An Analysis of Outcomes for FEMA Individual Assistance Recipients in the Context of Social Vulnerability

by

Gabriel Bann

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author ………………………………………………………………………………………....
Institute for Data, Systems, and Society
August 7, 2020

Certified by ………………………………………………………………………………
Jarrod Goentzel
Director, Humanitarian Supply Chain Lab
Thesis Supervisor

Accepted by ……………………………………………………………………………....
Noelle Eckley Selin
Director, Technology and Policy Program
Associate Professor, Institute for Data, Systems, and Society and Department of
Earth, Atmospheric and Planetary Sciences

An Analysis of Outcomes for FEMA Individual Assistance Recipients in the Context of Social Vulnerability

by

Gabriel Bann

Submitted to the Institute for Data, Systems, and Society
on August 7, 2020, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

Natural disasters such as hurricanes, floods, and wildfires have been increasing both in frequency and in the cost of damages. Different communities experience disasters differently, from the immediate losses to the long-term recovery process. Age, disability, race, housing status, and socioeconomic status are factors that can contribute to disparities in a person or community's vulnerability to disasters. Federal disaster aid policy has historically contributed to disaster vulnerability as well, at times enhancing disparities in survivors' ability to respond or recover. The Federal Emergency Management Agency (FEMA) has recently significantly expanded its efforts to individualize the survivor experience, and especially consider the survivor's social vulnerability. We received a large dataset with outcomes for Individual Assistance (IA) programs for registrants in eight different natural disasters from 2008-2017. My research has been to analyze these outcomes in order to explore potential disparities in the survivor experience based on demographic characteristics at the individual and community levels. This document covers my exploration of this data set in the context of social vulnerability. I employ high-level descriptive statistics as well as multi-level modeling techniques to depict the relationship between demographic variables associated with social vulnerability and outcomes in IA programs. Through these modeling techniques, I find that many individual and community-level demographic variables are strongly correlated with FEMA assistance outcomes including assistance amount and the number of contacts between the registrant and FEMA. Finally, I outline significant disparities in assistance levels for different demographic groups.

Thesis Supervisor: Jarrod Goentzel
Title: Director, Humanitarian Supply Chain Lab

Acknowledgments

There are several people who were instrumental in the creation of this document. First, I want to thank my advisor, Jarrod Goentzel, for his support over the past few years, starting with accepting me into the Humanitarian Supply Chain Lab. His expertise in disaster management and humanitarian aid logistics was invaluable to my progression in this work, and his advice made me a better researcher.

I want to thank Justin Steil for applying his extensive background in issues of social inequality to this work. Thanks for pushing me in the right direction.

I also want to thank Tim Russell and Michael Windle, who both welcomed me into HSCL and Center for Transportation Logistics at large. Our many informal conversations in the lab expanded my worldview and taught me about the various stakeholders in disaster management.

Some of the best parts about the Technology and Policy Program are the many, varied, and deep conversations with fellow students in the program. These conversations were so interesting, and I'm grateful for what you all taught me about politics, global issues, and technology. I hope to see everyone soon!

I want to thank in particular my roommates from the past year—Karan, Frank, and Nico. I could not have asked for a better quarantine crew. Thanks for the road trips, soccer games in the park, and sessions on the back balcony. Thanks for introducing me to new music, F1, the Premier League, and good food. I learned the most from you all, and can't wait to meet up someday, somewhere in the world.

Thanks to my parents for always answering my calls when I'm having a rough day, and for the morale boosts and care packages, and listening to my long rants about the news.

Thanks to my siblings—Seamus, Jonah, Dominic, William, and Lily, for being my best friends.

Finally, thanks to MK, for your unending support and encouragement. You made the last several months in Boston amazing, and I will be forever grateful for that.

Table of Contents

Chapter 1: Introduction

On November 27th, 2019, the top article on the New York Times website was titled "FEMA's Hurricane Aid to Puerto Rico and the Virgin Islands Has Stalled." It detailed FEMA's recovery efforts in the aftermath of 2017 Hurricanes Maria and Irma, and compared them to recovery efforts for other hurricanes that predominantly impacted mainland America. The New York Times found that in Puerto Rico, only 190 long-term recovery projects had been funded by FEMA, out of over 9,000 requests. After Hurricane Harvey in Texas, 3,700 projects had been funded two years later, similar to the result in Florida. The cause of such a disparity in aid has been the subject of significant disagreement. The government claimed the main issues were Puerto Rico's devastated economy, mismanaged funds, and disputes about the costs of recovery projects. But Puerto Rico thought that the government had been underestimating the recovery costs, and had deliberately delayed recovery efforts. San Juan mayor Carmen Yulín Cruz attributed the disparity to racism (Walker and Kaann-Youngs, 2019).

The unequal impact of natural disasters is a problem that is both current and historic, and has been well documented over the years. The poor recovery efforts in Puerto Rico and the Virgin Islands are not the first of their kind, and they will not be the last. In this section, I will first explain the increasing importance of studying the impacts of natural disasters. Then, I will put the study of natural disasters and their impacts in historical context. Next, I will discuss the various ways in which natural disasters impact people differently, from intersections with social vulnerability to government policy, including historical examples.

