CP-6542 Transportation and GIS

Final Project[1] (December 12, 2017)

Team member Spandana Ananda

Tamanna Goware

Duoduou Lin

Evan Walker

Alyas Widita

Examining the extent of Hurricane Harvey through social media-generated data, spatial assessment, and network analysis

Introduction

The main objective of this paper is to assess the extent of Hurricane Harvey and identify the appropriate aid center locations to help expedite the recovery process. In doing so, we employed a two-part analysis. The first part consists of assessing the damaged areas using 1) Social media-generated data, i.e., Twitter, and 2) spatial analysis based on information provided by Federal Emergency Management Agency (FEMA). The second part uses the information gathered in the first part to identify the appropriate locations for aid center using network analysis. That being said, each subsection of this paper will cover the methods, results, and discussion of the analyses. The overarching summary discussing the lessons learned and the recommendations is provided at the end of the paper.

Assessing Natural Disaster through Social Media-generated Data

One aspect that we wanted to look at was the impact of Hurricane Harvey on social media. To accomplish this objective, we utilized the Twitter API platform and R-Studio to access tweets that mentioned Hurricane Harvey and where these tweets were located. These were then plotted on a world map and the results were compared with more traditional forms of data regarding the hurricane. Table 1 describes the workflow.

---

[1] Pages : 10
Tables : 2
Figures : 6

Table 1. Workflow performed to obtain Hurricane Harvey-related tweets

Step	**Objective**	**Methods**
1st	**Collect Twitter data**	● Create app in Twitter developer site ● Use R-Studio and activated 'twitteR' package ● REST API
2nd	**Organize data**	● Convert to R data frame ● Convert to CSV ● Select results that contain latitude and longitude attributes
3rd	**Plot the results**	● Use 'ggplot2' package ● Download world map for reference ● Adjust attributes of plot and title

Method and Results

The first step in this process was to create a Twitter Application from the Twitter Development website. Using Evan Walker's personal Twitter account, we set up the application and named it "Hurricane Harvey Reaction". Once we set the app up we were able to acquire an API Key and Access Token. These were needed for the next step in R-Studio.

R-Studio is an open-source and very useful program that can be utilized for a variety of different analyses and data visualizations. R-Studio uses a variety of packages to execute the desired outcome. This process was done in three steps. The first in the process was to collect the tweets from the Twitter API. To accomplish this we used the twitteR package. The twitteR package was used because for this project we used what was called a "REST API", this means that a one time collection of tweets was done based on the specified parameters. There are a variety of parameters that can be entered when gathering Twitter data. For this project we wanted to focus on Hurricane Harvey and its effect on Texas. The initial twitter search was done for mentions of:, "#hurricane", "#hurricaneharvey", and "#harvey". They were then filtered by their geographic location. The geographic parameters were a 200 mile radius around the coordinates of the center of Houston. Then a filter for the dates was applied between August 17th and September 3rd, 2017. There were 10,000 tweets requested and they were all in English. Unfortunately this yielded zero tweets that fit the specified parameters. The second search did not include the specified date range and geographic range either the number of tweets requested was also increased to 15,000. This resulted in 15,000 tweets and led to the next step in the process. Table 2 describes the two searches we performed to obtain the relevant tweets based on a given criteria.

Table 2. Tweets obtained from two iterations with different search criteria

Stage	Criteria	Output	Geocoded
1st Search	● Date: August 17th - September 3rd, 2017 ● Location: 200 mile radius from the center of Houston	10,000	-
2nd Search	● Date: Not specified ● Location: Not specified	15,000	30 (0.2%)

The next step was to organize the output. The first part of this process was to convert the output into a dataframe and then furthermore converted into a CSV file. A CSV file can then be opened in Microsoft Excel. Upon inspection of the CSV file there were 15,000 items with 16 fields, so a very large dataset, two of the fields for each item were a latitude and longitude field. Unfortunately the vast majority of tweets are not geolocated. After the items with null values were filtered out only 30 items remained that were geocoded. The geocoded items were saved into a new CSV file and then loaded in R-Studio for the third step of the process.

