Title and Authors

U.S. Superfund Analysis by Milaud Nik-Ahd and Stanley Lai

Summary of research questions and results

We wanted to answer the following questions:

How many superfunds sites are there per state?

What are the highest/lowest average score per state?

What are the most frequent contaminants on all sites?

Is the site safe for human exposure?

Has the site been cleaned up?

Has the groundwater been stabilized for each site?

Is the site ready for industrial and/or recreational use?

We discuss the results of these questions down below in the ‘Results’ section.

Motivation and background

Before going over our data, it is important to note what a superfund site is. A superfund site is any land in the United States that has been contaminated by hazardous waste and identified by the EPA as a candidate for cleanup because it poses a risk to human health and/or the environment. These sites are placed on the National Priority List (NPL) and only sites with an HRS (Hazardous Ranking System) score of 28.5 or greater is eligible for placement on the NPL. The higher the score, the more polluted the site is. Our dataset consisted of information from all states as well as other U.S. territories such as Guam and the U.S. Virgin Islands. We wanted to see which areas of the United States were the most polluted and to see if anything was being done to remedy those areas.

Data set

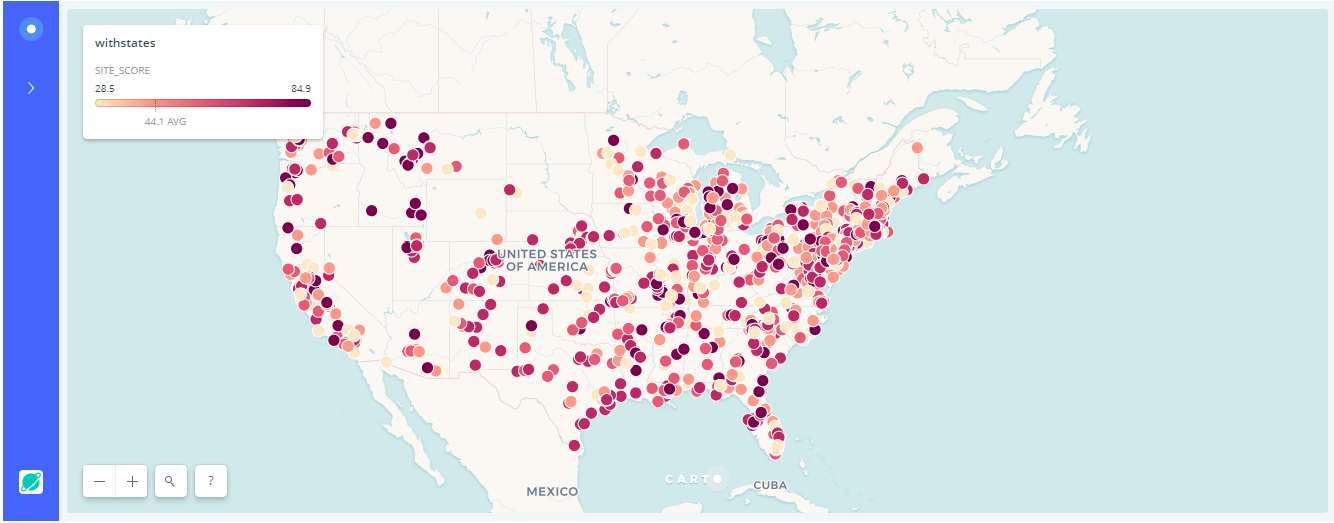
https://www.kaggle.com/srrobert50/federal-superfunds/data