Why Study Disasters?

First I want to provide some contextual basis for the growing relevance of natural disasters in society. Disasters are increasing in frequency, cost, and scale. The advent of global warming means that extreme weather, including natural disasters, will only increase in coming years. Over 170 peer-reviewed studies have already linked climate change to weather anomalies, and the field of attribution science has improved to the point that weather agencies will soon be able to say, in real-time, how much climate change increased the likelihood of an individual occurrence (Schiermeier, 2018). Disasters are also becoming much costlier. 2017 became the costliest year on record for natural disasters, with just three hurricanes, Harvey, Irma, and Maria, causing around $265 billion in damages (NOAA 2018). A 2013 study found that even if the number of yearly disasters stays constant, by 2025 damage costs will rise by a factor of 1.3-1.7, and by 2025 costs could quadruple (Preston, 719). Finally, disasters impact many more people than might be presumed. Data from the Spatial Hazard Events and Losses Database for the United States shows that 99.7 percent of US counties "have experienced notable property damage from natural hazards" (Howell and Elliott, 2018). Additionally, "over the past half-century, the average US county has experienced damages from five events *per year* totaling millions of dollars annually" (Howell and Elliott, 2018). In summary, disasters will impact more people, more often, and at greater cost.

These widespread impacts provide a rich opportunity for researchers to seek to understand potential disparities in these impacts along lines of race, class, gender, or other demographic distinctions. Given the increasing ubiquity of disasters and their impacts, those interested in social inequality should be concerned with how disasters intersect with social vulnerability and create or perpetuate inequality.

Definitions and Historical Background

One of the most famous definitions of disasters comes from Charles Fritz, who said that a disaster is an "event, concentrated in time and space, in which a society… undergoes severe danger and incurs such losses to its members and physical appurtenances that the social structure is disrupted and the fulfillment of all or some of the essential functions of the society is disrupted." The definition of social vulnerability more or less refers to the ability of people to respond to a shock to their society or their environment, from an unexpected job loss to a car breakdown. Different people and communities can have a harder time dealing with these issues, depending on disability, or socioeconomic status, or education level. There a few subtle but important aspects of these definitions. A key component of the definition is that there must be a society that the disaster happens to. A fire that consumes an uninhabited island, wouldn't really be considered a disaster because there is no human impact, or at least no immediate human impact. It is also easy to see how these definitions can be combined. Disasters can be thought of as societal disruption to the most extreme degree, and people that are more vulnerable to societal disruptions can be expected to be vulnerable to disasters (Tierney, 2019).

Therefore, the study of disaster vulnerability seeks to determine how different people and communities experience disasters differently, both in the immediate losses and more long term recovery. The origin of this research dates back to the late 1940s, when the Army Chemical Center contracted the National Opinion Research Center to study the impact of "extreme threat situations" on behavior in the aftermath of a toxic smog event in Pennsylvania. This study was the first of its kind, and one of the researchers, Henry Quarentelli, went on to found the Disaster Research Center in 1963, the first research center focused on how disasters impact social behavior (Tierney, 2019). Though its research was often limited to determining how people respond to stressful situations, the research was important nonetheless. It debunked the popular notion that people panic or turn violent when their environment is drastically changed—instead they found that people exhibit much more pro-social, positive behavior, and the community largely comes together after a disaster (Tierney, 2019).

Scientists did not study how vulnerability differs from group to group until later. Two contributions, *Interpretations of Calamity*, published in 1983, and *At Risk: Natural Hazards, People's Vulnerability, and Disasters*, published in 1994, were pivotal in moving disaster research toward vulnerability studies (Tierney, 2019). Though they focused on class rather than race or gender, they helped reconceptualize disasters as shocks to society that would impact people differently based on their societally determined ability to adapt (Tierney, 2019). In 1997, the volume *Hurricane Andrew:*

Ethnicity, Gender, and the Sociology of Disasters was published. It chronicled the response and recovery for Hurricane Andrew, which affected Miami and the surrounding area, and discussed differences in experience for people of various ethnic and racial groups. The authors found that African Americans had a much more difficult time recovering than Cuban Americans or white Americans, and this corresponded to the groups' relative political and social power in that area. The volume also discussed how gender impacted recovery, which added intersectionality to the academic discussion (Tierney, 2019).

Over time, researchers began to consider disasters as not just something that happen to society, but also a product of that society and of government policy. A disaster is not just an environmental hazard, but also the losses that come with it, and these losses are heavily determined by the social structure and economic and political factors. In this reading, an understanding of vulnerability is actually central to the understanding of disasters, because the losses and the unequal impact are not just components of the disaster, they constitute why we care about the disaster at all.