The third and final step of the process was to plot the points on a map of the world. R-Studio is very convenient for plotting graphics, in this case the X and Y axes were longitude and latitude. So for mapping in order to gain context of where the points were located we needed to upload a background shapefile that included a map of the world. When the points were plotted on the world map it showed where the points were located in the world. Figure 1 illustrates the result.

As the image shows, of the 30 points plotted the majority of them were in the Western Hemisphere with a concentration in the United States. The results were further categorized to show only tweets located in Texas. There were 4, all of them were located close to the coast where Hurricane Harvey initially struck.

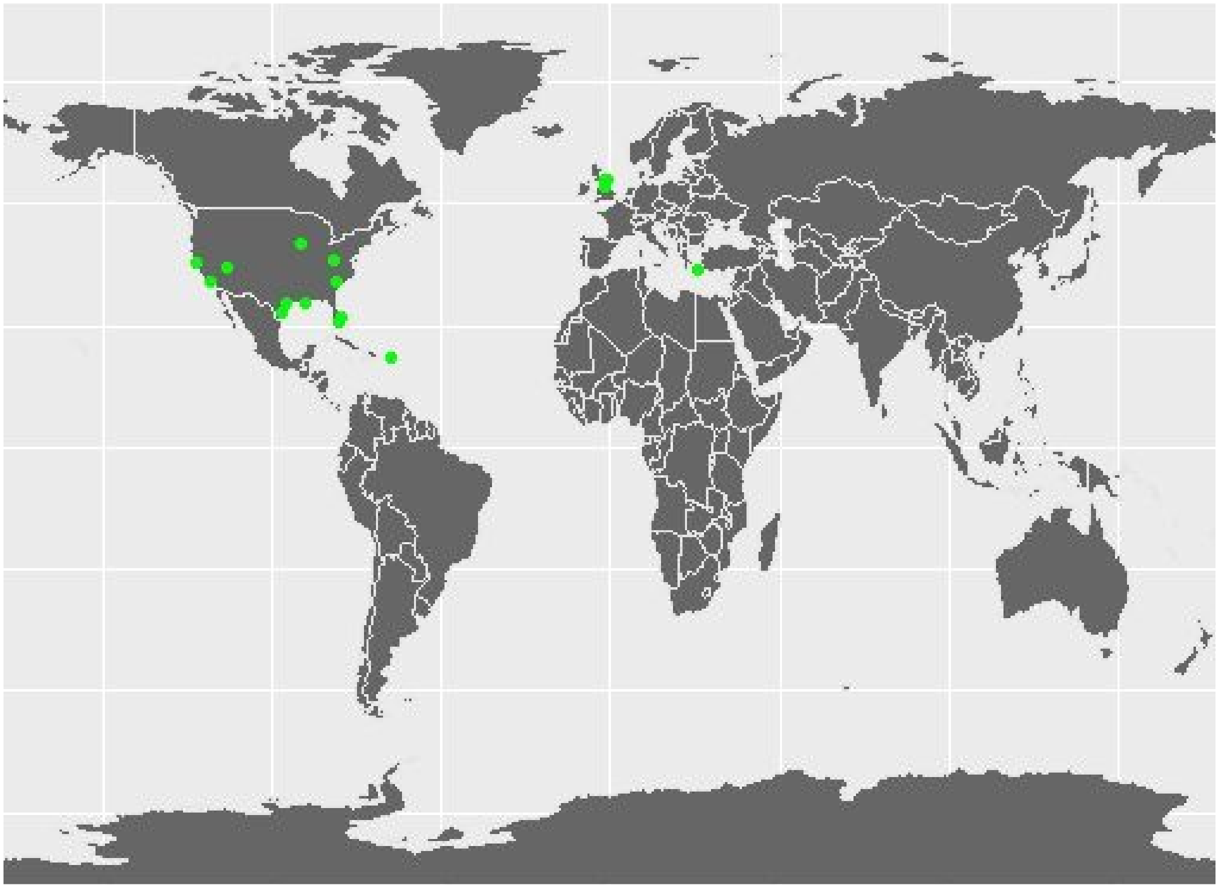

Figure 1. Hurricane Harvey tweet locations

Discussion

Prior to this study we wondered whether or not twitter data could be useful in a natural disaster like Harvey. Would Twitter be viable for assessing the impact of natural disasters and be helpful in a transportation context? In the immediate aftermath of a natural disaster this might be a useful tool. Social media is always changing and hot discussion topics change very quickly. Therefore during the disaster or its immediate aftermath the data one could gather may be very useful in the relief effort. However over time disasters fall out of the public eye and the density of mentions in social media, and more specifically Twitter will diminish. Since this data was pulled almost four months after the disaster, there were not very many mentions of the disaster. Furthermore there were even less tweets that were geocoded, 30 out of 15,000 is 0.2% of tweets. Were this study to be repeated one might need to do more background research to identify keywords that would get better results. In conclusion Twitter data is probably not the best way to assess damage of a natural disaster. However it could definitely be a useful tool in other areas that may not require a geographic location.

Identification of Damaged Areas using FEMA data

In addition to the non-traditional assessment method using social media data like what we described in the previous section, we also conducted a traditional assessment effort to identify the damaged areas throughout Texas. While the hurricane did affect most of the state, a few areas saw a higher magnitude of damage. We use the damage data provided by FEMA and the population count derived from the Census Bureau's website to come up with a damage index.

Method and Results

The damage designation dataset obtained from FEMA formed the basis of the analysis. FEMA categorizes the counties affected by level of damage, with IAPAA-G being the highest. Houston falls within this zone. The designations given to the counties in Texas is as shown in Figure 2.

