# Optimizing Locations for Hospitals and EMS Stations

Chloe Fortuna, Takumi House, Shahrez Jan, Karam Yang
CS591: Data Mechanics – Spring 2017

## Abstract

Motor vehicle accidents are an inevitable part of living in a busy city such as Boston. The number of accidents that occur throughout the year account for a lot of traffic in the emergency rooms of hospitals close to the site of impact. Our team collectively decided that, since thousands of accidents occur each year but there are only 24 hospitals in Boston that are able to take in patients of such events, we need to determine how to direct the ambulances to the nearest hospital using the most optimal route.

These ideas prompted us to ask these questions:

1. Given a specific crash site, is there an EMS station and hospital within a reasonable distance of the crash that will get the victims the care they need as soon as possible? Are they currently in the most optimal location considering all the crashes that occur in the area?
2. How do we determine the most optimal path from the nearest EMS station to the crash site and from the crash site to the nearest hospital so that it takes the least amount of time from one end to the other?

## Datasets

Boston Hospitals:
https://data.cityofboston.gov/Public-Health/Hospital-Locations/46f7-2snz

Car Crashes in Boston:
http://datamechanics.io/data/cfortuna_houset_karamy_snjan19/CarCrashData.json

EMS Stations in Boston:
http://datamechanics.io/data/cfortuna_houset_karamy_snjan19/EMSStationsData.json

Routing on Map:
Requests to maps.googleapis.com

Using the retrieveData.py file we implemented, we retrieve the data from each of the links above and store it in our database (mongoDB). The Boston hospitals data received from the City of Boston data portal contains the names, addresses, neighborhoods, and coordinates of the 26 hospitals in Boston.

We received the car crash data from the Highway Division of MassDOT as an Excel spreadsheet. However, the coordinates in this spreadsheet were in a different scale known as the Massachusetts Mainland State Plan NAD 83 meters. Because of this, we used a program known as ExpertGPS to map these coordinates into latitude and longitude before turning the spreadsheet into JSON format. This dataset contains the information such as the coordinates, time of day, date, weather conditions, road conditions, and number of injuries in car crashes.

The EMS stations dataset was received from the City of Boston. However, this dataset was in the form of a chart on their website with only the addresses of the EMS stations. Therefore, we took this data and turned it into a CSV, and calculated the latitude and longitude based on their addresses. From there, we turned this CSV into JSON format.

**Method**

As a team, we decided that our objective was to determine whether the locations of currently existing hospitals and EMS stations are optimal in relation to the crash locations throughout the city. Through our analyses we hope to determine which location is the most optimal to have such facilities and be able to suggest a move of facility or the possibility of opening a new, small scale facility to help those in the event of a crash near the heavily populated crash areas. We also determined that we would like to show the most optimal route to get from the EMS station to the crash site and from the crash site to the nearest hospital in order to provide the crash victims with the care they need as soon as possible.

After retrieving the datasets and storing them into our database, we ran the k-means algorithm using the crash sites and hospital locations and then the crash sites and EMS station locations to determine the optimal locations for the hospitals and stations. Since there were a large number of data points that were analyzing, we created clusters out of the locations using the existing hospital and EMS station locations and used the k-means algorithm to pinpoint each clusters' centers. We determined the optimal locations to be the locations near areas that had a high volume of crashes and was at the center of these clusters. An example of these clusters using the crashed in Boston and hospital locations are shown below in Figure 1. Running the k-means algorithm gave us the longitudes and latitudes of each of our optimal locations, so we then used the Google Map API in order to convert them into actual addresses for the sake of readability and stored the longitude and latitude along with the address in our mongoDB.

