Staying Connected: Analyzing Earthquakes through Twitter

Andrea Capshaw

Masters of Science in Geographic Information Science (MSGISci)

Department of Geography, California State University, Long Beach

Introduction

In today's world, social media has become the starting and ending point in a person's life, whether it be a way to document their everyday activities, or, a way to keep up with the world around them. Social media has become an influential tool in keeping people connected, while at the same time allowing a wide range of information to be shared and stored, which can be a very powerful tool. With so many people using social media, it now has the power to aid in events that could be a matter of life or death.

With the use of Twitter, this project has collected and mapped earthquake related tweets that have been generated by the public. In the event of a devastating earthquake, the goal is to provide first responders and emergency management personnel with quick and reliable on the ground data.

Figure 1. *Study area map. Project will generate data from the state of California and Hawaii.*

Data and Data Sources

All point data used in this project was collected manually in Twitter by using a Python script. All attribute data was then stored into a CSV file. Attributes that were collected from tweets include: coordinates, tweet itself, date, and any media files associated with them. Shapefiles for California and Hawaii provided by the United States Census Bureau.

Table 1. List of data and data sources used in the project

Dataset	Source
Point Data	Twitter
California Shapefile	United States Census Bureau
Hawaii Shapefile	United States Census Bureau

Methodology

Data acquisition was the core of this project, and needed to be collected manually by customizing a script in Python. Once all data was collected and stored into a CSV, it was then transformed into point shapefiles in ArcMap, symbolized for better visualization in QGIS, and exported as an interactive web map. Added functionality for the web map was produced by editing the document's HTML code. The steps below outline the methodology flow chart in more detail, excluding the process to gain access to Twitter's API. The entirety of this project required approximately six months to complete.

- ❖ Step 1(a): Python script extracts tweets carrying the hashtag 'earthquake', tailored to specified coordinates and 50km buffer.
- ❖ (b): Script stores all extracted data in a CSV file customized to pull only coordinates, tweet, media files and date.
- ❖ Step 2: Merge all CSV files into 1 for each state, include final header names.
- ❖ Step 3-5: Upload CSV's into ArcMap for creation of point shapefiles. Use QGIS to apply symbology.
- ❖ Step 6-7: Create web map with pop-up windows and the ability to geolocate users.
- ❖ Step 8: Edit HTML to add code that includes document title, navigation, heat maps, and user selectable map visibility.
- ❖ Step 9: Publish and host final interactive web map to Amazon service.
- ❖ Step 10: Fully functional, interactive web map!
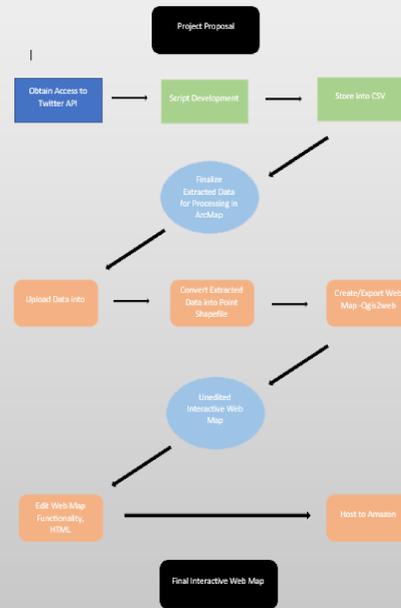
Figure 2. *Methodology flow chart for creation of interactive web map.*
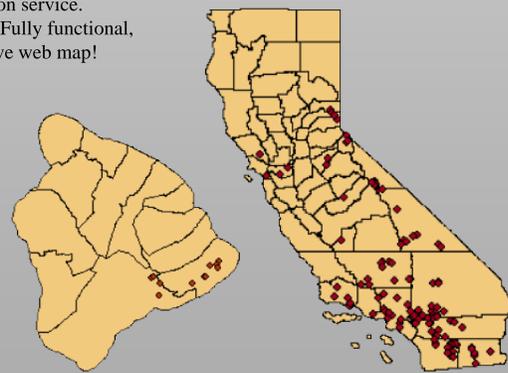
Figure 3. *Twitter data converted to points with true location parameters and placed onto map for shapefile creation.*

```
api = tweepy.API(auth)

for tweet in tweepy.Cursor(api.search, q="#earthquake OR #earthquakes',
                           lang= 'en', location='61.1077052,-150.0006914,50km',
                           include_entities = True).items():

    print tweet.created_at, tweet.text, tweet.user.id
```

Figure 4. *Python script used to extract tweets in a specified location.*

Results

The resulting map for California contained a total of 837points along with all necessary fields which included: county name, latitude, longitude, media, tweet text and the date. Due to the lower population associated with Hawaii, only 11 earthquake points were collected.
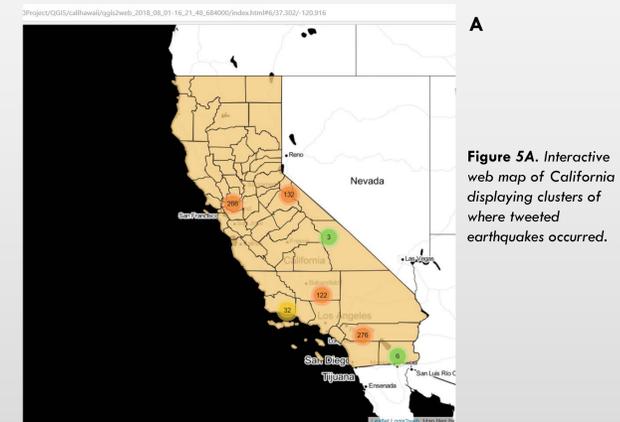