This transformation primed the academic community for heavy analysis of a few focusing events, the most important being Hurricane Katrina. Katrina presented a case filled with disparities, from access to resources and the immediate impact of the hurricane, to the robustness of the response and recovery effort. These disparities were studied extensively by sociologists, and formed the foundation of where disaster sociology and vulnerability research stands today (Tierney, 2019). In the next section, I will describe in more detail the ways in which disaster sociologists have come to consider how social vulnerability relates to disasters.

Hurricane Katrina: A Case Study in Disaster Vulnerability

Hurricane Katrina is an excellent case to understand the variety of ways in which disparities in disaster vulnerability arise, especially due to the level of social vulnerability prior to the disaster and the scale of the disaster itself. The first set of disparities in Katrina is related to how communities were immediately impacted, and how they were able to respond. In Katrina, the more socially vulnerable, especially poor Black people, often did not have the resources necessary to evacuate New Orleans before the hurricane hit. As much as 26 percent of New Orleans households did not own a car, while at the national level, this statistic is only nine percent. Others said they did not hear or understand the evacuation order (Fussell, 2015). When it came to returning home after the hurricane, the government first had to declare the neighborhood open. The first to be opened for return were the least damaged, where the most socioeconomically privileged lived, while the last were the poorest and least valuable, closer to the ocean and at lower ground. This most often disadvantaged Black survivors (Fussell, 2015). This ultimately meant that demographics shifted in the aftermath of Katrina as more socially vulnerable people returned to New Orleans at a much slower rate than more affluent people (Fussell, 2015).

The government response after Katrina was highly dependent on racial factors in other ways as well. One of the most egregious examples of this was the government's decision, spurred on by the media, to prioritize "public safety" over recovery and aid efforts. As explained earlier, contrary to popular opinion the history of disasters shows that communities respond in a positive, pro-social way. This was not the media framing of the aftermath of Katrina. The prominent media coverage of New Orleans played on negative stereotypes of Black people as criminals, and chose to highlight "looting and lawlessness" (Tierney, Bevc, and Kuligowski, 2006). This coverage was highly exaggerated, to say the least. Articles from *The New York Times* and *The Washington Post* included language like "opportunistic thieves" and "armed thugs." *The New York Times* reported later that by the Superdome, "rapes and assaults were occurring unimpeded in the neighborhood streets," and "America is once more plunged into a snake pit of anarchy, death, looting, raping, marauding thugs…" (Tierney, Bevc, and Kuligowski, 2006). The reports of rapes and murders in the Superdome were revealed later to be false (Tierney, Bevc, and Kuligowski, 2006). But a highly racist scene had already been set. This scene soon worsened as National Guard and law enforcement were sent in droves, subsequently leading the media to frame New Orleans as a "war zone" (Tierney, Bevc, and

Kuligowski, 2006). Over 72,000 troops were sent in (Tierney, Bevc, and Kuligowski, 2006). The military's presence in New Orleans was viewed by policymakers and the media as a positive step forward, one that was bringing "law and order" to the area (Tierney, Bevc, and Kuligowski, 2006). Instead, militarization hindered recovery efforts and diverted aid resources. At one point, on September 1st, 2005, Louisiana Governor Kathleen Blanco "called off emergency rescue operations" so that "public agencies could devote all their attentions to looting" (Tierney, 2006). This militaristic approach, based on a mythology of the criminality of Black survivors, can be seen as further disaster vulnerability for people of color (Tierney, Bevc, and Kuligowski, 2006). The higher the minority population, the higher the likelihood of a militarized response, and the worse the overall outcome for the population.

Further Disaster Vulnerability Research

There are other examples of disparities in recovery following natural disasters. After Hurricane Harvey, some Texas towns received highly disproportionate levels of aid. The Southeast Texas Regional Planning Commission, tasked with distributing federal aid to townships, determined aid levels based on the proportion of the town that experienced damages, with no regard for the population size of that town. That meant that smaller, wealthier, whiter towns with a high rate of damages received much more aid per household than larger, higher percentage minority towns. One 89 percent white town with 23 affected residents received $49,000 per resident. A town whose population is roughly 50 percent Black and had 92,000 affected residents received only $40 per resident (Capps, 2018).

This anecdotal evidence is born out more generally across time. A study from last year analyzed Panel Study of Income Dynamics data to determine longitudinal trends in recovery after natural disasters, and found significant racial disparities in wealth accumulation. White people living in areas with significant disaster costs gained on average $126,000 from 1999 to 2013, while Black people in these areas actually lost $27,000 over the same period of time. For Latinos and people of other race, the losses are $21,000 and $10,000, respectively. The researchers found similar trends along education and homeownership lines (Howell and Elliott, 2018). They also found that independent from disaster costs, more aid from FEMA increased wealth inequality as well. White people increased their wealth the higher their community's level of aid from FEMA, while for black people as FEMA aid to their community increased, their wealth decreased significantly (Howell and Elliott, 2018).

