Article

Storm Chasers: Synthesizing New England Weather Data on a Dashboard for Emergency Response Workers

Version December 21, 2022

Abstract: During a natural disaster, emergency responders have to quickly view many data types to decide how to react. Currently, there isn't a platform for the United States that contains all of this data. With the abundance of hazardous industrial sites in the New England (NE) region, there is a need for resources to guide emergency responders. We develop an interactive Shiny dashboard to help emergency responders in NE make data-driven decisions on how to target their resources. We compile, wrangle, and display open-source datasets with relevant geospatial, demographic, and weather information. We develop and integrate into our dashboard a real-time machine learning framework to predict, at a county level, whether or not a flash flood will occur with 93% accuracy, given date/time and current weather conditions. Using Worcester County, MA we show our dashboard can help emergency responders understand how environmental hazards and social factors interact within a region.

1. Introduction

During environmental disasters it is critical that emergency relief personnel are able to distribute supplies to areas in need quickly and efficiently. These situations are time sensitive so it is important that people are able to predict what areas will be affected and where relief efforts should be focused. Combining weather alerts and background data on one platform allows emergency relief personnel to avoid scanning weather channels themselves and keep track of information such as the locations of hospitals and warehouses [1]. As climate change continues, natural disasters will become more frequent and worsen [2]. Therefore, it is imperative that there is a system in place to assist relief workers during these natural disasters and make their jobs easier. Hopefully it will also benefit populations, especially vulnerable populations, by ensuring that the supplies they need get to them as soon as possible in critical instances. Company has developed detailed dashboards for North Carolina, Florida, Texas, and Louisiana, as well as a country-wide dashboard with some weather alerts. While this is a good start, most states lack data on the vulnerability of populations, points of interest, or other data that would provide response workers with background information to guide their responses in the event of a disaster. Here, we compile this data for the New England region (Massachusetts, Connecticut, Rhode Island, Maine, Vermont and New Hampshire). These states have some different risk factors than the southern states with preexisting dashboards. For example, southern states are mostly concerned with tropical cyclones, while the northern states are more susceptible to winter storms. We design a dashboard that fixes many of these issues, focusing on data points relevant to the Northeast.

New England experiences many extreme weather events including hurricanes, flooding, winter storms, and droughts. As climate change progresses these events will become more frequent and severe Between 1958 and 2012 there was more than a 70% increase in the amount of rainfall in heavy precipitation events in the Northeast, which is more than anywhere else in the United States, and projections indicate that precipitation will continue to increase [3]. Flooding events have also become more common due to the increase in precipitation and extreme weather events [4]. The severity of these flooding events are increasing, with 100-year flooding events now happening every 60 years, and it is projected they will become even more frequent and occur every 10-20 years for the Atlantic Coast in 2050 [4]. The Northeast also has some of the oldest buildings and infrastructure in the United States [5]. This can be a compounding factor when combined with extreme weather events and lead to more disastrous effects on local populations. Events with heavy precipitation can cause sewer-stormwater systems in the Northeast to overload and discharge wastewater into bodies of water used for drinking water [3]. The Northeast also contains hundreds of EPA-designated Superfund sites [6]. When these sites are hit by weather events such as hurricanes and flooding the toxic chemicals in them can contaminate waterways, affecting communities and farms [7]. Thus, increasingly extreme weather events and their potential for contamination make New England a location of interest for disaster preparedness work.

Climate change will have far reaching effects on human health, agriculture, and the ecosystems, yet it will not affect all populations equally. Natural disasters have a disproportionate long-term impact on vulnerable communities [8]. Low-income communities of color are often not able to evacuate and their communities are more vulnerable to flooding due to worse infrastructure [8]. Additionally, EPA Superfund sites are disproportionately concentrated near low-income communities of color [9]. Furthermore, even after damage occurs, FEMA often gives more aid to white victims of natural disasters versus people of color, even when the damage is the same [10]. Due to this disparity, we also focus on compiling data into our dashboard that can help emergency personnel locate and direct resources to socially vulnerable populations.

2. Data Sources

We gather data sources with variables relevant to our three main categories of interest: environmental landmarks, flood risks, and social vulnerabuility. All of the data sources we choose to

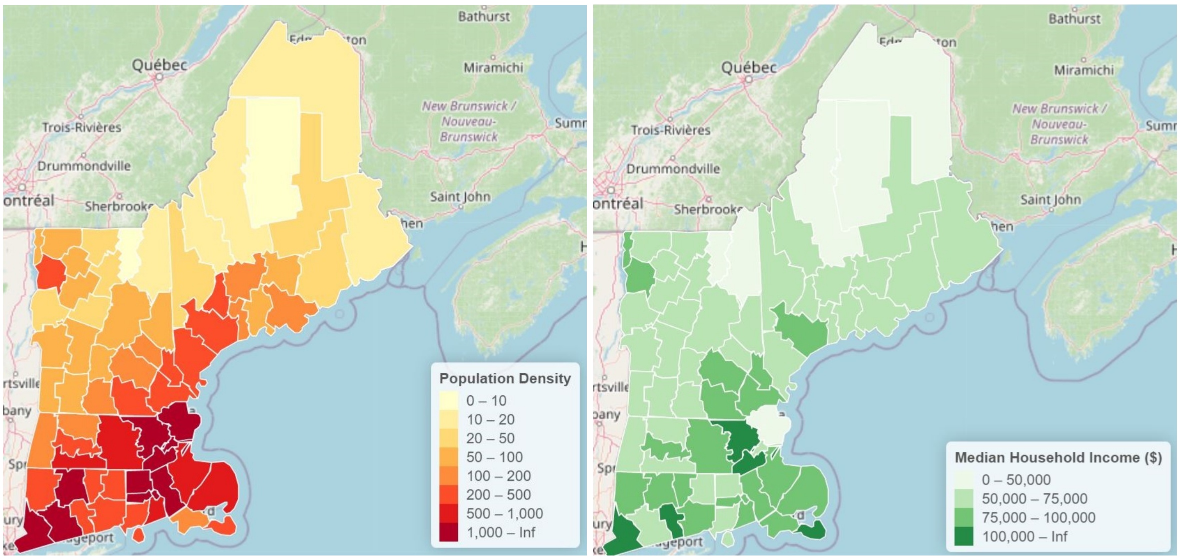

Figure 1. Population Density and Median Household Income from ACS

display are open source data sources, so they are accessible to anyone who wants to use them. Below we describe the data sources and which categories their variables fall into. In general, we use data that reports information at the county level for the six New England states: Massachusetts (MA), Connecticut (CT), Maine (ME), New Hampshire (NH), Rhode Island (RI), and Vermont (VT).