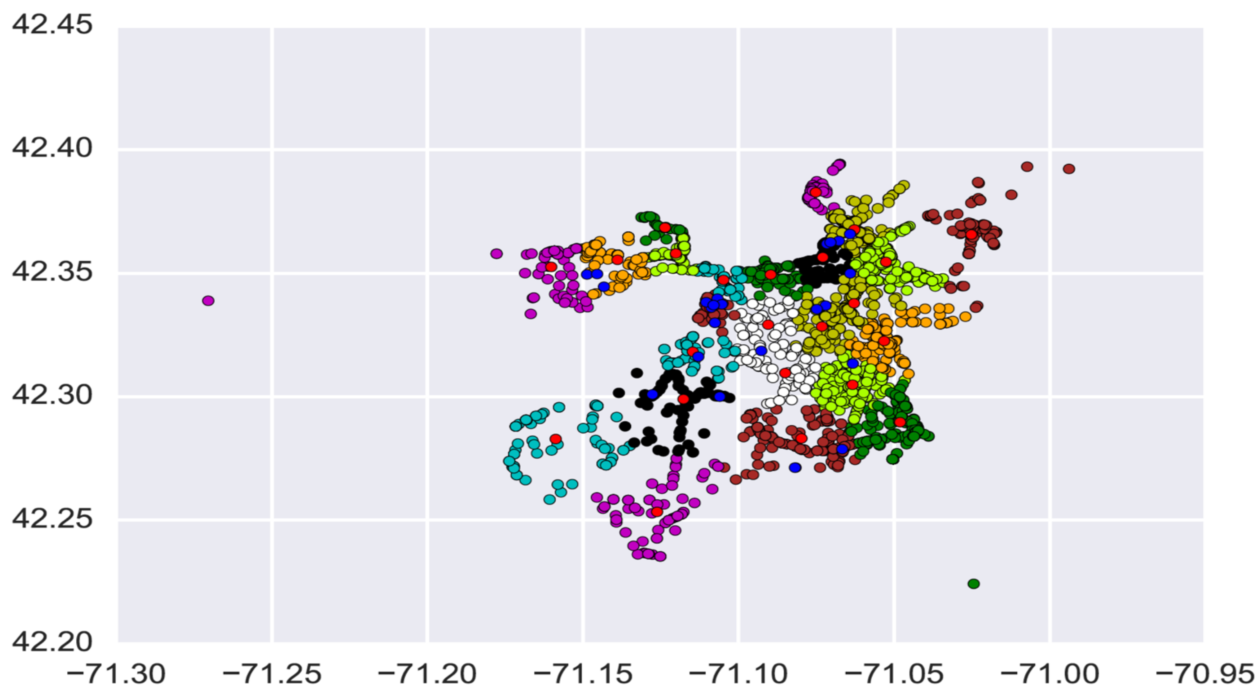


**Figure 1.** *The clusters of crashes throughout Boston in relation to existing hospital locations*

## Results

After determining the locations of our optimal locations for the hospitals and EMS stations, we plotted them on a map using D3.js by plotting them by their longitude and latitude. For the sake of having a variety of visualizations, we used the hospitals data and the crashes for one of our visualizations (hospitals.html) and the EMS stations and crashes for the other (ems.html). Our first visualization, hospitals.html, plots all our crash locations, existing hospital locations, and our calculated optimal hospital locations to give the user a grasp of what the data distribution is like. The user has the ability to filter out which data points they want to see on the map and which they want to hide. An example of what the user sees on the hospitals.html page is given below in Figure 2.