This paper built on other research covering examples of disaster vulnerability stemming both from damages due to the disaster itself and from aid and recovery policy. In "Places as Recovery Machines: Vulnerability and Neighborhood Change After Major Hurricanes," Pais and Elliott showed that disasters furthered inequalities, and the vulnerability of certain communities could counteract gains in population and housing relative to nearby, less vulnerable communities. In "Loan Request Outcomes in the U.S. Small Business Administration Business Disaster Loan Program," Dahlhamer found disparities among demographic groups in approval for Small Business Administration loans. In "Beyond Disasters: A Longitudinal Analysis of Natural Hazards' Unequal Impacts on Residential Instability," Junia and Howell found that in particular for Black and Latina women, natural disasters increase residential instability.

Dimensions of Disaster Vulnerability

Due to this research, those interested in disaster vulnerability have sought to develop metrics that can measure and compare vulnerability across different communities. There are many examples of disaster vulnerability metrics. In the following section, I will provide overviews of some of the most prominent. Each includes metrics of vulnerability that fall primarily into five categories: socioeconomic status, household composition and disability, minority status and language, housing and transportation, and social capital. There is a fair amount of overlap across different indexes, but the differences are important. For instance, while all four indexes seek to express the potential for losses for different areas in the aftermath of the disaster, not all focus solely on social vulnerability.

The FEMA Preparedness organization uses a set of metrics that focus on resilience. There are some subtle differences between vulnerability and resilience measurements. Resilience is sometimes thought of as the ability to overcome vulnerability, and for this reason social capital metrics are included in resilience measurements. Religious affiliation is included in the Preparedness indicator variables because of the belief that more well-connected communities will be better-equipped to handle adversity. However, researchers creating social vulnerability metrics would be unlikely to include religious affiliation, because less religious areas are not necessarily more vulnerable to disasters. Similar, but opposite reasoning explains the inclusion of race in vulnerability measurements but not in resilience measurements. Part of the difference between vulnerability and resilience is the application of each metric. Social vulnerability indexes help government agencies target certain communities that may need more assistance following a disaster. Resilience metrics are often used as tools for communities to identify their own shortcomings. In this way, the framing of each concept is slightly different (Tierney, 2019).

The following table provides color labels for variables in each list, based on the five broad categories discussed earlier

Category	Color
Socioeconomic status	
Household composition/disability	
Minority status/language	
Housing/transportation	
Social capital	

Table 1: Categories of Metrics for Disaster Vulnerability

Methodologies for Measuring Disaster Vulnerability

POST Demographic Data

Prioritizing Operations Support Tool (POST) was developed by FEMA in an attempt to help prioritize response and recovery operations in the immediate aftermath of a disaster, in certain areas based on three main criteria. The first two criteria are based on the likelihood and severity of the environmental hazard hitting the area, while the third is a social vulnerability index that uses thirteen different community-level demographic statistics taken from the American Community Survey 5-year estimates. These variables are used to rank communities by overall vulnerability level, although it is not yet clear exactly what process will be used to create this rank (Goldblatt, 2019).

POST Social Vulnerability Index
Population age 16+ and unemployed
Population age 16+ not in labor force
Number of households in poverty
Number of households on disability and food stamps
Number of households on disability and no food stamps
Number of households with food stamps/SNAP
Number of households with public assistance
Population age 65 and over
Population American Indian or Alaska Native
Number of housing units that are mobile homes
Number of households
Number of housing units
Population

Table 2: POST Variable List

FEMA Preparedness Data

The FEMA Preparedness Indicator Variables were selected by a research group at Argonne National Laboratory. Their methodology for selection began with a review of six meta-analyses of community resilience assessment, which covered a total of 73 different methodologies. These were reduced to 8 methodologies that met these criteria: "they used a unit of analysis that corresponded to U.S. county-level data, applied to multiple hazards, had a pre-disaster focus, used quantitative measures, used a publicly available methodology, and used publicly available data sources." The 8 selected methodologies included a total of 100 variables. These were reduced to a list of 20 variables which were found in at least 3 separate methodologies. The variables selected cover both individual and community-level data points, and most are found in the American Community Survey.

The primary use of these variables is to "assist in prioritizing locations for Technical Assistance investment and in informing community resilience Technical Assistance content" (Edgemon et. al., 2019).

FEMA Preparedness Indicator Variables
Percent of population without high school diploma
Percent of labor force unemployed
Median household income
Gini Index
Percent of population without health insurance
Percent of population with disabilities
Percent of population 65 years and older
Percent of single-parent households
Percent of with limited English-speaking households
Percent of owner-occupied housing units
Percent of occupied housing units with no vehicles available
Number of hospitals per 10,000 people
Number of health-diagnosing and treating practitioners per 1,000 population
Percent of housing units that are mobile homes
Number of public schools per 5,000 population
Net migration of individuals
Number of hotels/motels/casinos per 5,000 population
Percent of vacant rental housing units
Percent of population that are religious adherents
Number of civic and social organizations per 10,000 people

Table 3: FEMA Preparedness Variable List

CDC Social Vulnerability Index

The Centers for Disease Control and Prevention Social Vulnerability Index was first proposed by Flanagan, et. al., in "A Social Vulnerability Index for Disaster Management." It was developed by ATSDDR's Geospatial Research, Analysis & Services Program to "help public officials and emergency response planners identify and map the communities that will most likely need support before, during, and after a hazardous event." The SVI data is used to rank Census tracts in order based on each variable. Census tracts on the more vulnerable end of the spectrum can be targeted by public health officials for more supplies, aid, or other resources before, during, or after a hazardous event. I use the CDC SVI 2018 documentation for precise calculations for each variable from the American Community Survey data.

