# Identifying Vulnerable Healthcare Facilities Due to the Effects of Precipitation

Pronoy Pant, Karan Bhanot, Varad Pahak, Devanshoo Jain, Chirag Sahni, Lauren Roy, Pritesh Maheshwari, Vandita Patidar, Diyanko Bhowmik, Kayla Baysuik, Meixin Liang

Thilanka Munasinghe (munast@rpi.edu)

Rensselaer Polytechnic Institute, Information Technology & Web Science (ITWS), Troy, NY, United States.

## Abstract and Problem Area

Climate change can have a prominent effect on the health care facilities vulnerability to get negatively impacted. Environmental factors such as extreme precipitation pose increased flood risk, thereby increasing the vulnerability of health care facilities in these areas. In this poster, we explore the **effect of extreme monthly averaged precipitation on the risk of healthcare facilities from the years 2010 to 2020** by leveraging data from 2005 to 2020. We applied KMeans, DBSCAN, and Agglomerative clustering to identify the top areas of extreme vulnerability.

Healthcare facilities are the most important part of our society, as they are the backbone on which the whole system is based. As witnessed by the recent snowstorm and the subsequent power outages in Texas (February 2021), damages to these facilities can result in loss of human lives and a disruption of the community. These facilities can be most disrupted and brought to a halt by being flooded. A very recent example for this is the flooding that happened in Massachusetts where heavy rainfall triggered a flash flood emergency in the area and dozens of patients had to be transported to different hospitals because of the hospital becoming unable to function. This project aims to identify and reduce the severity of these events by assigning hospitals and other healthcare facilities with a risk factor, so that the hospitals can assess their vulnerability to the aforementioned events and take necessary precautions.

## The Data

- NASA GPM IMERG Monthly Precipitation Data
- Department of Homeland Security HIFLD hospital locations
- Department of Homeland Security HIFLD urgent care center locations
- Department of Homeland Security HIFLD pharmacy & dialysis center locations

## Exploratory Data Analysis

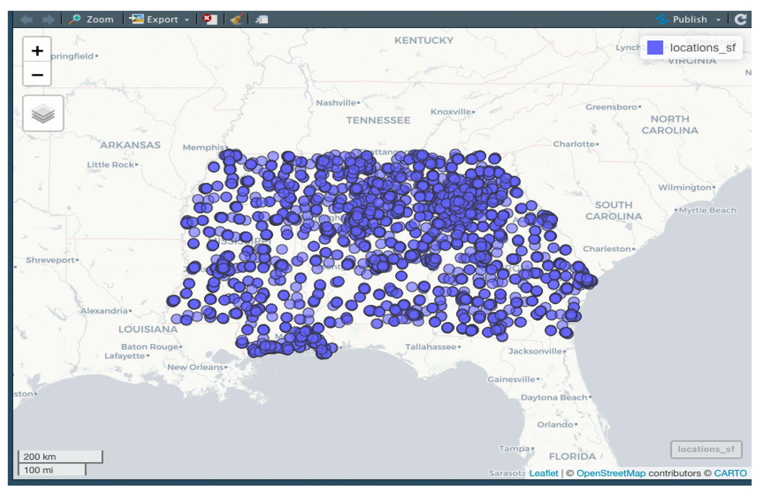

Fig 1. All The Healthcare Facilities Mapped using their coordinates

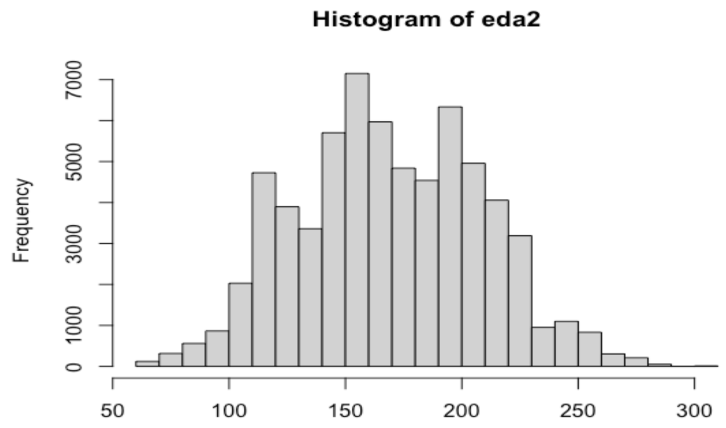


Fig 2. Histogram of Rainfall averages from days Jan-10, Dec-20 (days shown in x-axis)

