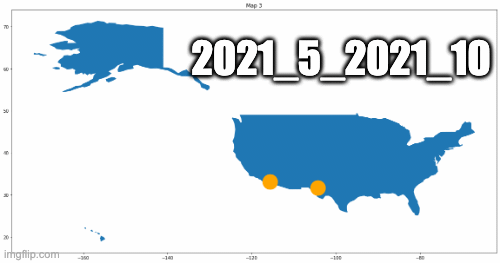
### **D1 Maps / Hotspots Gifs:**

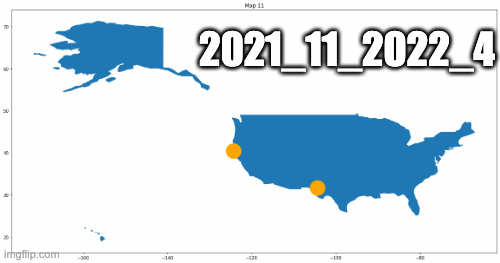
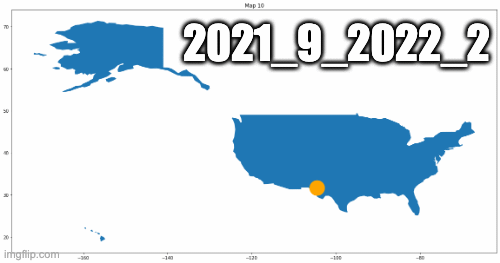
( If this is viewed in pdf format, the GIFs are not working. You can check this link attached to see how hot spots changed over time. <https://docs.google.com/document/d/1P4KsecWikLW1qB3zly-kxuHFvyjL9ncCaliPfOgH6yM/edit#heading=h.uyrgzu6jzize> )

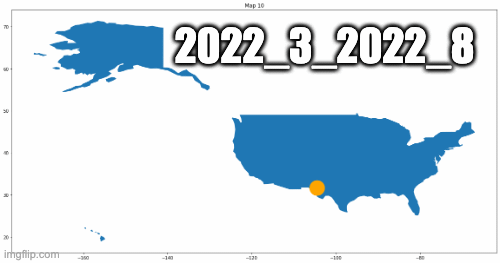
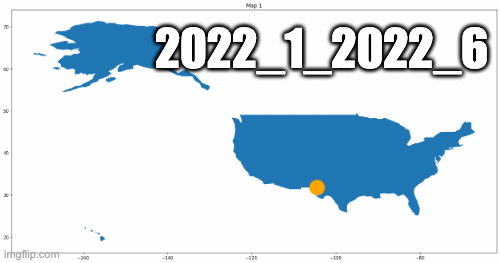
Since the earthquake dataset is subdivided into 12 batches for time intervals September 2020, October 2020,…,August 2022 and, assuming a window size of 3 batches, the system creates hotspots/warnings for the citizen who live inside the hotspots for batch1+2+3, batch 2+3+4, …,batch 10+11+12. Each GIF below determines each 3 months and iterates until August 2022.







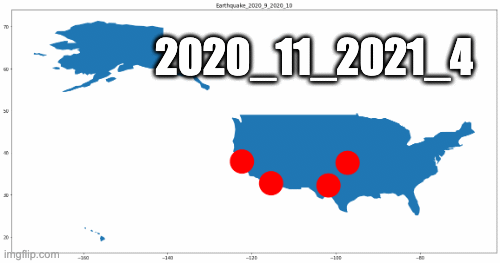
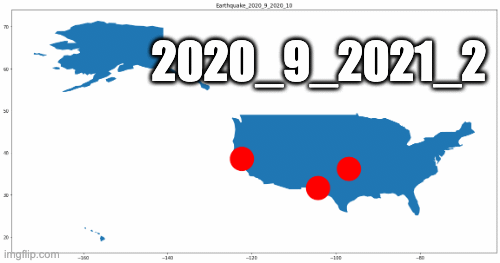