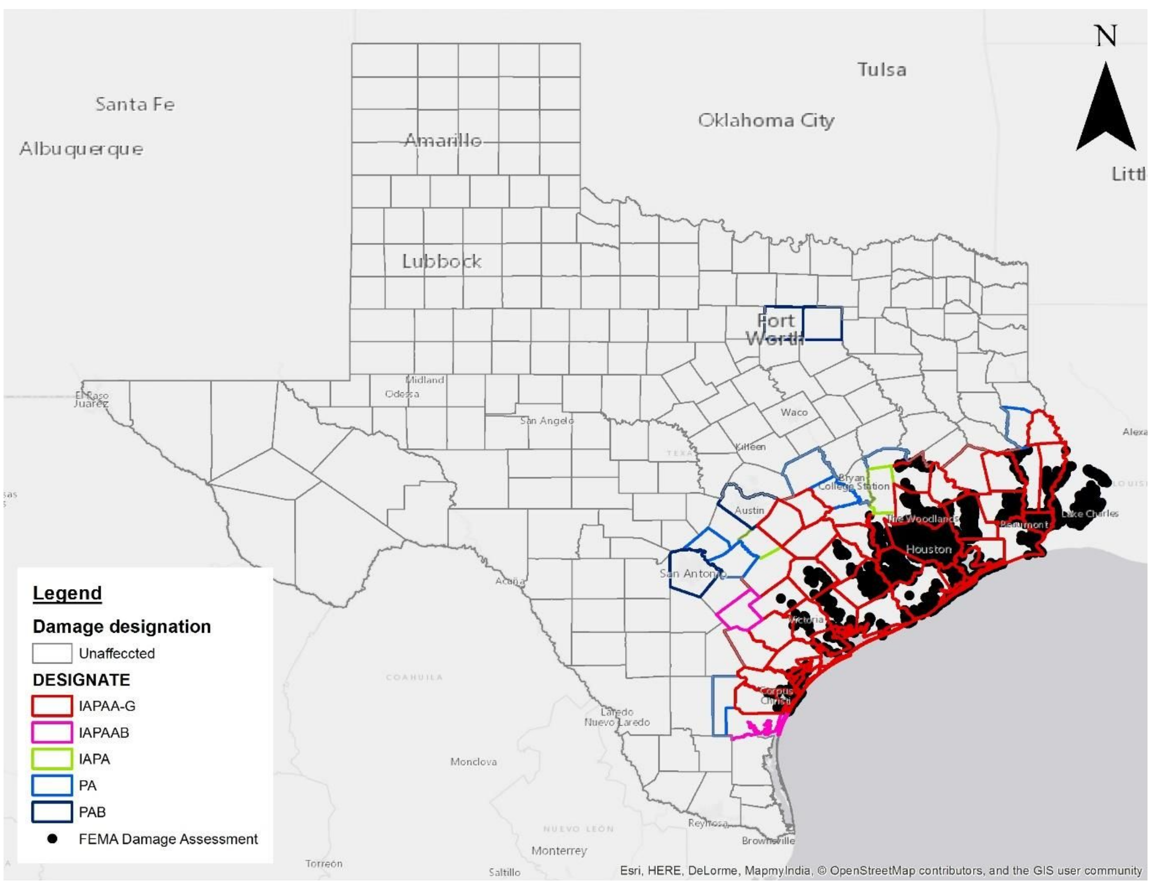

Figure 2. FEMA designated damaged areas and its corresponding damage level

The coastal areas were the ones most badly affected by the hurricane as we can see from the image above. The black dots represent the incidents of damage identified by FEMA. We weigh the number of incidents reported per census tract with the population density data provided by the Census Bureau to arrive at our final damage index, also calculated per census tract, as represented in Figure 3.

According to the damage index, Houston was the most affected by hurricane, but some of the other areas along the coast, such as Beaumont and Port Arthur, also show high damage indices. San Antonio and Austin also see a moderately high index due to their population, but they're not as badly affected as they lie in the low damage designated zones (PA, PAB). A close up of some of the areas that have been badly affected has been shown in Figure 3.