CDC Social Vulnerability Index
Percent individuals below poverty
Percent civilian unemployed
Per capita income
Percent persons with no high school diploma
Percent persons 65 years of age or older
Percent persons 17 years of age or younger
Percent persons with a disability
Percent male or female householder, no spouse present, with children under 18
Percent minority
Percent persons 5 years of age or older who speak English less than "well"
Percent multi-unit structure
Percent mobile homes
Crowding
No vehicle available
Percent of persons in group quarters

Table 4: CDC SVI Variable List

HVRI Social Vulnerability Index

The social vulnerability index was first proposed by Cutter, Boruff, and Shirley in "Social Vulnerability to Environmental Hazards." Initially, 42 variables were selected as having an impact on social vulnerability. Since then, the Hazards & Vulnerability Research Institute have modified the original list to include these 29 variables (SoVI Evolution).

HVRI SoVI
Percent poverty
Percent households receiving social security benefits
Percent households earning over $200,000 annually
Per capita income
Percent with less than 12th grade education
Percent civilian unemployment
Median housing value
Median gross rent
Percent population under 5 years or 65 and over
Percent children living in 2-parent families
Median age
Percent female
Percent female headed households
People per unit
Percent Asian
Percent Black
Percent Hispanic
Percent Native American
Percent speaking English as a second language with limited English proficiency
Nursing home residents per capita
Hospitals per capita
Percent of population without health insurance
Percent renters
Percent mobile homes
Percent employment in extractive industries
Percent employment in service industries
Percent female participation in labor force
Percent of housing units with no car
Percent unoccupied housing units

Table 5: HVRI SoVI Variable List

Research Questions

In the present study, there are a few main research questions I will seek to answer. First, how do the experiences of survivors seeking aid in FEMA's Individual Assistance (IA) programs vary? In which dimensions, among numbers of contacts, amount of assistance, application rate, and approval rate, does the experience vary most? Second, what individual or community-level demographic factors from IA, POST, FEMA Preparedness, the CDC SVI, or the HVRI SoVI impact the experience of survivors in IA programs? Finally, how should FEMA use demographic information at the individual and community levels to better manage IA programs in the future? Answering these questions require definition of the units of analysis.

Units of Analysis

There are two main levels of data used in this analysis—the individual level and the community level. The initial data set received from FEMA was comprised of data about every registrant for various IA programs from eight selected natural disasters. These disasters include Hurricanes Irma, Matthew, Maria, Harvey, and Ike, as well as the 2014 California earthquake, the 2017 California wildfires, and the 2016 Louisiana floods. For each disaster, FEMA provided information about individual survivors of these disasters who applied to FEMA directly for aid. The IA data is reported at the individual household level. One unit will therefore be the individual—comparing individual demographic characteristics that we know, such as age, gross income, and household composition, with outcomes in Individual Assistance programs. The other unit of analysis is the community. The highest geographic spatial resolution in the IA data is the reported ZIP code. Merging ZIP code-level data with the IA data allows us to compare outcomes to the community-level demographic statistics. However, many of the ZIP codes are poorly represented by registrants.