In October, November, December and January, there is a high rainfall in most of the places
In February and March, there is moderate to high rainfall in most places
In April and May, there is moderate rainfall in most places
In June, July, August and September, there is moderate to low rainfall in most places

**Ackowledgement**: We would like to thank the NEIH and NASA team including Dr. John Balbus, Dr. Helen M. Amos, Dr. Binita K.C., Dr. Jennifer Wei for their extended support and motivation for this data analysis and project formalization.

## Models

### 1. K- Means Clustering

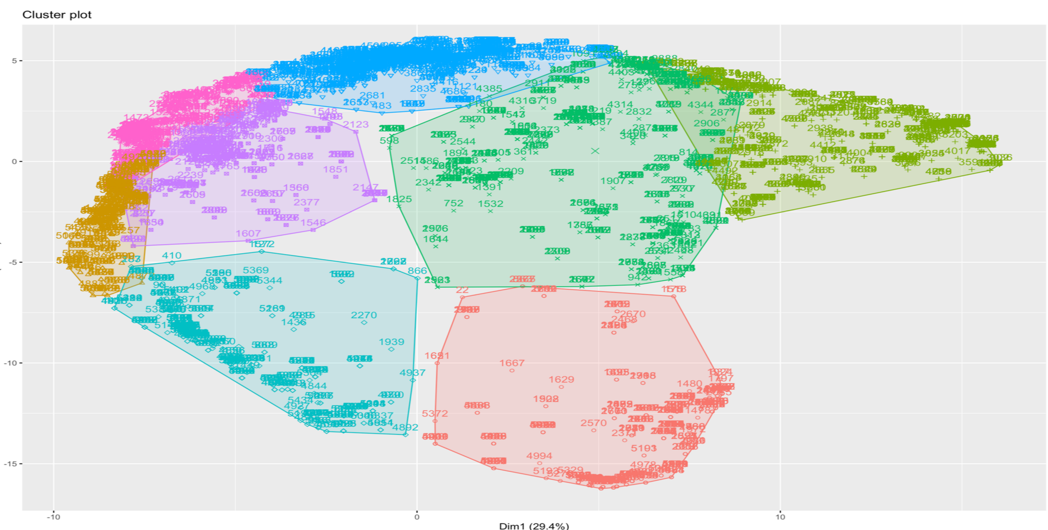

Fig 3. Clusters formed after applying K means

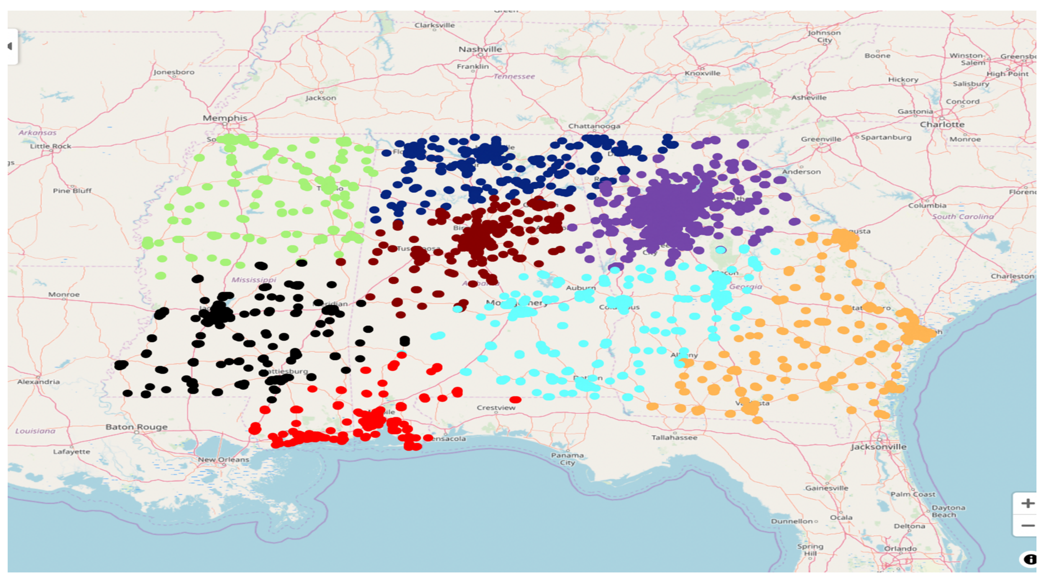

Fig 4. The clusters formed when visualized on a map

| Cluster | Color | Rank |
|---|---|---|
| 1 | Red | 1 |
| 2 | Lime Green | 5 |
| 3 | Orange | 8 |
| 4 | Light Blue | 7 |
| 5 | Black | 2 |
| 6 | Purple | 6 |
| 7 | Maroon | 4 |
| 8 | Dark Blue | 3 |

Table 1. KMeans clustering with Ranking (Higher the rank value shows, the more increased the precipitation)