2.1. 2019 American Community Survey

The American Community Survey (ACS) is an annual nationwide survey that helps guide federal spending [11]. It collects information related to age, ancestry, place of birth, disability, educational attainment, race and ethnicity, health insurance coverage, income, occupation, employment status, housing and rent costs, sex, and housing, among other variables. We gather county-level information to help guide our understanding of demographics and social vulnerability in New England. Key variables in this dataset include county name, total population, population density (measured as number of people per square mile), median household income (in 2021 inflation-adjusted dollars), unemployment rate for the population 18 years and older, proportion of the population with a high school diploma or equivalent, number of renter-occupied housing units, and the proportion of the population that identifies with different racial and ethnic backgrounds (Table 1). The categories from the ACS related to race and ethnicity that we use are: White alone, Black or African-American alone, American Indian or Alaska Native alone, Asian alone, Native Hawaiian and other Pacific Islander alone, Some other race alone, and Two or more races. This dataset has 68 observations (one per county) and 31 variables (Table 1).

2.2. CDC Social Vulnerability Index

The Centers for Disease Control (CDC) assigns a Social Vulnerability Index (SVI) to each county in the United States. The CDC defines social vulnerability as the resilience of communities (the ability to survive and thrive) when confronted by external stresses on human health, stresses such as natural or human-caused disasters, or disease outbreaks [12]. This metric draws from 15 different variables recorded in the U.S Decennial Census Survey that relate to socioeconomic status, household composition and disability, minority status and language, and housing type and transportation [12]. We obtain county-level SVI measures for each New England state, resulting in a dataframe with 68 observations and two key variables: county name and SVI (percentile from 0-1) (Table 1).

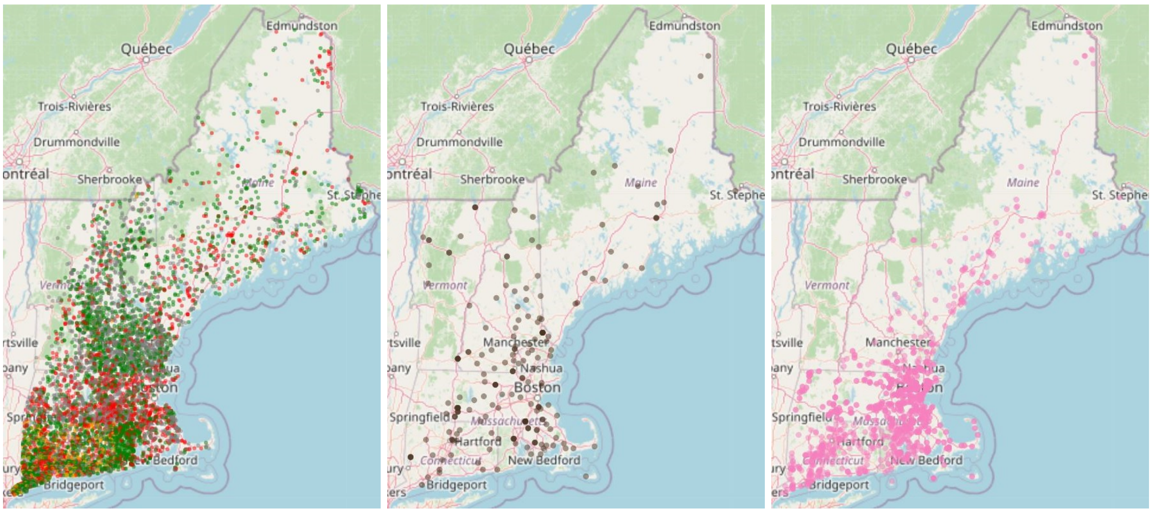

Figure 2. Visualzation of the dams, landfills, and EPA Superfund sites data layers respectively

2.3. New England Dams Database

The New England Dams Database draws information from state environmental databases,the Nature Conservancy's Northeast Aquatic Connectivity Tool, the National Hydrography Dataset Plus, the USGS National Land Cover Database, and the American Rivers' Removed Dams Database [13]. Dam information is relevant for understanding flood risk, since dam failures can cause severe flooding and aggravate other environmental hazards if floodwaters reach contaminated sites. There are 7,437 dams recorded in the current version of the database (downloaded 10/1/22) (Figure 1) and the relevant variables for each dam are dam identification and location (in the form of coordinates), dam status (Existing or Removed) and hazard classification (Negligible, Low, Moderate, Significant, or High) (Table 1).

2.4. EPA Project and Landfill Database

This database tracks key information for landfill gas energy projects and municipal solid waste landfills in the United States [14]. Landfill locations are relevant when floods or other similar disasters occur, since damage to the landfill site can cause contamination in the local groundwater or drinking water supply. We gather and join landfill locations for each New England state. Across New England, there are 201 landfills recorded in the databases (Figure 2). The key information recorded for each landfill includes landfill name, county, point coordinates, landfill status (Open or Closed), and waste in place (measured in tons) (Table 1).

2.5. EPA Superfund Sites Database

For each New England state, we also gather point locations of EPA-designated hazardous sites [6]. Hazardous sites fall into three main categories: Superfund sites, Brownfield sites, and RCRA Corrective Action sites. Superfund sites are toxic or hazardous locations designated through the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 that gives the EPA license to clean up toxic sites and hold responsible parties financially accountable [6]. Brownfields are properties that cannot be redeveloped or expanded because of environmental contamination [15]. RCRA-designated sites include hazardous and non-hazardous waste sites that the Resource Conservation and Recovery Act gives that EPA the right to oversee and manage [16]. Here, we focus on Superfund sites since they generally pose the greatest environmental and human health risks of all three categories. Our dataset of New England Superfund sites includes 1,338 observations and 9 variables, where each observation is a site and the key variables are site name, county name, latitude, and longitude (Table 1).

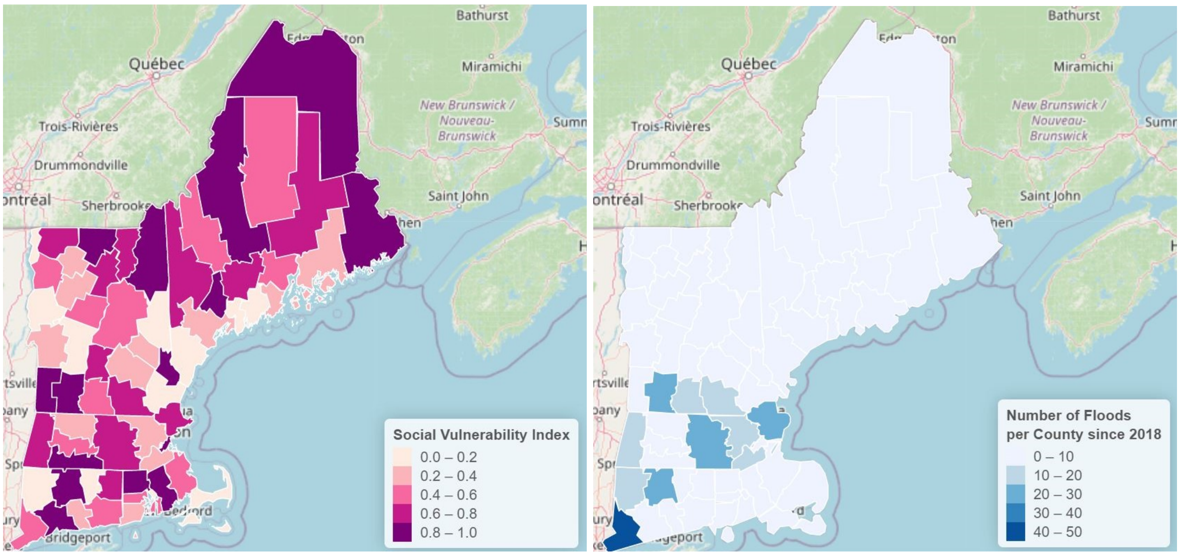

Figure 3. Social Vulnerability Index from CDC, floods per county calculated from NOAA Storm Events Database