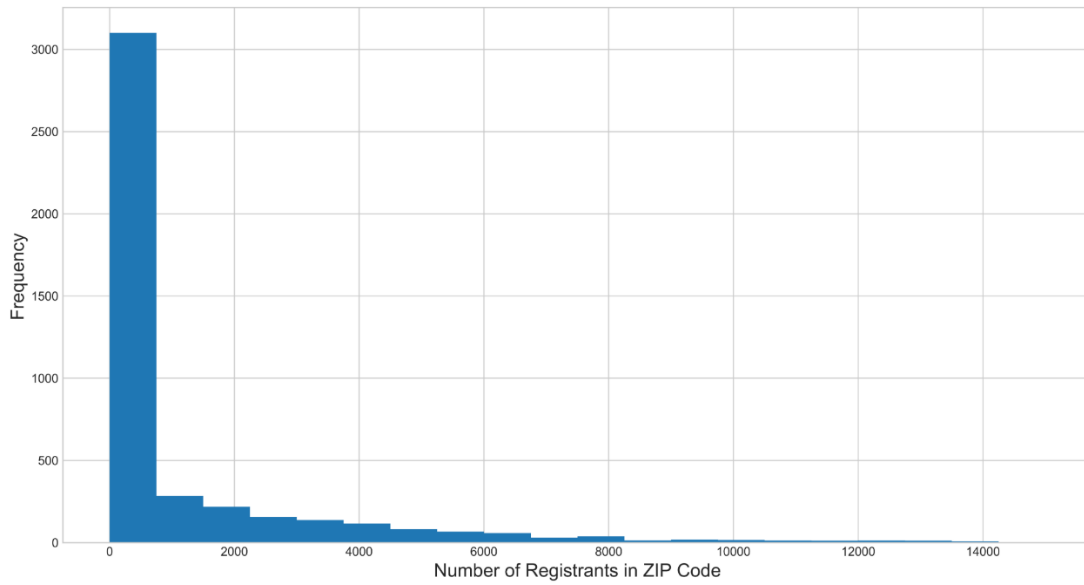


Figure 1: Histogram of Registrants Per ZIP Code

Figure 1 shows that most ZIP codes have very few registrants applying. There are 4424 ZIP codes across all 8 disasters, covering over 5.6 million registrants. But 25% of the ZIP codes have only 1 registrant. To avoid selection bias in the aggregated ZIP code data, I will select a cutoff for representation in ZIP codes.

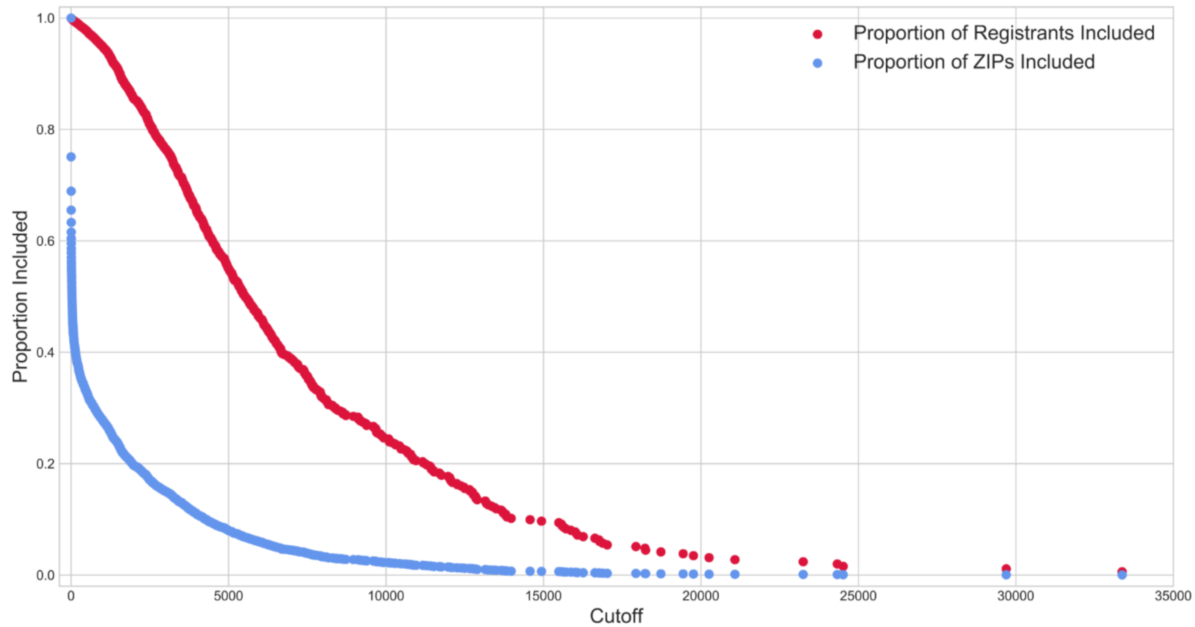


Figure 2: Comparing Impact of Cutoff Selection on Proportion of ZIP Codes and Registrants Included

Figure 2 shows how selecting different cutoffs vary proportion of ZIP codes and registrants included in the analysis. Increasing the cutoff drastically reduces the number of ZIP codes included, but does not drastically reduce the number of registrants included. Setting the cutoff at 30 registrants per ZIP codes includes 50% of the ZIP codes, but covers 99.8% of the registrants because of the high-population ZIP codes.

One methodology for evaluating cutoff selection is to compare the Census data at the ZIP code level the with the aggregated individual data self-reported in the IA data. This chart shows how the proportion of the population age 65 or older as determined by the Census relates to the proportion of IA registrants age 65 or older, for different cutoffs of ZIP code populations.