### 2. Agglomerative Hierarchical Clustering

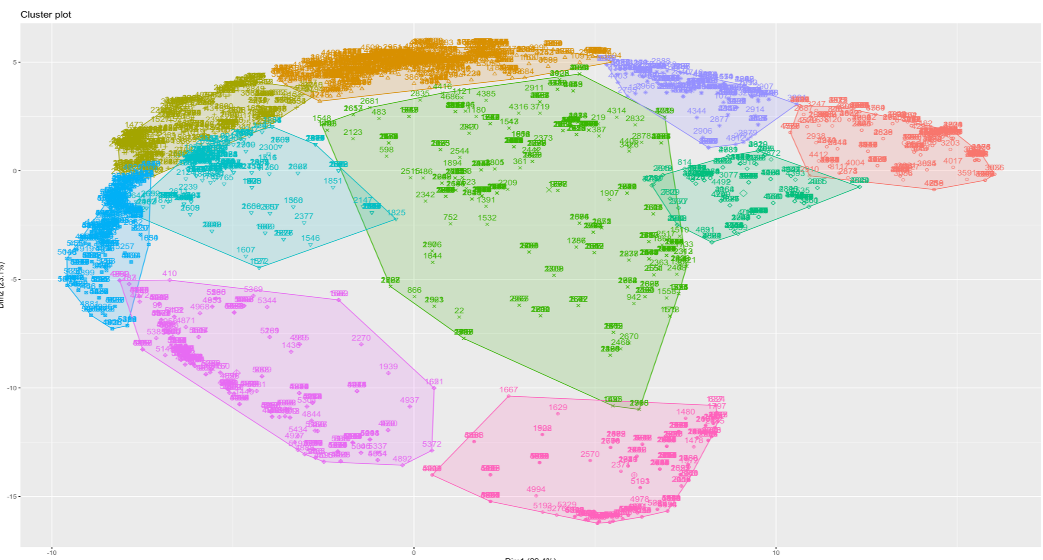

Fig 5. Clusters Visualized after model application

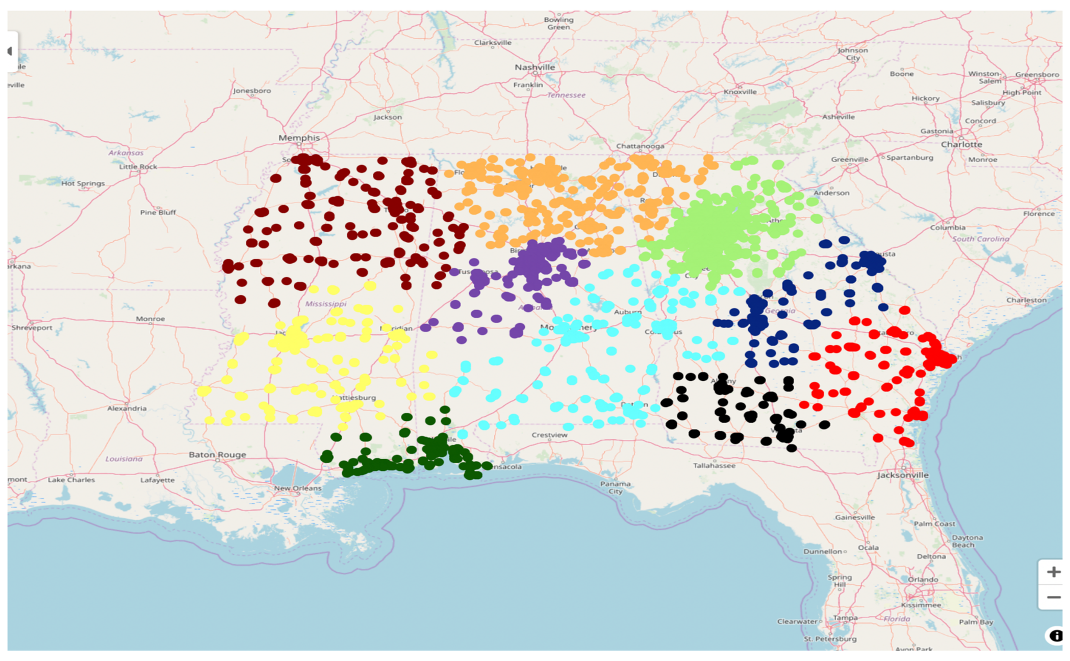

Fig 6.The clusters formed when visualized on a map

| Cluster | Color | Rank |
|---|---|---|
| 1 | Red | 10 |
| 2 | Lime Green | 7 |
| 3 | Orange | 5 |
| 4 | Light Blue | 6 |
| 5 | Black | 9 |
| 6 | Purple | 4 |
| 7 | Maroon | 1 |
| 8 | Dark Blue | 8 |
| 9 | Light Yellow | 2 |
| 10 | Dark Green | 3 |

Table 2. Agglomerative clustering with Ranking (Higher the rank value shows, the more increased the precipitation)