2.6. NOAA Storm Events Database

The NOAA Storm Events Database records the occurrence of storms and other significant or rare weather events that have the potential to cause economic damage or loss of life [17]. The database contains storm records dating back to 1950, though to limit the amount of missing records and to recognize that climate change is quickly altering weather patterns, we restrict the database to only include records from January 1st, 2018 to January 1, 2022. This dataset contains 8,324 observations and 18 variables. Each observation is a weather event in New England within this date range, and key variables for our analyses include county, state, year, month, day of the month, beginning time, and event type.

2.7. NOAA Climate Data Online Database

The NOAA Climate Data Online database provides access to NOAA's archive of global climate and weather data [18]. We use this database to obtain daily summaries for each county from all weather stations in New England between January 1, 2018 and January 1, 2022. We restrict the data to this date range so we could join it with data from the Historic Storms Database. Since the observations in the Historic Storms Database represent one county on a given day, we average the observations from all weather stations in a given county for a given day to facilitate data joining. There are 40,874 observations and eight variables in this dataset. The key variables are state, county, year, month, day of the month, daily precipitation in inches, daily minimum temperature, and daily maximum temperature.

2.8. MRLC Land Statistics Dataset

We obtain land statistics on a county level for New England counties from a dataset from MRLC that was preprocessed to aggregate variables by county [19]. This dataset contains statistics gathered in 2019, and contains 67 observations and 10 variables. The variables are: county, land area in square feet, water area in square feet, latitude, longitude, mean land slope in the county, mean land elevation in the county, percent of the county area covered by water, percent of the county area covered by impervious surfaces, and percent of the county area with tree cover.

2.9. NWS API Web Service

We retrieve current temperature and precipitation conditions within Massachusetts via the open-source National Weather Service API [20]. After processing the data retrieved (see Methods), this dataset contains 14 observations—one per county—and three key variables: county, precipitation within the last hour, and temperature.

Table 1. Descriptions of data sources.

Data Source	Number of Observations	Number of Variables	Key Variables: Character	Key Variables: Numeric
2019 American Community Survey [11]	68	31	County, State	Total Population, Population Density (persons/sq. mile), Median Household Income, Unemployment Rate, Educational Attainment: High School or Higher, Renter-occupied Housing Units, Race
CDC Social Vulnerability Index [12]	68	3	County, State	SVI
New England Dams Database [13]	7,437	80	Dam Name, Town, State, Dam Status, Dam Hazard	Latitude, Longitude
EPA Project and Landfill Database [14]	201	16	Landfill Name, Landfill Address, County, State, Current Landfill Status	Latitude, Longitude, Waste In Place (Tons)
EPA Superfund Sites Database [21]	1,338	9	Site Name, Site Address, City, County, State, Interest Types	Latitude, Longitude
NOAA Historic Storm Events Database [17]	8,324	18	County, State, Event Type, Event Narrative	Year, Month, Day, Begin Time
NOAA Climate Data Online Database [18]	40,874	8	County, State, Year, Month, Day	Daily Precipitation, Daily Minimum Temperature, Daily Maximum Temperature
MRLC Land Statistics Dataset [19]	67	10	County	Latitude, Longitude, Land Area, Water Area, Mean Slope, Mean Elevation, Percent Water Coverage, Percent Impervious Surfaces, Percent Tree Cover
NWS API Web Service [20]	14	3	County	Precipitation in the past hour (inches), Temperature (°F)

3. Methods

3.1. Displaying data

To display the data layers that an emergency responder could need, we create a Shiny [22] dashboard. The dashboard has a main panel containing a Leaflet map of New England and a sidebar with a list of all the different data layers that users can add or remove from the map. Leaflet is a open source Javascript library used to build maps, we utilized the Leaflet R package for our project [23]. The layers we choose to display on our map of New England are dams, EPA superfund sites, landfills, social vulnerability index (SVI), population density, median household income, and floods per county since 2018. Dams, EPA Superfund sites, and landfills are all point data layers that we get from the New England Dam Database [13], EPA Superfund Sites Database [21], and EPA Project and Landfill Database [14] respectively. SVI, population density, median household income, and floods per county since 2018 are all polygon data layers. The SVI data is from the Center for Disease Control, which defines social vulnerability as the resilience of communities (the ability to survive and thrive) when confronted by external stresses on human health, stresses such as natural or human-caused disasters, or disease outbreaks [12]. The values for population density and median household income are from 2020 US census data [11], which we joined to a US county boundary shape file [24] after cleaning the census data. To calculate the number of floods per county we use NOAA historical data sets [17] between January 1, 2018 and January 1, 2022 and filter for all flash flooding events, which we then sum per county, and join the results to a US county boundaries shape file. Finally, we developed a layer that displays the predictions of our flash flood machine learning model based on current API weather data for Massachusetts counties. The second tab of our dashboard contains a list of all the different data sources with a description of each and where they can be found.

3.2. Modeling flash flood risks

3.2.1. Training on historical weather events

Data

We use binary classification techniques to predict, given county-level weather conditions with date/time, precipitation, and temperature information, whether or not that county is at risk of a flash flood. Fig. 4 shows an overview of our predictive modeling workflow. We train our classification model using historic storm and weather data from New England between January 1, 2018 and January 1, 2022. We obtain historical datasets from NOAA [18, NOAA [17]]. Each observation in the training dataset is a weather event. The target variable in the dataset is event type, which we recode a binary variable that indicates that the weather event is a flash flood (1) or is not a flash flood (0). The other variables in the dataset include geospatial information such as county FPS code, state FPS code, latitude, and longitude, information on the event's timing such as year, month, day of the month, and begin and end time, and weather information such as average county-level precipitation on that day and average minimum and maximum temperatures across the county on that day. We join this dataset with land usage datasets from MRLC [19] which contain county-level statistics such as mean elevation, mean slope, land area, water area, percent of land area with tree cover, and percent of land area with impervious surfaces. We choose to include these variables in our analysis because factors like elevation can influence which areas are prone to flash flooding, e.g. valleys or hollows, and the percentage of area covered by impervious surfaces impacts the effectiveness of water absorption. The dataset we begin the training process with has 14,832 observations and 20 variables.