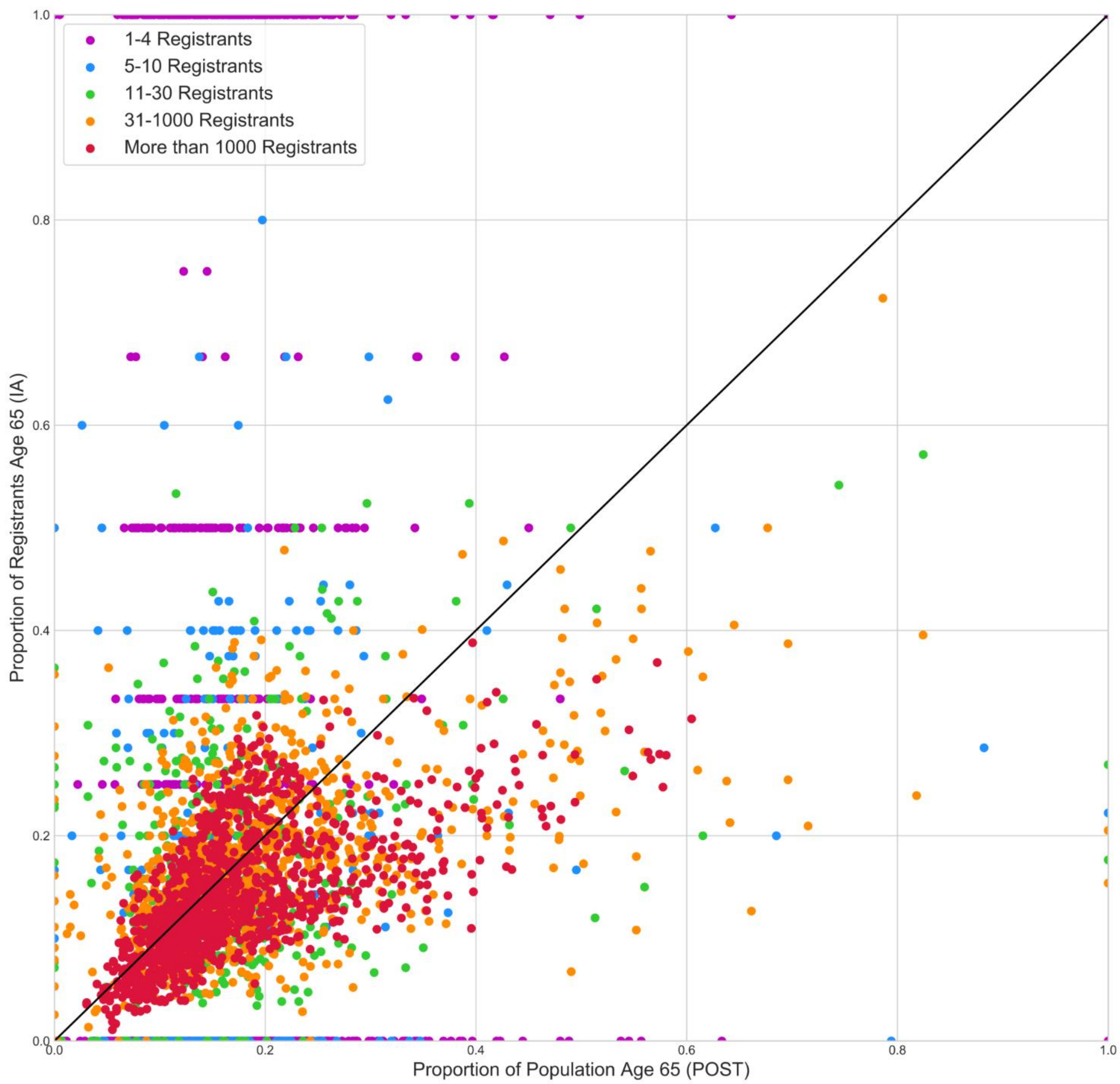


Figure 3: Evaluating Accuracy of Cutoff Selections

Figure 3 shows that choosing a higher cutoff yields better results, with diminishing returns. The proportion of IA registrants age 65 or older in ZIP codes with 11 to 30 total registrants related to the Census estimate in those areas in a similar way when comparing the proportion of IA registrants age 65 or older in ZIP codes with 31 to 1000 total registrants. This reveals a few important insights. The aggregate demographics of well-represented ZIP codes resemble demographic estimates from the Census. Therefore, including more well-represented ZIP codes in the analysis better justifies the comparison of community-level characteristics to individual-level outcomes.

Research Design

The two broad fields of analysis deployed in this research are descriptive statistics and regression analysis. Descriptive statistics provide an overview of the data, and clue us in to broad trends that may exist. It also helps to narrow down which specific distinctions, within demographics or the disasters themselves, may correlate with disparities in assistance outcomes. Regression analysis shows much more robustly the relationship between the independent and dependent variables.