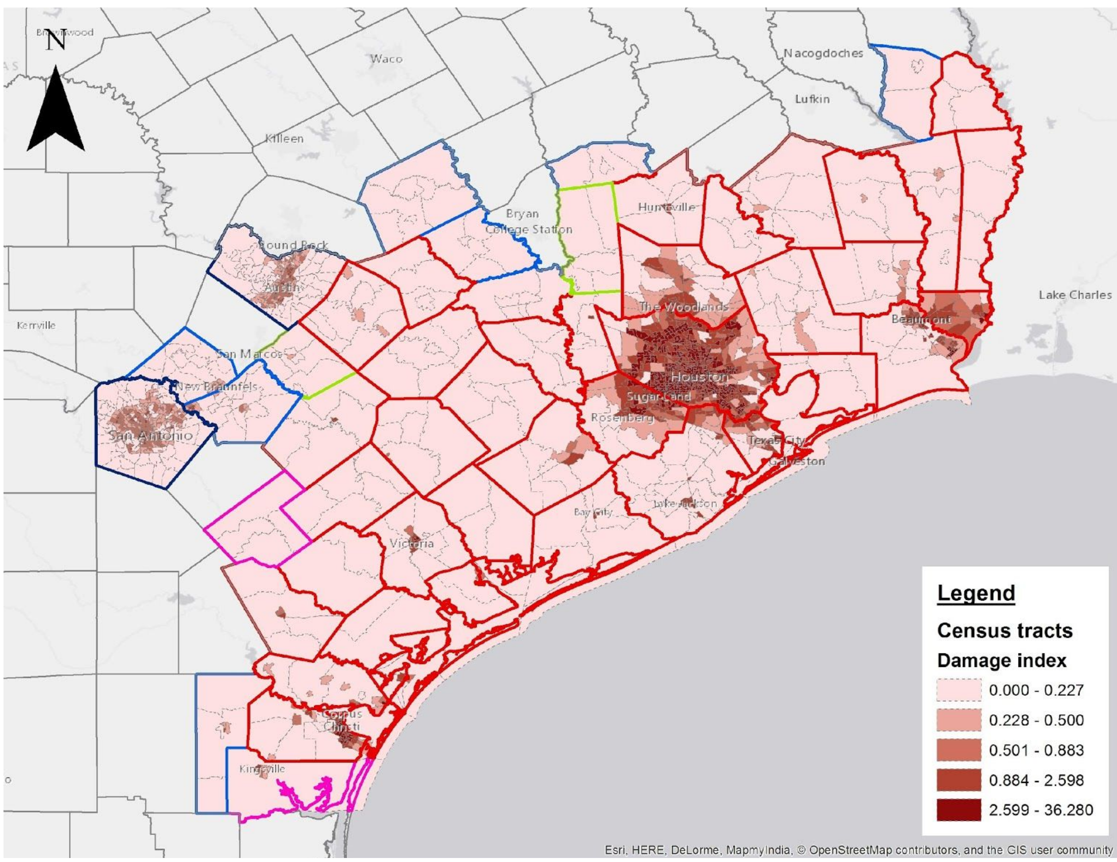

Figure 3. An overview of damaged areas in the southeastern region of Texas

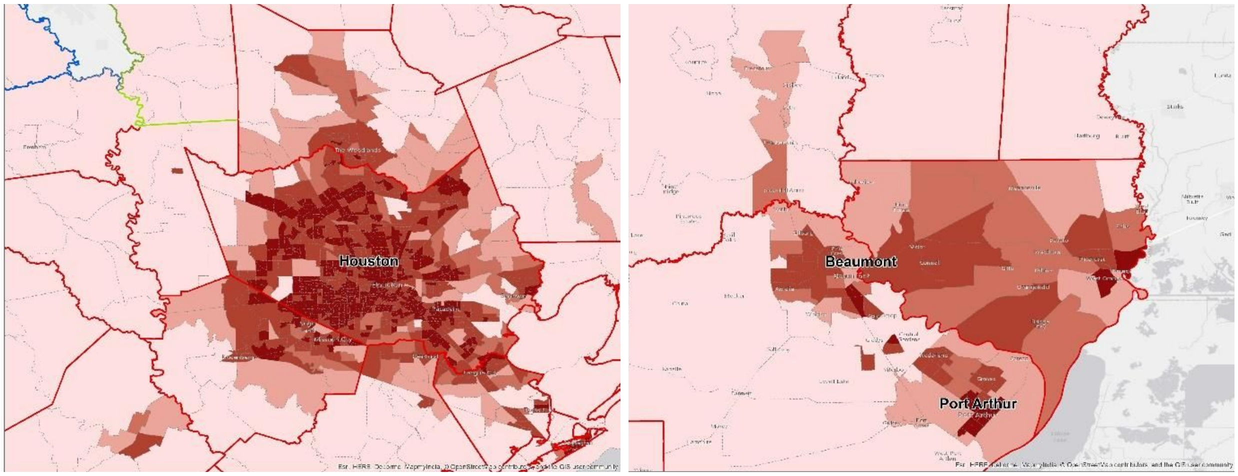

Figure 4. Examples of considerably small communities along the coastal areas that were heavily damaged by the hurricane

The city of Houston was the worst hit in the Hurricane due its high population density. The figure below shows the distribution of the damage index in the areas surrounding the city, with the black spots representing the FEMA damage assessment points, and the blue line showing the city limits.

Since we've calculated the damage based on both population density and points of incidence, the census tracts with the highest amount of damage are distributed all over the region. A few census tracts that lie just outside the city limits have also been badly affected. The population density does seem to have outweighed the incident points in the final calculation as evidenced by the low damage index in the areas surrounding Houston.