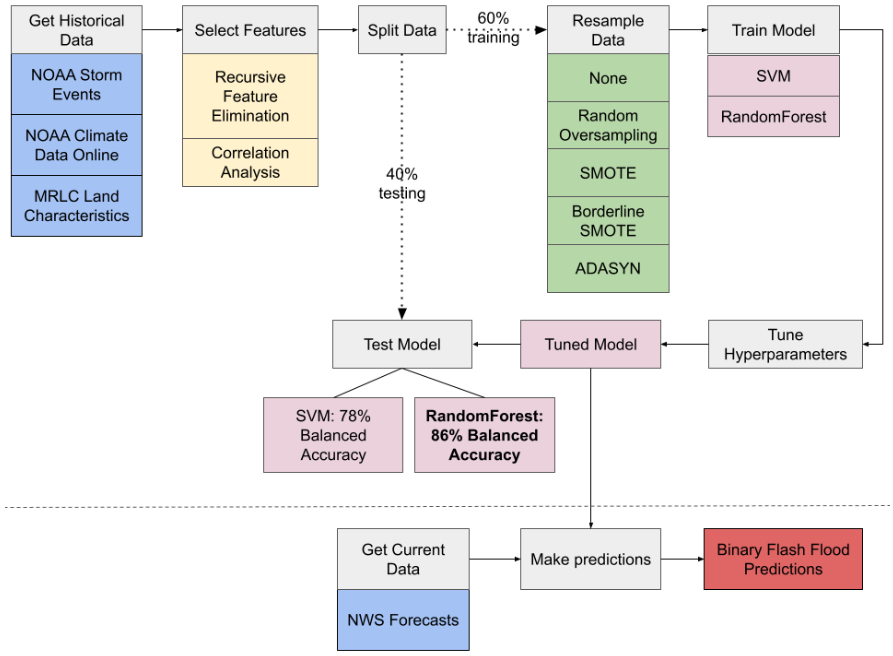

Figure 4. Overview of machine learning workflow. Gray boxes denote process steps, while colored boxes underneath represent particular data sources or techniques used at each step. Blue boxes represent datasets, yellow boxes represent feature selection techniques, green boxes represent resampling techniques, and pink boxes represent models. Red box denotes final output used on dashboard. Dashed horizontal line separates steps performed on historical weather datasets from steps performed on current weather dataset.

Feature Selection

We use two methods to select the optimal set of features from our historical weather dataset to predict flash floods. First, we compute a correlation matrix among all features and identify highly correlated features—which we define as features with $|r| > 0.9$—using the `highlyCorrelated()` function from the `corrplot` R package [25]. We remove highly correlated features from the historical weather dataset. Second, we use a recursive feature elimination algorithm to identify the features with the highest predictive power. Recursive feature elimination (RFE) works iteratively by fitting a machine learning model, ranking features' importance, and removing the least important features until a specified number of features is reached. We implement RFE on our historical weather dataset using the `rfe()` function from the caret R package [26]. We identify the optimal features for models with between one and 12 features and compute the model accuracy on the historical weather dataset with 10-fold cross-validation. We retain the optimal features from the model size that yields the highest cross-validation accuracy to produce a training dataset.

Resampling

The initial training dataset is highly imbalanced; only 7% of observations come from the positive class, which reflects the fact that most weather events in New England are not flash floods. Imbalanced datasets make classification tasks more difficult since most models often struggle to predict observations from the minority class correctly. One way to mitigate the effect of class imbalance is through resampling techniques, which balance the class distribution in the training dataset either by undersampling the majority class or oversampling the minority class. Since we do not want to reduce the size of our training dataset, we choose to use oversampling techniques. We test four oversampling techniques: Random oversampling, Synthetic minority oversampling technique (SMOTE), Borderline-SMOTE, and Adaptive synthetic oversampling (ADASYN), which we implement using the `smotefamily` R package [27]. Random oversampling samples with replacement from the minority class so that there are equal numbers of observations from both classes. SMOTE balances the class distribution by creating synthetic minority class observations [28]. For a given minority observation, it identifies the five nearest neighbors that are also minority observations. Depending on the degree of class imbalance in the dataset, it randomly selects a subset of the neighbor observations and creates synthetic points along the lines in feature space between the original observation and the neighbor observations [28]. Borderline-SMOTE is a SMOTE variant that works similarly, except that it only creates synthetic examples from observations that are near the border between the majority and minority classes in feature space [29]. ADASYN also creates synthetic observations in a similar manner to SMOTE, except that the number of synthetic observations generated per minority observation depends on the class distribution of its surrounding observations [30]. First, ADASYN finds the class distribution among the five nearest neighbors to a minority example and calculates the proportion of majority examples in the neighborhood. This proportion controls the number of neighbor minority observations that are sampled to create synthetic observations, such that more synthetic observations are created around isolated minority observations. Intuitively, this means that the ADASYN algorithm balances the class distribution by focusing on 'hard to learn' observations [30].

Model fitting

We test two types of supervised classification models on the training dataset. The first is a support vector machine (SVM). For a dataset with N features, an SVM attempts to find a hyperplane in N-dimensional space that separates the two classes in a dataset [31]. The second classification model is a random forest classifier. Random forest is an ensemble method, meaning that its prediction for an observation is an aggregate of multiple individual models' predictions for the same observation. The individual models in Random Forest are decision trees, which are flowchart-like structures in which each node represents a feature, each branch represents a decision rule, and each leaf represents an outcome [32]. Random Forest builds a specified number of decision trees to make predictions for observations in the training dataset, and averages the outcomes to obtain a final prediction for a given observation. We apply each model to the training dataset and make predictions on the testing dataset using 10-fold cross validation. We measure the true positive rate, true negative rate, false positive rate, false negative rate, and balanced accuracy for each model. True positive rate is the proportion of positive testing examples that are correctly predicted, and true negative rate is the proportion of negative testing examples that are correctly predicted. Conversely, false positive rate is the proportion of negative testing examples that are incorrectly predicted, and false negative rate is the proportion of positive testing examples that are incorrectly predicted. Balanced accuracy, which is the average of the true positive and true negative rates, is a more useful metric for imbalanced datasets than true accuracy, since it captures performance on both classes.

Hyperparameter Tuning

We define the optimal hyperparameters for a model as those which maximize its balanced accuracy. We select optimal hyperparameters using a grid search, which tests all possible combinations of supplied hyperparameter values. For the SVM classifier, we optimize three hyperparameters: kernel, gamma, and cost. The kernel choice determines the shape of the hyperplane that forms the decision boundary. The gamma parameter controls the curvature of the decision boundary, and is applicable only to non-linear kernels. The cost parameter controls the strictness of the model penalty for misclassification. For the Random Forest classifier, we optimize two hyperparameters: number of trees and mtry. Number of trees controls the number of decision trees that are built and averaged to determine a final prediction. Mtry controls the number of features that are sampled at each split in the decision tree.