### 3. Divisive Clustering

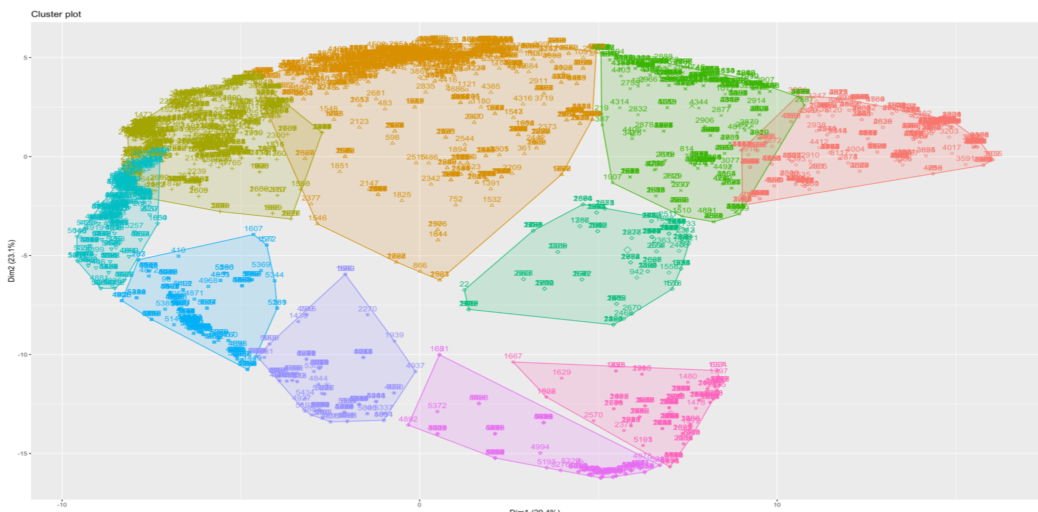

Fig 7. Clusters Visualized after model application

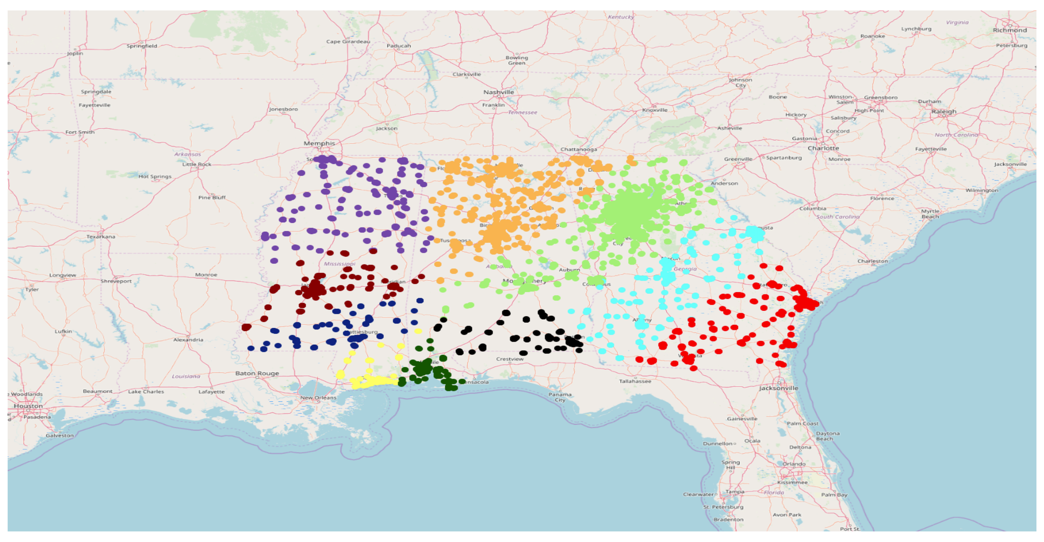

Fig 8 .The clusters formed when visualized on a map

| Cluster | Color | Rank |
|---|---|---|
| 1 | Red | 10 |
| 2 | Lime Green | 8 |
| 3 | Orange | 6 |
| 4 | Light Blue | 9 |
| 5 | Black | 7 |
| 6 | Purple | 2 |
| 7 | Maroon | 1 |
| 8 | Dark Blue | 4 |
| 9 | Light Yellow | 3 |
| 10 | Dark Green | 5 |

Table 3. Divisive clustering with Ranking (Higher the rank value shows the more increased the precipitation)