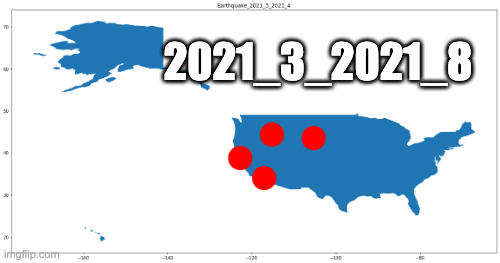
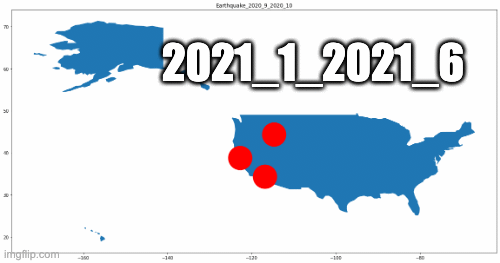


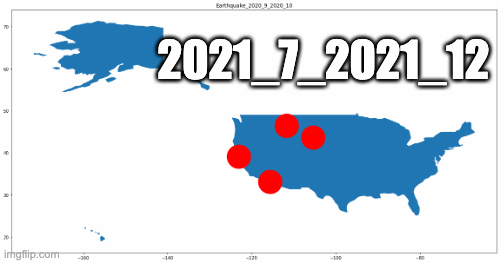
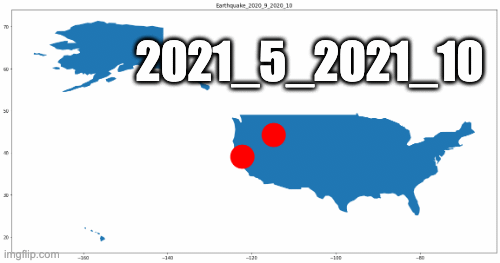
Looking at our plots made with D1 density mapping, it appears the densest area with visualized earthquakes is mostly located in central to southern California, as well as west Texas. Occasionally, during the fall season (August to December), we observed D1 hotspots in the north west area around Idaho and northern California. During the 2022 year, we only observed D1 spots purely in west Texas.

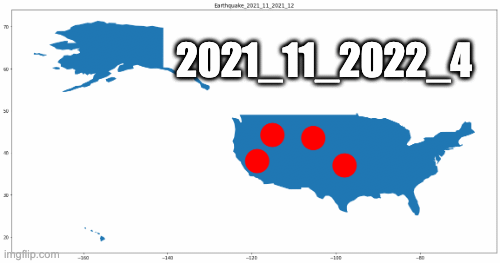
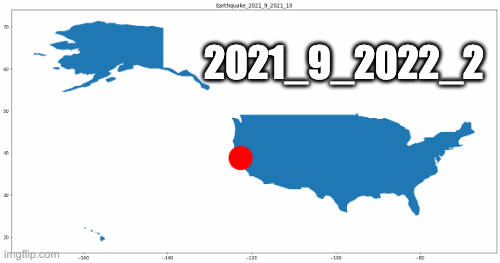
**D2 Maps / Hotspots Gifs:**

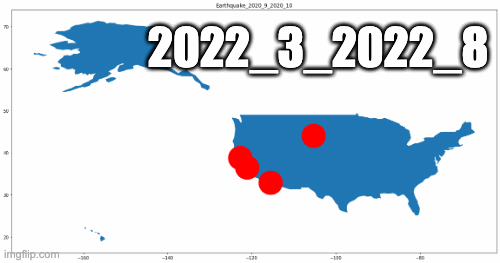
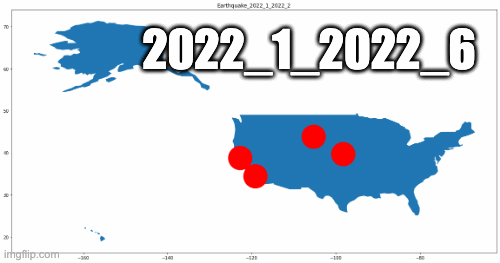
Likewise, all D2 hotspots Gifs below work the same as D1’s. If we look closely, there are more regional hotspots with less density than small and very hot spots over a year.