3.2.2. Integrating current weather conditions

In order to integrate the current weather conditions we utilize the National Weather Service's (NWS) API, as detailed in the following section.

Accessing National Weather Service API

To apply our predictive model to real time weather data, we use the National Weather Service's weather observations API. We call the API on each of Massachusetts' 26 weather observation zones. Each API call for a zone yields the current weather conditions for each of the many weather stations within the given zone. We specifically retrieve two different attributes for each station: temperature (which we convert from Celsius to Fahrenheit), and precipitation in the last hour. Since our goal is to predict flash flood risks at the county level, we average the observations for every weather station in a county. After obtaining and processing current weather conditions from the NWS API, we add in variables for location and date and time that the data was gathered. Thus, we have a dataset with the same features as our training dataset for the predictive model.

Predicting flash flood risks

We generate real-time flood risk predictions at a county level in Massachusetts. When a user loads the dashboard, we call a single function that completes the API calls and data processing tasks described above to output a dataset with temperature and precipitation in the past hour by county. We add additional variables relating to event timing such that the current weather dataset has the same set of features as our flood modeling training dataset. We load the tuned classification model optimized to historical storm data and generate a binary prediction per county, denoting whether or not the current weather conditions pose a flash flood risk in that county. We add this prediction variable to the dataset with timing and weather conditions and return this dataset so it can be used as a layer on the dashboard.

4. Results

4.1. Displaying data

To display data regarding natural disasters to our users we develop a Shiny dashboard with an integrated Leaflet map. Figures 5 and 6 show the two main tabs of the dashboard. Fig. 5 depicts the main panel where users can interact with the different map layers, and Fig. 6 shows the second tab where users can learn more about the data sources incorporated into the dashboard. The user can interact with the check boxes on the left side of the panel and add/remove the different data layers to the map. For all of the layers, users can interact with it by clicking on a data point on the map and a popup will appear showing the location of the data point and the specific value of the data point for that layer. An image of each of the data layers can be found in the Data section of this paper along with its description. We also published our dashboard on the shinyapps.io server so it is accessible to the public. After analyzing our dashboard, we propose a few areas of significance to disaster relief personnel. These areas might either be more prone to a natural disaster or be more susceptible to damage if a natural disaster did occur. Fairfield CT, Essex, MA, Middlesex, MA, and Worcester, MA have a high population and have had frequent floods since 2018, which makes them potential areas of interest. It's worth noting that we are only measuring the frequency of floods, not how much damage floods do. Different magnitudes of floods can have drastically different effects, so this is only an approximation. There are also some counties throughout New England with high SVIs. This means that these communities are less resilient in the face of natural disasters and diseases, and could be more impacted by severe weather, so they are also points of interest for disaster relief personnel.

4.2. Modeling flash flood risks

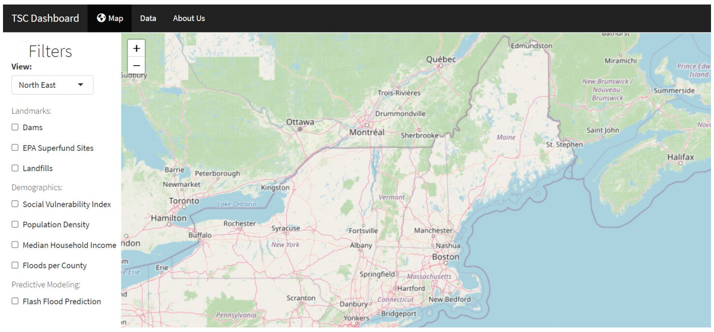

Figure 5. First tab of TSC Dashboard containing a Leaflet map and filters of the data

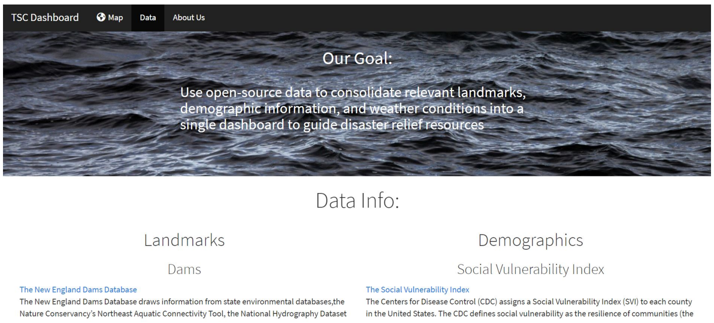

Figure 6. Second tab of TSC Dashboard with information about data sources

4.2.1. Training on historical weather events

We develop a predictive model that can predict, at a county level, whether or not current weather conditions pose a flash flood risk. Our audience for this model are first responders who want to prepare their flood responses in advance once a weather event is imminent. Ideally, we want a model with high balanced accuracy that performs well for both classes in the dataset. A model with a high true positive rate, or recall, is desirable because it means first responders will enter the fewest flash flood events unprepared. A model with a high true negative rate is also desirable because it means that first responders will not waste resources or unnecessarily alarm the public when there is no flash flood imminent.

Data

Our initial training dataset contains 20 variables and 1,889 observations, each of which represents a weather event. Of all observations, 7% are flash floods, and 93% are not flash floods. All of the flash floods in this dataset occurred in either Connecticut, Massachusetts, or Maine, likely because these states have the largest coastlines. 31 flash floods occurred in Massachusetts, 82 in Connecticut, and 8 in Maine. Of the four years of weather events included in this dataset, the most flash floods happened in 2021 (60), followed by 2018 (45). The most common times for flash floods to occur were during the months of September (57), July (35), August (22), and June (10). Only two flash flood events occurred outside these months, both during April. Given that flash flood events are restricted to only a subset of states and a subset of months during the year, we restrict the training dataset to weather events that occur during April through October and in Massachusetts, Connecticut, or Maine, reasoning that our model will likely yield more meaningful flash flood prediction results if it is trained on a more representative dataset of weather events that might be flash floods. After restricting this dataset, we are left with 1,202 observations.

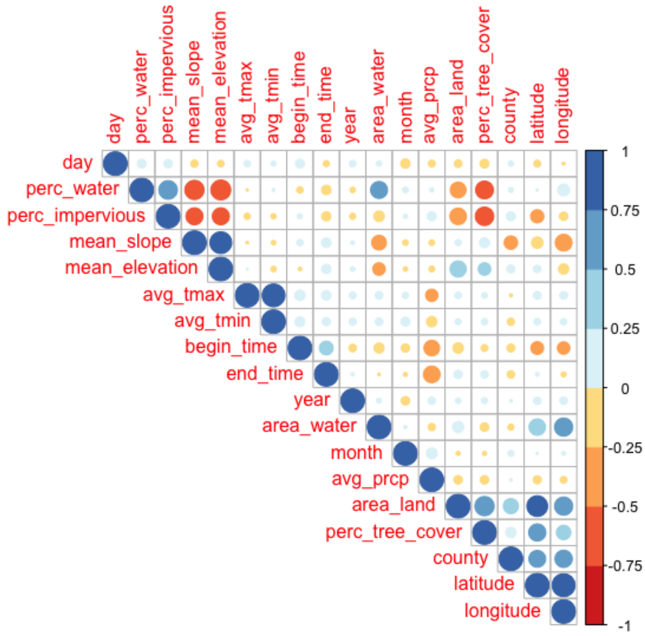

Figure 7. Correlation matrix of training data features.