Methodology

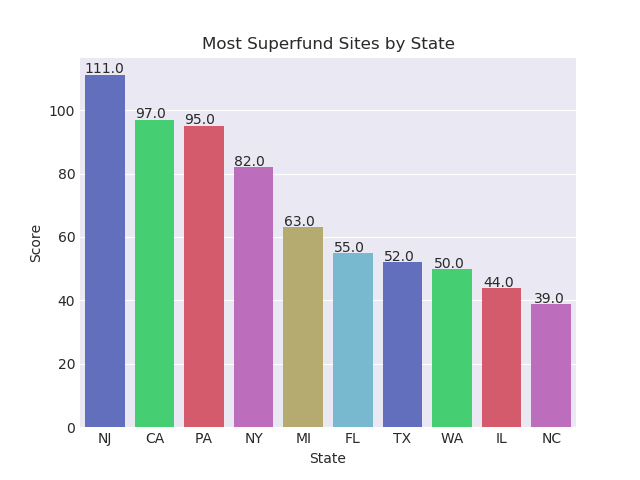
Our methodology was relatively simple. We imported our source csv file into a Pandas Dataframe and extracted relevant columns used for the analyses including address, longitude, latitude, and score. We used longitude and latitude to geocode the data and create a CARTO map output showing the distribution of Superfund sites across the US by score. The addresses were stripped for their state abbreviation and the resulting states were added as a new column. Additionally, we used BeautifulSoup to crawl the EPA Superfund Progress sites given in the csv file to extract site contaminants data and Performance Measures data. This new data was also added as a new column in the Dataframe, then plotted using matplotlib.

Results

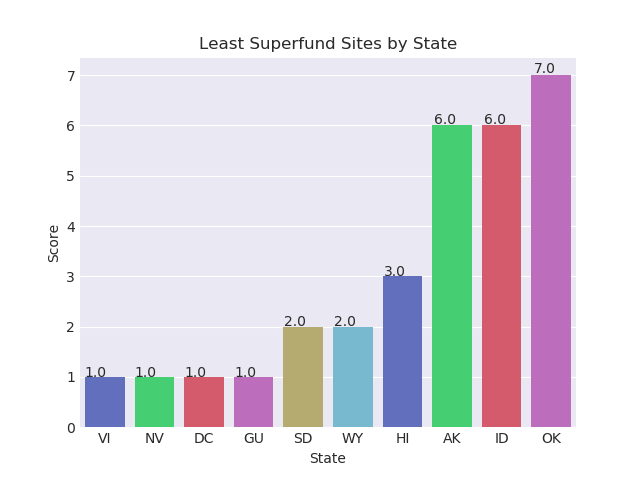
We first wanted to plot out a map of where each superfund site is located in the United States. As you can see there is a fair amount of distribution across the country, but many of the sites are seen along both coasts, as well as in central and the eastern parts of the country.



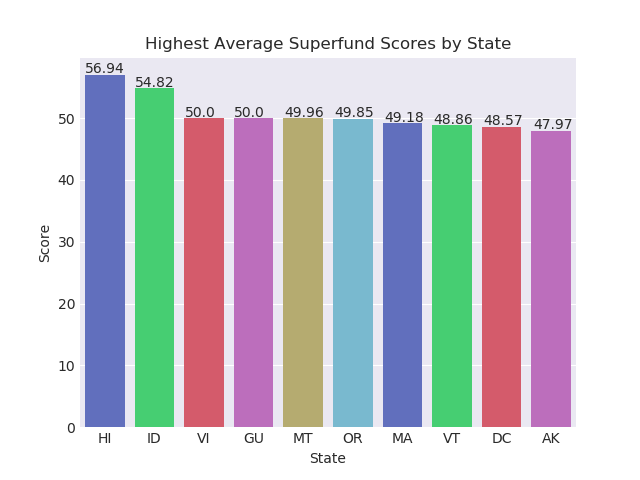
When looking at the number of each superfund sites per state, we found that the highest number of superfund sites belonged to New Jersey with 111. It was very surprising to see that New Jersey had so many given that it is a small state but when we looked into it more we found that there is a heavy industrial presence in New Jersey and the state has developed a reputation for dumping stuff anywhere. It was also surprising when comparing New Jersey to California which is a much larger state but holds the second-most superfund sites with 97. Pennsylvania and New York are the next highest with 95 and 82 superfund sites, then it begins to significantly drop off with the rest of the states.