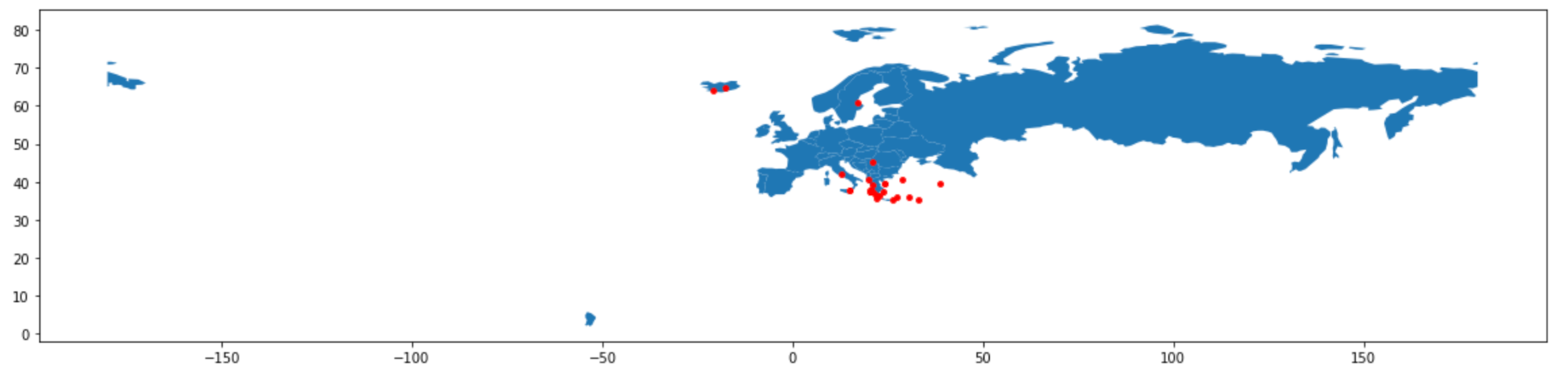


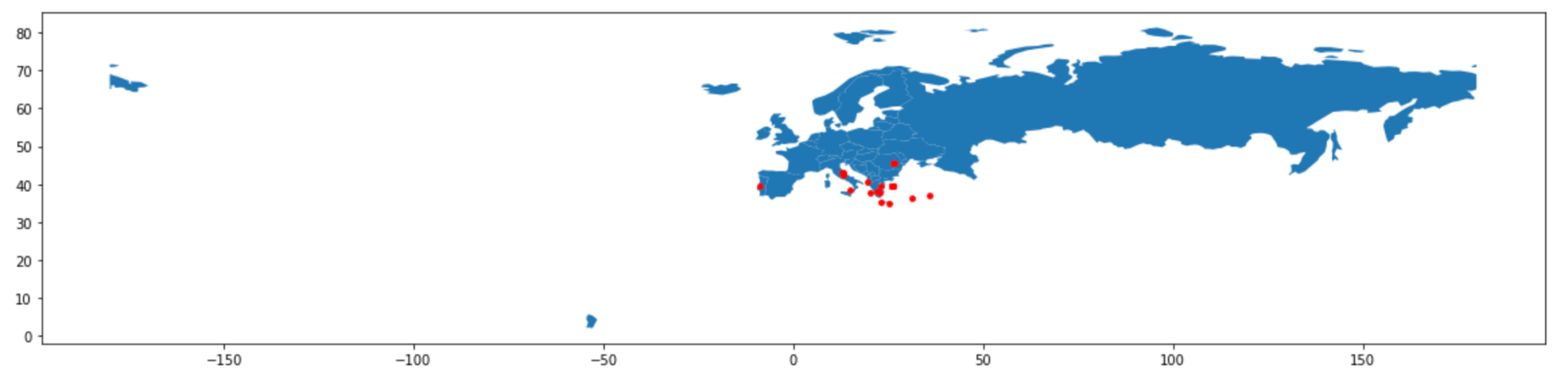


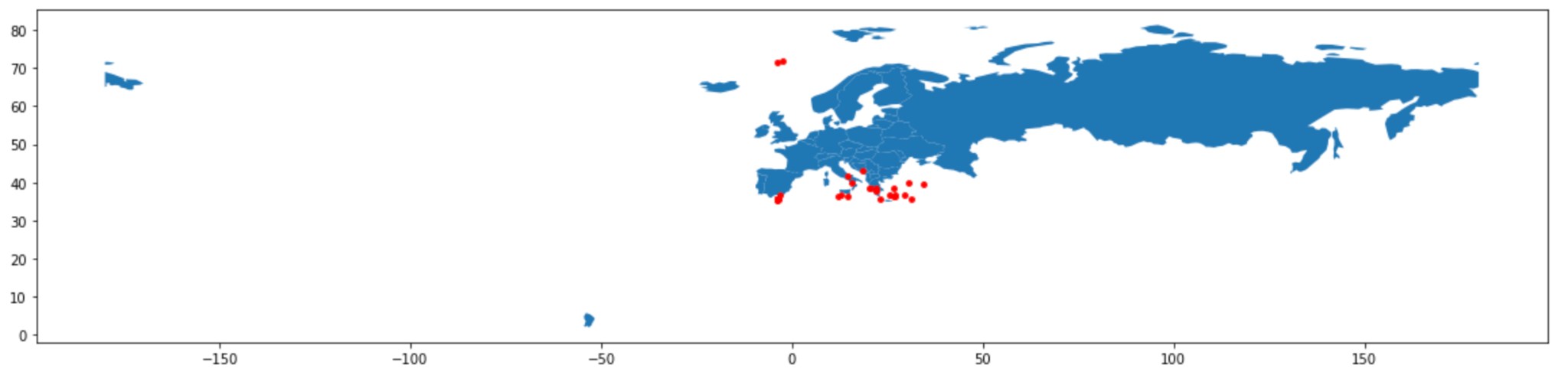
Looking at our plot maps made with the D2 density mapping, we observed more hotspots happening in the central United States, whereas in our D1 density mapping, there was very infrequent mapping in the central United States. There are only 2 occurrences of D2 hotspots in west Texas, this could mean that west Texas usually receives small dense hotspots rather than large regional hotspots. Similar to the D1 maps, California exhibits multiple D2 hotspots.

### **Europe Data (for testing purpose)**

We did not count the grid cells to find D1 and D2 hotspots because Europe Dataset is used only for testing purposes. We estimated the hotspots based on 3 maps below as representatives. All maps dotted with hotspots of Europe Dataset between 2016-2018 were generated in TeamOGP2022.ipynb.







Above are 3 plots extracted from the Europe Dataset that represent the pattern or regions that most earthquakes are usually formed from 2016 - 2018. Most of the earthquakes were centered in southern Europe near Italy and Greece. We also noticed that around Spain/Portugal as well as in the north around Iceland they also received a good amount of Earthquakes, but not as much as southern Europe. For the POIMAGIC tool we have developed, it was easy to transition between different datasets to plot the earthquakes on the geopandas map. However, certain adjustments must be made in the code to take into account values for D1 and D2 along with the cell grid size parameters.

**Conclusion**

Our thought process when creating our POIMAGIC detection system is we took a look at previous earthquake data as our training data and develop thresholds that can detect certain spots that has frequent earthquakes. Our high density areas, labeled D1, can detects a large concentration of earthquakes happening in a specific area. Our low density area, labeled D2, can detect earthquakes happening in a wider area. As for Europe dataset using the existing system above, we observe that hotspots occur mostly in the southern part of Europe but no significant hotspots can be seen in eastern Europe.

### 

### References:

* “Mapping and Plotting Tools#.” *Mapping and Plotting Tools - GeoPandas 0.12.1+0.g195f70b.Dirty Documentation*, https://geopandas.org/en/stable/docs/user\_guide/mapping.html.
* *Fast and easy gridding of point data with geopandas* (2020) *James Brennan*. Available at: https://james-brennan.github.io/posts/fast\_gridding\_geopandas/.
* *Plotting geospatial data using GeoPandas* (2020) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/plotting-geospatial-data-using-geopandas/ (Accessed: November 22, 2022).