Feature selection

We begin with a training dataset with 20 variables and 1,202 observations. When we compute a correlation matrix across all features (Fig. 7), we find that the variables denoting longitude, state, and minimum daily temperature have correlations greater than |0.9| with other variables. This makes sense intuitively because longitude is naturally highly correlated with latitude since we are focused on a small geographic area, and similarly, states are associated with particular counties and minimum temperatures are correlated with maximum temperatures. We concatenate state and county into a single numeric variable, since county codes are repeated across states and we want a unique identifier for each county. We remove the other highly correlated features from the training dataset, along with the ending time variable, which has a similar correlation pattern with the beginning time variable. We also remove the variable denoting the year of the weather event; this is ultimately not relevant since we hope to make predictions on current weather events beyond the end of 2021.

We further restrict the feature set to the most informative features by performing recursive feature elimination with 10-fold cross validation. We find that a model with five features produces the highest cross-validation accuracy of 96% (Fig. 9). The five features are: daily precipitation, month, maximum temperature, day of the month, and beginning time. Intuitively, daily precipitation makes sense as an important feature for flash flood modeling, and as we described above, flash floods are usually restricted to a subset of months. The median daily maximum temperature is slightly lower for days with flash floods than days without flash floods (75º F vs. 80º F), although the overall distributions are similar across both classes. Furthermore, more floods overall have occurred at the beginnings and ends of the month in the past four years than dates in the middle of the month. Intuitively, since all of the flash floods in our training dataset occurred in states with large coastlines, it is possible that the timing of floods across the month is linked to tidal cycles. We display the relationships between each individual predictor and flash flood occurrence in Fig. 9.

Modeling

To develop our model, we split the dataset 60/40 into a training set and a testing set (Fig. 4). Each set has the same features. We develop and tune all combinations of two classification algorithms and

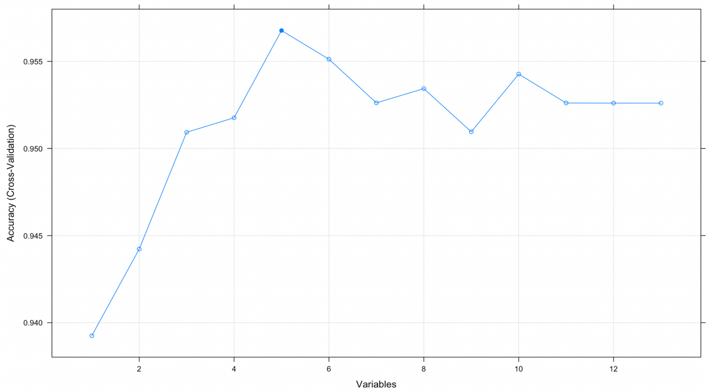

Figure 8. Cross-validation accuracy as a function of number of features in a model. The particular features at each model size were selected using Recursive Feature Elimination

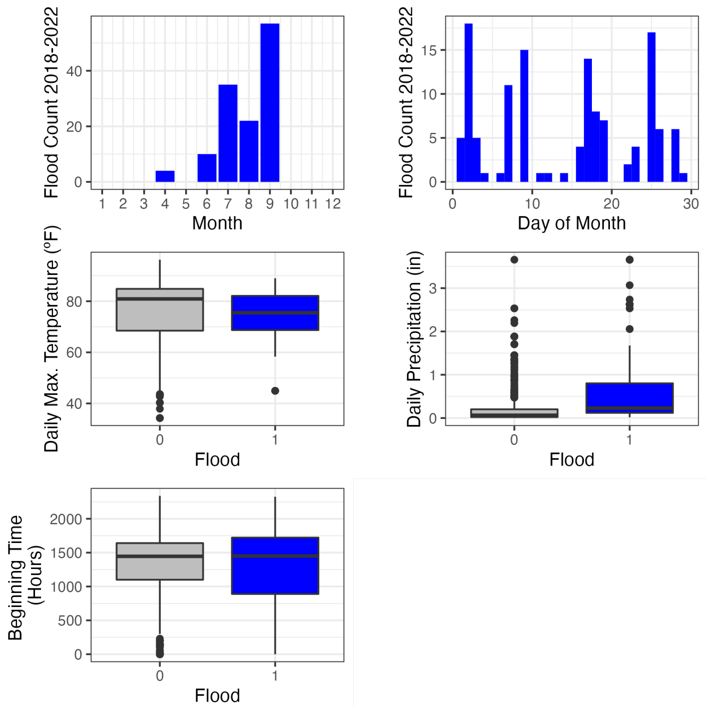

Figure 9. Individual relationships between each model feature and flash flood occurrence. Barplots display flash flood occurrence as a function of a given categorical feature, and boxplots compare the distributions of numerical features between flash flood events and other weather events

Table 2. SVM Classification Results. We test two different kernels—radial and linear—and tune the gamma and cost hyperparameters; in all experiments we achieve the best accuracy with a radial kernel and a gamma value of 10, and we achieve the best accuracy with a cost of 100 for all experiments except SVM with Borderline-SMOTE resampling, where we use a cost of 10.

Resampling	Balanced Accuracy	True Positive Rate	True Negative Rate
None	71%	42%	99%
Random Oversampling	72%	48%	97%
SMOTE	75%	52%	98%
ADASYN	78%	58%	98%
Borderline-SMOTE	78%	58%	98%

Table 3. Random Forest Classification Results

Resampling	Balanced Accuracy	True Positive Rate	True Negative Rate	Mtry	Number of Trees
None	86%	75%	96%	3	700
Random Oversampling	85%	74%	96%	4	200
SMOTE	75%	61%	96%	1	100
ADASYN	80%	65%	96%	2	300
Borderline-SMOTE	85%	75%	95%	2	200

four resampling techniques to find the model that will have the highest flash flood prediction accuracy. For SVM classifiers, we find that the radial kernel always produces superior performance, indicating that our data does not have a linear decision boundary. We achieve maximum performance when we use the ADASYN algorithm to balance the class distribution in the training dataset. The model achieves an 78% balanced accuracy, with a 58% true positive rate and 98% true negative rate (Table 2). In this context, the model correctly predicts 58% of flash flood events and correctly predicts 98% of non-flash-flood weather events. The optimal parameters for this model are a gamma value of 10 and a cost value of 100. With a RandomForest classifier, interestingly, we achieve maximum performance when we do not use resampling techniques. Without resampling, with an mtry value of 3, and with 700 decision trees, we achieve 86% balanced accuracy, with a 75% true positive rate and 96% true negative rate (Table 3). While the true negative rate is slightly lower for RandomForest than it is for SVM, RandomForest achieves a substantially better recall, which is important for our use case. Since it produces a higher balanced accuracy, we use the optimal RandomForest model to make flash flood predictions for current weather conditions.