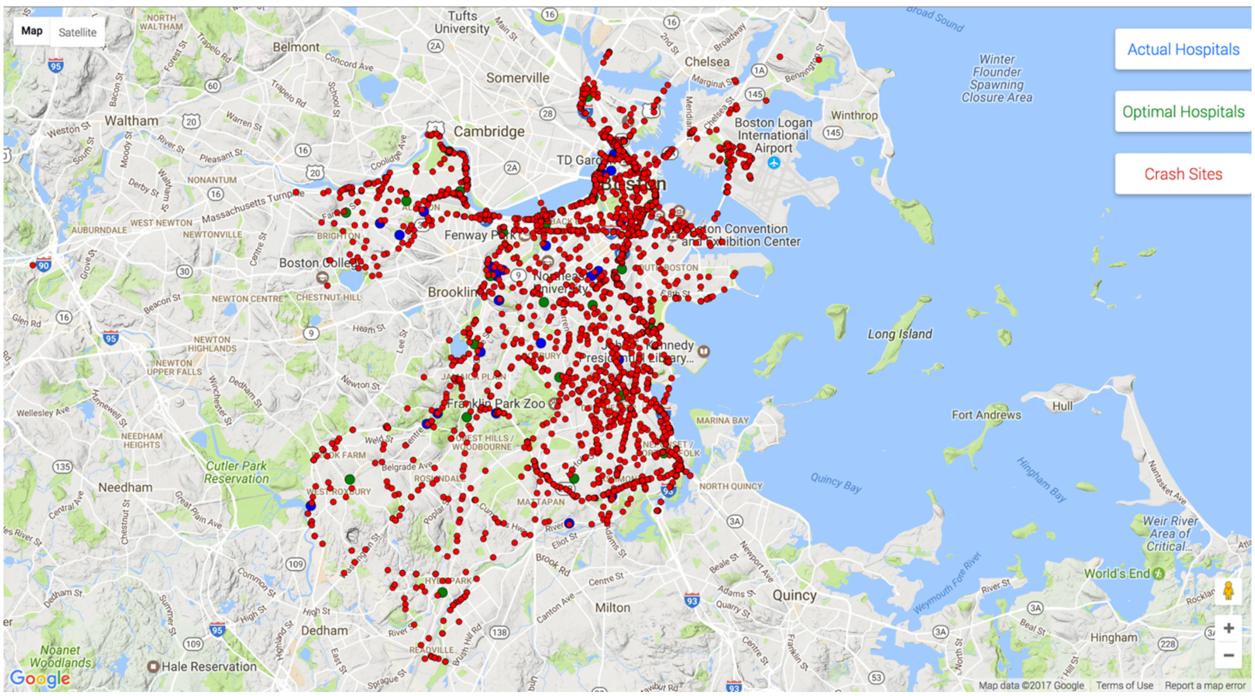

**Figure 2.** *hospitals.html: filterable map to see the actual hospitals, optimal hospitals, and crashes*

Our next visualization is also a map using D3.js, and like our previous visualization, maps the actual EMS stations, optimal EMS stations, and the crashes, but it also enables the user to choose any crash site and through a request to the Google Map API, it maps a route to the crash site from the existing EMS station closest to it. It also returns a route from our optimal location closest to the crash as well in order to show the user the comparisons in the existing and optimal locations. An issue/limitation that we ran into through the use of the Google Map API, however, was how the Google Map API's Distance Matrix Service geocoded latitude and longitude coordinates into addresses. The way that the API geocodes affected the actual locations of the EMS stations and crash sites and these locations have been distorted on the map. As a result, routing the EMS stations to crash sites may have some variance.

From our analyses of the locations of the hospitals and EMS stations throughout Boston in relation to the car crash locations, we have concluded that the current locations of these facilities are not optimal. When we picked several different crash sites on our ems.html to see the difference in the routes to the crash site, we noticed that on average, the distance between our optimal EMS station and the crash site was shorter than the existing location to crash site. A trial run of this is shown in Figure 3 below.

Click on a red crash site to get the route from the closest optimal (blue) and actual (green) EMS stations.

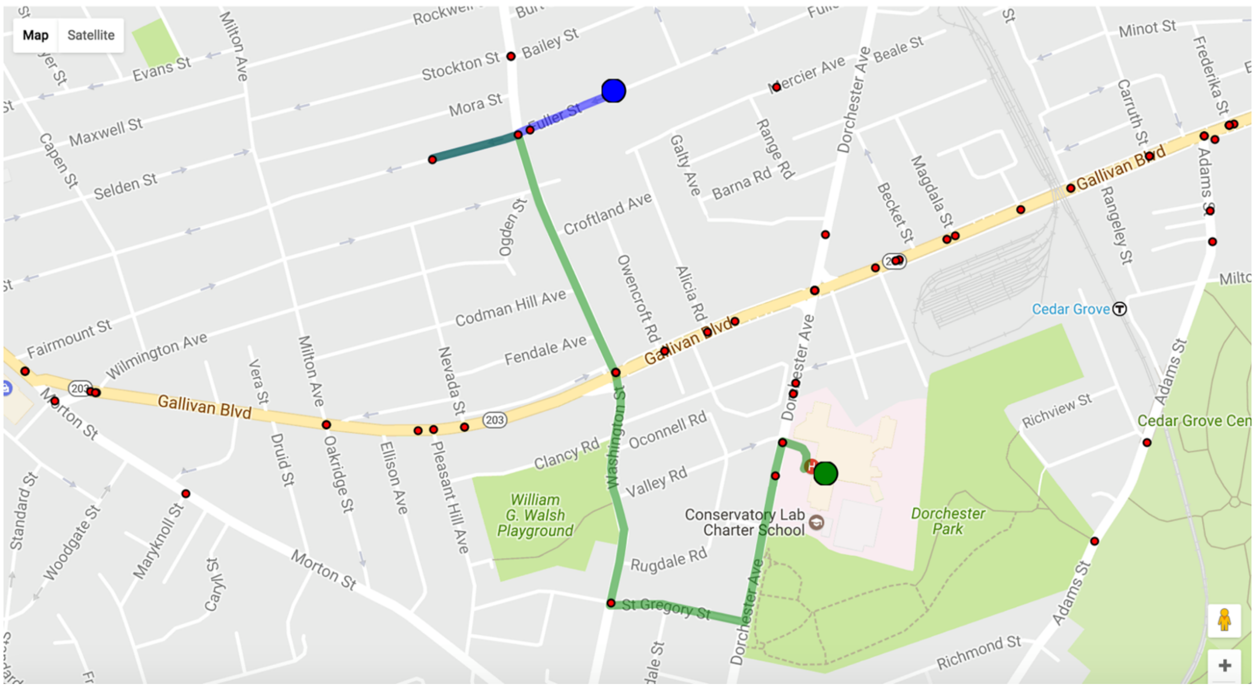


**Figure 3.** *ems.html: interactive map that returns routes from nearest existing and optimal stations to a given crash site*

## Future Work

Our next steps would be to take in consideration other factors to determine the optimal locations of these facilities instead of using only car crash sites in our algorithm. Some factors we would have liked to use in our analysis of the optimal locations include health conditions, population size, average income of area, traffic, and ease of access through public transportation. Using these additional factors, we could attempt to put hospitals in locations that have dense populations, have more frequent cases of illnesses, and can be accessed easily even by those without access to private transportation. Traffic data can be used to determine where the optimal EMS stations should be placed by analyzing locations with less congestion during peak accident times, so ambulances can easily get to crash locations and transport patients as soon as possible. We can also use the average income data to determine if the hospitals and EMS stations will have enough funding to be developed in these areas that we choose.