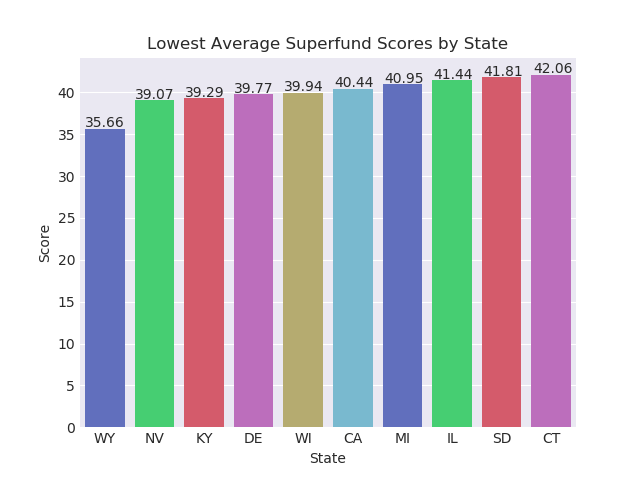
On the contrary, when looking at the states with the least number of superfund sites, we found that a lot of the smaller states/territories with a smaller population have the least number of superfund sites. Surprisingly, we found that Alaska has one of the lowest number of superfund sites with only 6 even though it is a relatively large state.



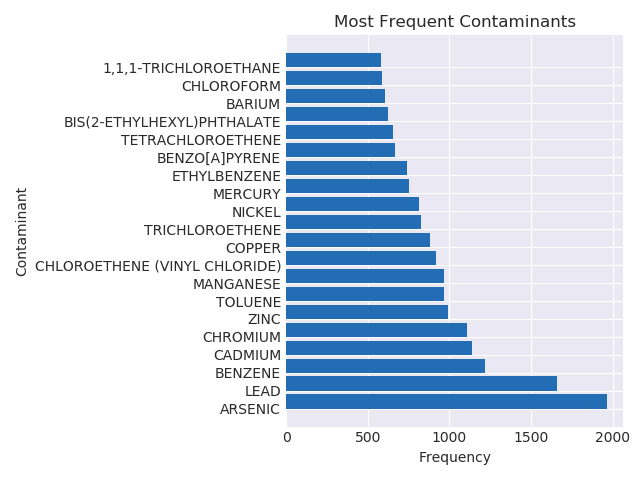
However, when looking at the average scores of superfund sites per state, many of the states with the smallest number of superfund sites had the dirtiest. Hawaii, Indiana, Alaska, U.S. Virgin Islands and Guam are in the group that had the lowest number of superfund sites, but the average scores for those sites ranged from around 48 to upwards of 56.



When looking at the lower average scores per state, we found that California may have had one of the most superfund sites in the country, but their average score was among the lowest with 40. South Dakota had some of the fewest superfund sites with only 2, and their average score was also among the lowest in the country with 41.8.