4.2.2. Integrating current weather conditions

Each time a user loads the dashboard, we call a single function that retrieves a dataset of current temperature and precipitation conditions in Massachusetts from the NWS API. The function then averages these conditions by county, and we add additional date/time variables so that we are left with a dataset containing the same features as our training dataset of historical data. Using the tuned RandomForest model described above, we predict whether or not each Massachusetts county is currently at risk of a flash flood. The binary prediction results are a layer that users can view on the dashboard. During development, we validated the temperature and precipitation measurements that we retrieve against reports from The Weather Channel to ensure that our API calls are accurate. During the development period (late November - early December 2022) our model did not predict that any county was at risk of a flash flood; this validates the low false positive rate we obtained during our training process. In order to ensure that our model can make positive predictions, we generate a simulated dataset that reconstructs the weather conditions on July 17, 2018, during which flash floods occurred in Worcester and Suffolk Counties. We find that our model correctly predicts flash flood risks in these counties (Fig. 10).

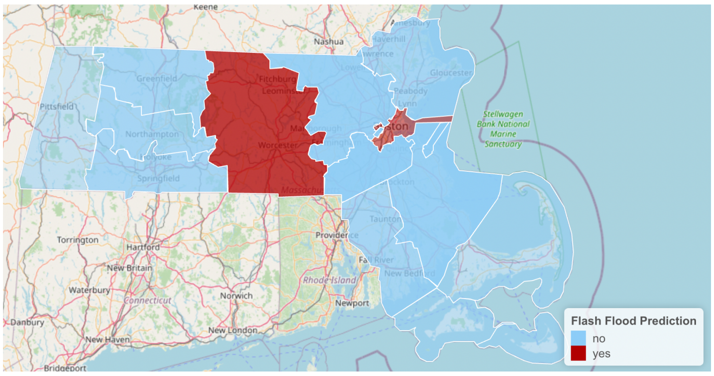

Figure 10. Reconstruction of the flash flood prediction dashboard layer on July 17, 2018, during which flash floods occurred in Worcester and Suffolk counties.

4.3. Case Study: Worcester County, MA

A potential use of our dashboard is to highlight possible areas of interest for natural disasters that emergency relief workers should pay attention to. Overlaying the layers on the dashboard can point out vulnerable areas. For example, Worcester County, MA has a SVI of .6923, 27 floods since 2018, a population density of 570.68 people per square mile, and a median household income of $79,142. Worcester has over 100 high-hazard dams (indicated by red dots on the dashboard), 45 EPA Superfund Sites, and 10 landfills.

On July 17, 2018, Worcester County had a devastating flash flood, shortly before the evening rush hour commute. Storm drains failed leading to flooded streets, sweeping away debris and dirt, and surrounding cars. With mass power outages impacting more than 3,000 households, civilians were left stranded and unprepared.

If any of the many high-hazard dams failed on this day, the consequences would have been devastating. Landfill sites, of which Worcester has many, have increased ground erosion, as well as an increased likelihood of leaching waste into the surrounding area. Superfund sites are already vulnerable to flooding, as they contain some of the most contaminated environments. As a county, Worcester has a high number of all of these potentially destructive areas. After inputting the weather data from the day of incidence, our dashboard's model would have predicted the Worcester County flash flood, and could have better prepared households and emergency response workers to mitigate the effects on their neighborhoods and the environment.

5. Discussion

Here, we consolidate relevant landmarks, demographic, and weather datasets into an interactive dashboard designed to help emergency responders in New England make data-driven decisions on where to direct time and resources during a natural disaster. By combining different datasets containing locations of environmental hazards and social vulnerability metrics, we can show first responders where multiple features overlap and elevate risks from severe weather. We also develop a RandomForest model that provides accurate, real-time predictions of flash flood risks at a county level in Massachusetts. Lastly, we use Worcester County, MA as a case study to illustrate the benefits that our product can provide to local emergency workers.

We overcame several challenges in the making of this dashboard. We experimented with other map layers that we ultimately did not incorporate because of missing data; for instance, we hoped to visualize hurricane/tsunami evacuation routes in each NE state but these are only available for Connecticut. This reflects a broader dilemma that other developers face, which is that the rich data sources needed to build a compelling dashboard do not always exist in regions that could benefit most from this product. Additionally, we navigated a data wrangling challenge when integrating current weather conditions from NWS into the dashboard. This dataset is quite large given the number of weather stations within New England, which is why we focus on flash flood risks within Massachusetts as a proof-of-concept. Weather observation zones (which can contain many weather stations) do not always fit neatly within county boundaries, and the number of zones per county varies. By researching the locations of weather observation zones, we were able to match the relevant Massachusetts zones to a county, and average weather conditions across stations to yield county-level estimates.

We note that there are several limitations and ethical considerations that are important to keep in mind when using the dashboard. For example, we use static datasets to visualize the landmarks and demographic variables on the dashboard rather than continuously retrieving current data via an API to reduce computational burden. While dam and landfill locations are constant, dam hazard status or waste in place at a landfill may change over time, so we cannot guarantee the most up-to-date information for those variables at all times. There are also several caveats to our predictive model. First, during the training process we restrict the training dataset to weather events within a subset of months and states—based on the distributions of where and when flash floods occurred between 2018-2022—in order to reduce class imbalance. This makes the model's predictions more accurate for the majority of flash floods, but it may result in poorer performance on rare flash flood events. For instance, under our model a flash flood in August on an 85° day has a better likelihood of detection than a flash flood in January on a 45° day. Second, since we average all current weather observations within a county and make predictions at a county level, our model may be less sensitive to localized extreme weather. Thus, we recommend that users keep these limitations in mind when preparing for flash flood conditions, and supplement our dashboard with more granular weather forecasts.

There are some ethical considerations in the usage of our dashboard. The primary purpose of this dashboard is to inform relief workers of relevant weather conditions and environmental hazards, and to make sure that resources are going to communities that face the greatest health and property risks. If the data sources and/or models are flawed, we risk promoting an inequitable distribution of resources. This is one of the reasons why we chose to use open-access datasets, primarily from governmental sources where we can access information about how the data were collected. However, we cannot guarantee that the sampling or surveying processes used to collect these datasets were unbiased. There is also the low possibility of false positives under our flash flood model, which could waste emergency responders' time and resources and create unnecessary public alarm. This is why we optimized our models to have high prediction accuracy for flash flood and non-flash-flood events. Finally, there is the consideration of how responders would use the data, possibly ignoring which communities are most vulnerable and instead choosing to prioritize areas where they are more concerned about monetary losses.