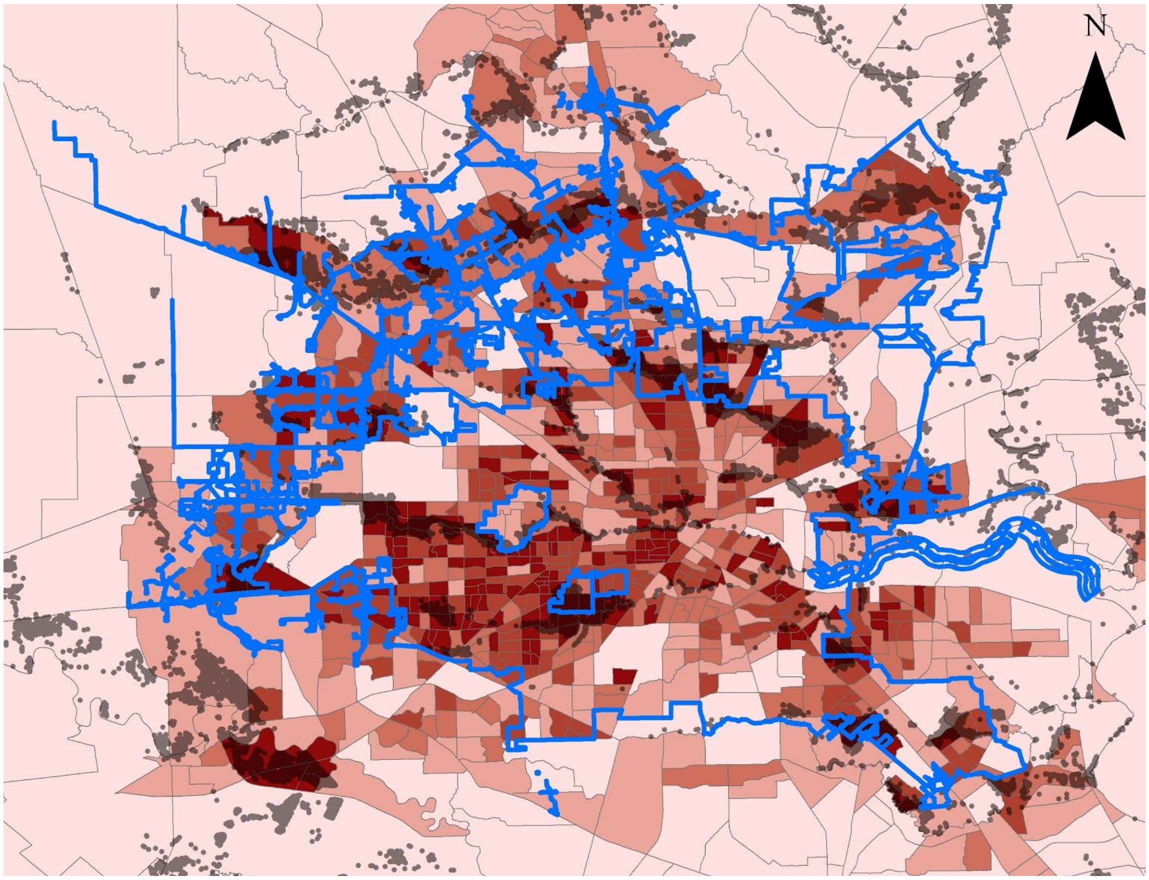

Figure 5. High-damaged areas scattered throughout the region

Discussion

The FEMA dataset has enabled us to produce a reasonably fine-grained index at census tracts level to illustrate the spatial distribution of the damage in light of Hurricane Harvey. Utilizing this spatial index, in the following section we will describe how we developed a location-allocation model as the basis to locate 10 chosen aid centers out of 20 candidates.

Identifying Appropriate Aid Center Locations using Network Analysis

Method and Results

The first step in network analysis was to create a network dataset of Houston in ArcCatalog. The network dataset was created based on the Texas Department of Transportation Roadway Inventory.

Using this network dataset the 20 aid center location candidates were located in areas with high damage index. A buffer of 2 miles was added to each candidate aid center locations, screening census tracts centroids within accessible distance of the candidate aid centers. To choose the best aid center locations that can capture the greatest demand in the minimum distance, a location-allocation model was built to allocate demand points (census tracts centroids) to their nearest aid center based on their location.

Discussion

The results as presented in Figure 6 illustrates the chosen aid center locations. As can be seen, most of these chosen aid centers are evenly located across the region. The southwestern parts of the region were among the areas that had highest damage index; therefore, it is understandable that a considerable number of aid centers were most appropriate to be located in that area. The model, however, seems to fail to consider the outlying, northern part of the region as none of the chosen aid centers are located there. This might be attributed to the less damage these particular areas suffered from; yet, it doesn't mean that aid center should not be located there.