### 4. DBSCAN Clustering

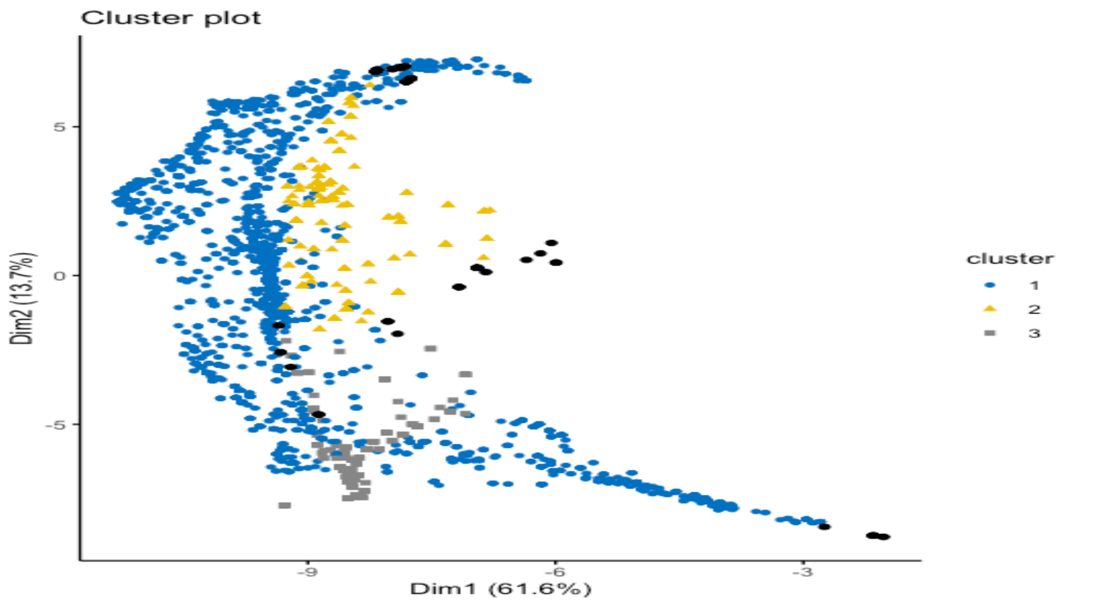


Fig 8. Clusters Visualized after model application

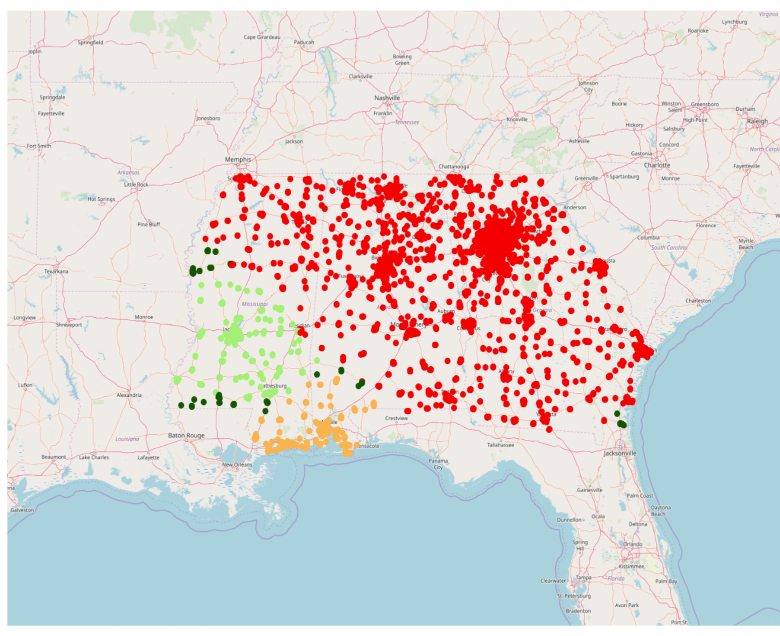

Fig 9.The clusters formed when visualized on a map

| Cluster | Color | Rank |
|---|---|---|
| 1 | Red | 3 |
| 2 | Lime Green | 1 |
| 3 | Orange | 2 |

Table 4. DBSCAN clustering with Ranking (Higher the rank value shows, the more increased the precipitation)

## Conclusion/ Future Work

Algorithms worked very well to provide us with clusters with their corresponding risks.
We gave the clusters a uniform coloring scheme while visualization and chose to give the clusters a rank based on their risk.
The rank was calculated by first taking the average of the cluster precipitation values
The models hold up as when we check the facilities against each other, the rank in which they fall are mostly similar.
In the future, we could look at making more models for this data set like Spectral Clustering and Meanshift clustering.
Apart from that, other factors like soil moisture and population density can also be considered for calculating an overall risk for the facilities.
Currently, we are only working on few selected States, Mississippi, Alabama and Georgia.. etc, over the course of time we plan expand this into the whole of United States, which will be the end goal of the project.

**Glossary:**

R – A program language to process data and perform statistical analysis
Package (P) or Library (R): software package to be loaded to perform extra tasks
Histogram - A histogram is a graphical display of data using bars of different heights.
DBSCAN – Density-Based Spatial Clustering of Applications with Noise. Finds core samples of high density and expands clusters from them. Good for data which contains clusters of similar density.

**Resources:**

Hierarchical Clustering : https://uc-r.github.io/hc_clustering
K-Means Clustering : https://uc-r.github.io/kmeans_clustering#kmeans
R visualization: https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/
NASA GPM IMERG Dataset : https://doi.org/10.5067/GPM/IMERG/3B-MONTH/06
Hospitals Dataset : https://hifld-geoplatform.opendata.arcgis.com/datasets/6ac5e325468c4cb9b905f1728d6fbf0f_0/data?geometry=10.030%2C-16.829%2C-40.946%2C72.120