Future directions for this project include expansion of our flash flood modeling, seasonal customization, and integration of social media trends. We would like to extend our flash flood predictions, which we currently restrict to MA counties, to the rest of NE states. It would also be interesting to make more granular predictions, such as flash flood risks at a zip code level, which Company employees have successfully implemented in Florida using similar model features. From a technical standpoint, this extension is feasible for NE states. However, we anticipate that our training data may be insufficient, since flash floods are rare in most NE areas, especially compared to Florida where tropical storms are more frequent. Additionally, since New England experiences extreme seasons, there are distinct types of natural disasters that relief workers must prepare for. Here, we primarily highlight variables relevant for flooding events, which tend to occur in warmer months. We would like to add additional layers and filters to the dashboard so that users can customize it to the season. For instance, we could implement a similar predictive model to visualize which regions are at risk of heavy snowfall. We could add more landmarks such as the locations of shelters where people can find warmth and resources in the event of a wintertime power outage. Lastly, since socially vulnerable communities face greater risks during a disaster and may be overlooked by mainstream media coverage, social media updates could be a valuable source of insight into these communities' needs. We would like to mine geotagged posts via the Twitter API to display which hashtags and topics are trending in a given region during extreme weather events.

Abbreviations

The following abbreviations are used in this manuscript:

NOAA	National Oceanic Atmospheric Administration
NE	New England
MRLC	Multi-Resolution Land Characteristics Consortium
NWS	National Weather Service
SVM	Support Vector Machine
SVI	Social Vulnerability Index

References

1. Hall, T. A Vital IBM Technology Helps Nonprofits Prepare for Disaster Season, 2020. URL: https://newsroom.ibm.com/ORI-nonprofits-disaster.
2. US EPA. Climate Change Indicators: Weather and Climate, 2016. URL: https://www.epa.gov/climate-indicators/weather-climate.
3. US EPA. Climate Impacts in the Northeast. URL: https://climatechange.chicago.gov/climate-impacts/climate-impacts-northeast#Reference%201.
4. Ning, L.; Riddle, E.E.; Bradley, R.S. Projected Changes in Climate Extremes over the Northeastern United States. *Journal of Climate* **2015**, *28*, 3289–3310. doi:10.1175/JCLI-D-14-00150.1.
5. US. Climate Resilience Toolkit. Infrastructure and the Built Environment. URL: https://toolkit.climate.gov/regions/northeast/infrastructure-and-built-environment.
6. US EPA. What is Superfund?, 2017. URL: https://www.epa.gov/superfund/what-superfund.
7. McKenna, P. Climate Change Threatens 60% of Toxic Superfund Sites, GAO Finds, 2019. URL: https://insideclimatenews.org/news/20112019/superfund-flooded-climate-change-toxic-health-risk-sea-level-rise-wildfires-gao-report-epa/.
8. Walz, K. What Natural Disasters Reveal About Racism and Poverty, 2017. URL: https://www.povertylaw.org/article/what-natural-disasters-reveal-about-racism-and-poverty-2/.
9. Carter, J.; Kalman, C. A Toxic Relationship, 2020. URL: https://www.ucsusa.org/resources/toxic-relationship.
10. Flavelle, C. Why Does Disaster Aid Often Favor White People? *The New York Times* **2021**. URL: https://www.nytimes.com/2021/06/07/climate/FEMA-race-climate.html.
11. US Census Bureau. About the American Community Survey. URL: https://www.census.gov/programs-surveys/acs/about.html.
12. US CDC. CDC/ATSDR Social Vulnerability Index (SVI), 2022. URL: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.
13. New England Dams Database. URL: http://ddc-dams.sr.unh.edu/about/project_description/.
14. US EPA. Project and Landfill Data by State, 2017. URL: https://www.epa.gov/lmop/project-and-landfill-data-state.
15. US EPA. Brownfields, 2013. URL: https://www.epa.gov/brownfields.
16. US EPA. Resource Conservation and Recovery Act (RCRA) Overview, 2015. URL: https://www.epa.gov/rcra/resource-conservation-and-recovery-act-rcra-overview.
17. NOAA. Storm Events Database - Search Page | National Centers for Environmental Information. URL: https://www.ncdc.noaa.gov/stormevents/choosedates.jsp?statefips=23%2CMAINE#.
18. NOAA. Climate Data Online (CDO) - The National Climatic Data Center's (NCDC) Climate Data Online. URL: https://www.ncei.noaa.gov/cdo-web/.
19. MRLC. Data | Multi-Resolution Land Characteristics (MRLC) Consortium. URL: https://www.mrlc.gov/data.
20. US Department of Commerce, NOAA. API Web Service. Publisher: NOAA's National Weather Service.

21. US EPA. Cleaning Up in New England, 2015. URL: https://www.epa.gov/cleanups/cleaning-new-england.
22. RStudio, Inc. Easy web applications in R., 2013. URL: http://www.rstudio.com/shiny/.
23. Cheng, J.; Karambelkar, B.; Xie, Y.; Wickham, H.; Russell, K.; Johnson, K.; Schloerke, B.; library), j.F.a.c.j.; library), V.A.L.; library), C.L.; library), L.c.L.; plugin), B.C.l.m.; plugin), J.D.L.T.; plugin), B.B.L.M.; plugin), N.A.L.M.; plugin), L.V.L.a.m.; plugin), D.M.L.E.; plugin), K.A.P.; plugin), R.K.l.l.; plugin), M.l.o.; Bostock (topojson), M.; RStudio. leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library, 2022.
24. ArcGIS Hub. US County Boundaries. URL: https://public.opendatasoft.com/explore/dataset/us-county-boundaries/export/?disjunctive.statefp&disjunctive.countyfp&disjunctive.name&disjunctive.namelsad&disjunctive.stusab&disjunctive.state_name.
25. Wei, T.; Simko, V. *R package 'corrplot': Visualization of a Correlation Matrix*, 2021. (Version 0.92).
26. Kuhn, M. *caret: Classification and Regression Training*, 2022. R package version 6.0-93.
27. Siriseriwan, W. *smotefamily: A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE*, 2019. R package version 1.3.1.
28. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **2002**, *16*, 321–357. arXiv:1106.1813 [cs], doi:10.1613/jair.953.
29. Han, H.; Wang, W.Y.; Mao, B.H. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. Advances in Intelligent Computing; Huang, D.S.; Zhang, X.P.; Huang, G.B., Eds.; Springer: Berlin, Heidelberg, 2005; Lecture Notes in Computer Science, pp. 878–887. doi:10.1007/11538059_91.
30. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 1322–1328. ISSN: 2161-4407, doi:10.1109/IJCNN.2008.4633969.
31. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory; Association for Computing Machinery: New York, NY, USA, 1992; COLT '92, pp. 144–152. doi:10.1145/130385.130401.
32. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32. doi:10.1023/A:1010933404324.