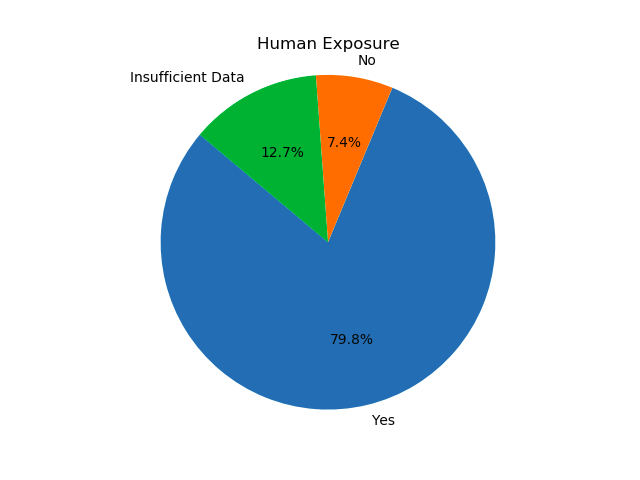


We originally wanted to find out the most common contaminants per site, and then find the most common ones per state, but that didn’t give us any valuable information as each site isn’t correlated with another even if they’re in the same state. Instead we wanted to find the most frequent contaminants among all sites across the country. We found that Arsenic, Lead, Benzene, Cadmium, and Chromium were the most frequent occurring in at least 1000 sites.

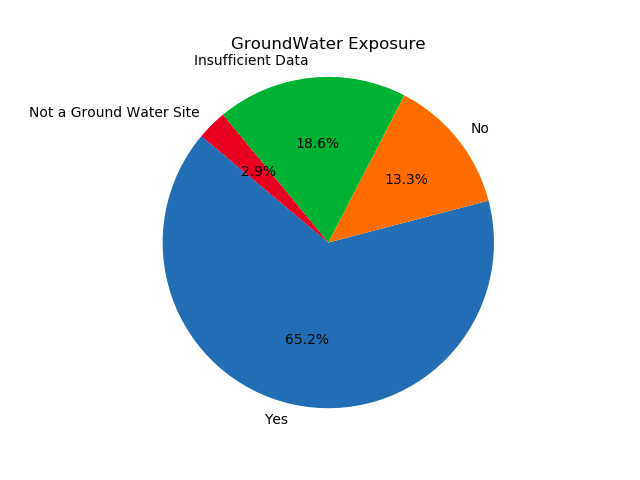


After analyzing the number of superfund sites, the scores among those sites, and the continents among them, we wanted to see what was being done to remedy those sites. We found another dataset that had information on whether the site is safe enough for human exposure, whether the site had been cleaned up, whether the groundwater has been stabilized to be ready for usage, and whether the site is ready for industrial and/or recreational use.

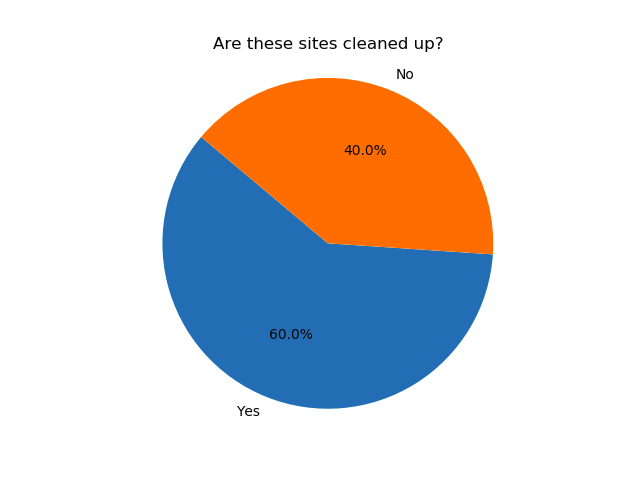
We found that most sites weren’t contaminated enough where it would be unsafe for people to be exposed to.



We also found that the groundwater in most sites have been stabilized enough where the EPA has reviewed all information on known and reasonably expected groundwater contamination and have concluded the migration of contaminated groundwater is stabilized and there is no unacceptable discharge to surface water.



We also found that most of the sites have completed the physical construction of the cleanup for the entire site.



However, even though most of the sites have been cleaned up to standards that meet EPA requirements, a vast majority of sites have not been cleared to be used for industrial or recreational purposes. To be cleared for usage, there are three factors that have to be met:

1. All cleanup goals affecting the current and reasonably anticipated future land uses of the entire site have been achieved, so there are no unacceptable risks.
2. All required land-use restrictions or other controls have been put in place.
3. The site has achieved Construction Complete status (meaning they have to be cleaned up as shown in the previous pie chart).