Figure 5A. *Interactive web map of California displaying clusters of where tweeted earthquakes occurred.*
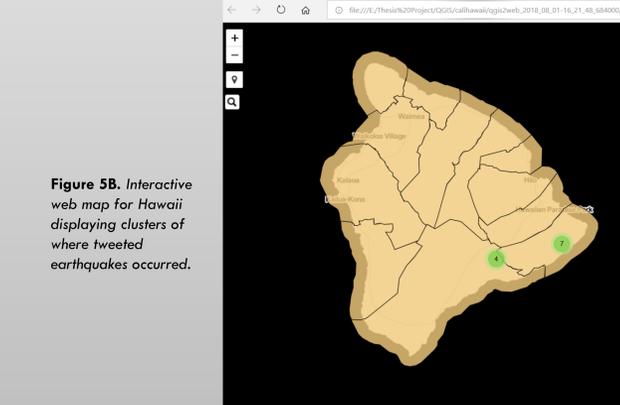
**Figure 5B.** *Interactive web map for Hawaii displaying clusters of where tweeted earthquakes occurred.*
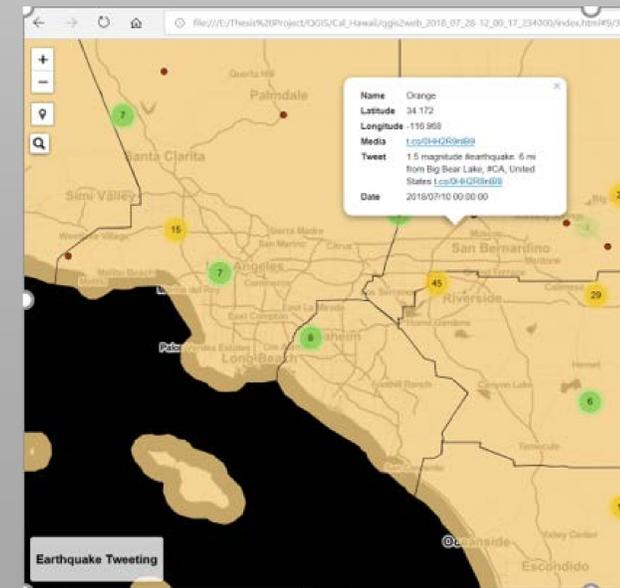
Figure 6. *Zoomed in photo displaying the pop-up window information associated with each point and added title to the bottom left corner of map.*
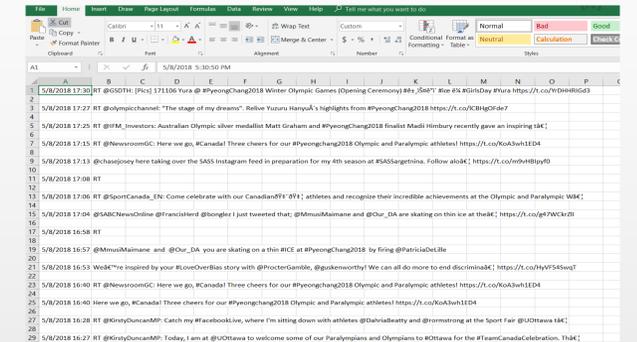

Figure 7. *Example of Twitter streaming data.*

Discussion

After all components were complete, the significance of this project relied heavily within the data acquisition stage. Since all earthquake points had to be collected via Python, producing a successful script was the most challenging part. Below are a few other areas where issues were encountered.

Python:
- ❖ Errors within code spacing and wording would not retrieve tweets specified to coordinates, tweets were being retrieved from all parts of the world.

Chosen Study Areas:
- ❖ The study areas chosen proved to be a limitation to the amount of data that would be retrieved. Hawaii resulted in only 11 extracted points, this is due to their population not being as great as other regions.

Twitter API:
- ❖ Twitter only allowed for the collection of tweets within a 10 day span, nothing further, resulting in fewer points in the final map.
- ❖ Twitter only permits a small fraction of tweets to be retrieved by the public, resulting in less data that can be extracted

QGIS:
- ❖ Creation of the interactive web map was very basic, further functionality had to be established through the HTML document.

Conclusion

The final result establishes a great way to visualize where earthquakes have occurred and how well the public responds to them via Twitter. For the purpose of this project, it was concluded that although very interesting, relying on tweet data alone to respond to these events is not the suggested method due to the limitations in data collection. For future work, this project could gain more reliability by using Twitter's premium API as opposed to the standard. Creating a continuous script that would retrieve and update earthquake points daily is also suggested for future work. With the implementation of both suggestions, the final deliverable could be a great tool in responding to disaster type events in California.