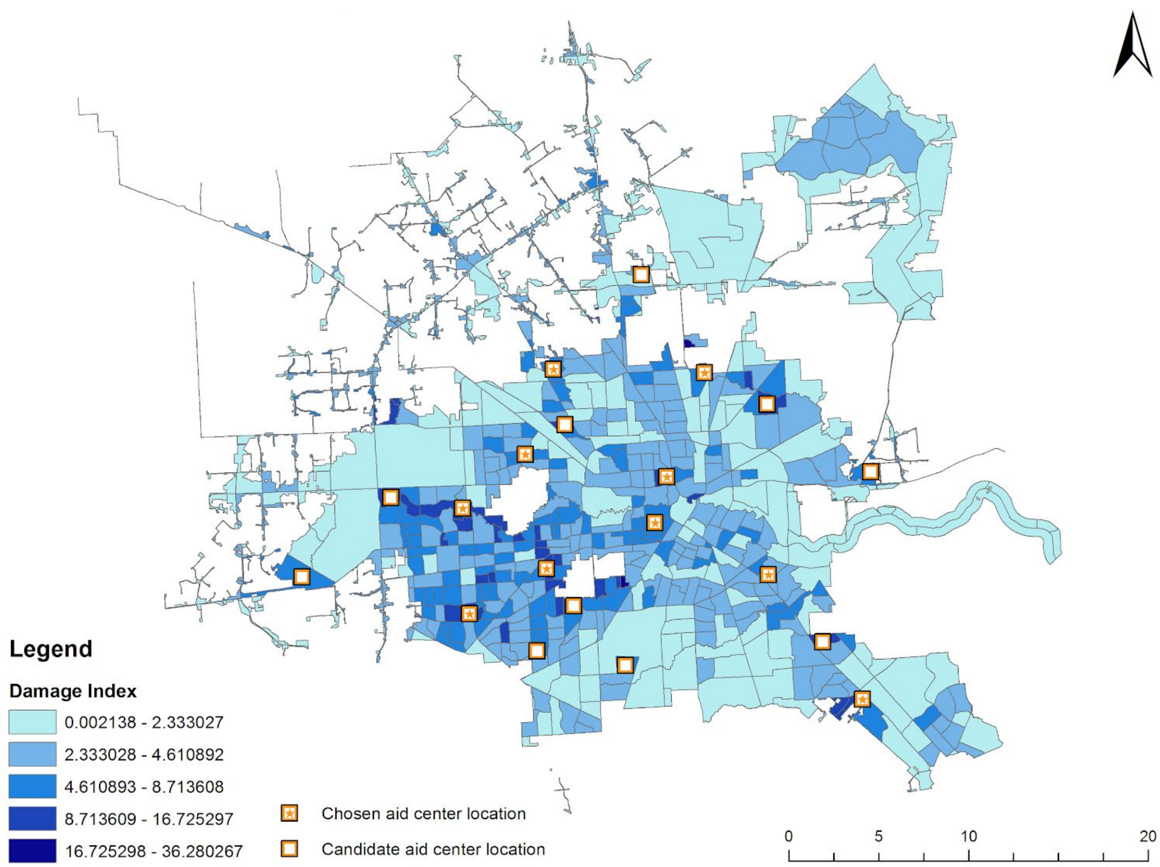

Figure 6. Chosen aid center locations spread somewhat evenly throughout the region

Summary and concluding remarks

In this paper, we examine the extent of Hurricane Harvey using social media-generated data, i.e., Twitter, spatial assessment, and network analysis. This series of analyses leads to our recommendation of where to locate aid centers in light of the devastating storm. The analyses as conducted provide a valuable lessons-learned regarding the strengths and weaknesses of each approach. First, non-traditional data sources like Twitter has the potential to provide a user-based and subjective assessment of the damages during adverse events like the Hurricane Harvey. However, such analysis remains susceptible as the information provided relies entirely on the users. As our analyses indicate, only 0.2% of the tweets containing the words we are interested to were geocoded, significantly limit our objective to assess the spatial distribution of the damages based on tweets.

Second, given that we couldn't rely on users-generated data, the spatial assessment using FEMA dataset might as well remain the most appropriate approach yet. FEMA dataset provides an objectively measured damages assessment as the staple of our analysis on the census-tract damage index across Houston, TX.

Third, utilizing the census tract-based damage index, we were able to develop a location-allocation model and provided a recommendation of the 10 chosen aid centers out of 20 possible locations. While the results of the model could provide insightful suggestions, policymakers need to approach these suggested chosen locations with caution. For instance, the most visible issue we encountered is the fact that the model appears to ignore a certain part of the region entirely, solely because that considerably large area wasn't severely damaged relative to other portions of the region.

Data source

Twitter

United States Census Bureau

Texas Department of Transportation

Federal Emergency Management Agency

Texas Natural Resources Information System

National Aeronautics and Space Administration