Chapter 2: Data

There are two main levels of data used in this analysis—the household level and the community level. The initial data set received from FEMA was comprised of data about every registrant for various Individual Assistance programs from eight selected natural disasters. These disasters include Hurricanes Irma, Matthew, Maria, Harvey, and Ike, as well as the 2014 California earthquake, the 2017 California wildfires, and the 2016 Louisiana floods. For each disaster, FEMA provided information about individual survivors of these disasters who applied to FEMA directly for aid. One important note is that while these programs are under the umbrella of Individual Assistance, registrants apply for aid on behalf of their household, which may comprise other survivors. This incongruity is reflected in the data. For instance, self-reported demographic information includes age, gross income, number of people in the household, homeowner's insurance status, and their city and ZIP Code. Age and income are expected to be individual-level data points, while clearly the number of people in the household depends on household-level information. The other data cover different aspects of the survivor journey, including every time there was a phone contact between the survivor and FEMA and the general subject of that call, how many hotels they stayed in and for how long, and every form of assistance the survivor applied for. Some programs are simple loan programs and provided is an amount that was approved for provided in the data, and an approval date. Other programs like the Direct Housing program have much more information including dates for every point of the application and approval process, which can total up to 10 different dates. It is from the FEMA Individual Assistance data set that the dependent variables for the analysis, as well as some independent variables, are drawn. Figure 4 displays the disasters provided, along with their four-digit DR code, the number of registrants in the data set, and the year that the disaster occurred.

Disaster	DR	Number of Registrants	Year
Hurricane Ike	1791	734,386	2008
California Earthquake	4193	7,399	2014
Louisiana Flooding	4277	153,408	2016
Hurricane Matthew	4285	81,981	2016
Hurricane Harvey	4332	895,617	2017
Hurricane Irma	4337	2,644,443	2017
Hurricane Maria	4339	1,122,545	2017
California Wildfires	4344	25,425	2017

Table 6: Disasters Provided

Dependent Variables

The dependent variables for this analysis are comprised of outcomes for registrants in FEMA's IA programs. There are many data points in the IA dataset that fit this criterion. We attempted to address four outcome variables in this analysis, two that are binary, and two that are continuous. The two binary dependent variables include a measure of whether the registrant applied for general assistance, and of those who applied for general assistance, whether the applicant was approved. For the applicants who were approved for assistance, we included their assistance amount in dollars as a third dependent variable. Finally, we included the number of phone calls between the registrant and FEMA throughout the assistance process as a continuous dependent variable. Again, this is not an exhaustive list of all the potential outcome variables that can be used. However, these four variables were the least complex to understand and predict at this stage of the study. These variables also allow the inclusion of the maximum number of registrants. Outcomes for specific programs would limit the number of data points significantly, but could be used in future studies.

FEMA IA Outcomes
Applied for assistance (binary)
Received assistance (binary)
Assistance amount ($)
Number of phone calls to/from FEMA

Table 7: Dependent Variables List

Independent Variables

Household-Level Variables

The FEMA IA data included self-reported demographic characteristics for each registrant and their household. This included information about whether the registrant was a renter or homeowner, whether or not the registrant had homeowner's insurance, and where the registrant lived down to the city and ZIP code level. It also included the age, yearly gross income level, and the number of people living in the registrant's household. Finally, it contained the registration method and the initial application date for individual assistance. Five of these individual-level demographic characteristics from the IA data will be used as independent variables in this analysis. This included all data points except for the registrant's location. Again, while this data is described as being at the household level, some characteristics such as age are at the individual-level.

Demographic Data from FEMA IA
Age
Homeowner or renter
Gross income
Number of people in household
Homeowner insurance status

Table 8: Demographic Data from FEMA IA

Community-Level Variables

In this analysis we are also including independent variables at the community level. The highest geographic spatial resolution in the Individual Assistance data is the reported ZIP code. Data at the ZIP code level from the American Community Survey (2013-2017 5-Year estimate) was merged with the Individual Assistance data. Here, I selected variables by developing my own social vulnerability index based on the literature.

From the FEMA Preparedness Indicator Variables, the CDC SVI, and the HVRI SoVI, I selected several variables that may reflect community vulnerability, and will be used as independent variables in the modeling phase. The methodology for selecting these variables was to take the CDC SVI and the FEMA Preparedness Indicator Variables and combine them, while removing all variables where data was not available at the ZIP code level. This included data on migration, religious affiliation, public schools per capita, hospitals per capita, and civic organizations. From the HVRI SoVI I included four variables describing the racial demographics of the area in greater detail. The final list includes 23 variables covering socioeconomic status, household composition and disability, minority status and language proficiency, housing status, and transportation.

Proposed Community Variables
Percent individuals below poverty
Percent of labor force unemployed
Per capita income
Percent of population without high school diploma
Percent of population without health insurance
Gini Index
Percent of population 65 years and older
Percent of population 17 years of age or younger
Percent of population with a disability
Percent male or female householder, no spouse present, with children under 18
Percent minority
Percent Asian
Percent Black
Percent Hispanic
Percent Native American
Percent of limited English-speaking households
Percent of owner-occupied housing units
Percent mobile homes
Percent of occupied housing units with no vehicles available
Percent multi-unit structure

Figure 4: Proposed Variable List

The following table describes the relationship between the chosen community variables and the various existing variable lists.

Proposed Community Variables	POST SVI	Preparedness Indicators	CDC SVI	HVRI SoVI
Poverty	X		X	X
Unemployment	X	X	X	X
Income		X	X	X
Education Level		X	X	X
Health Insurance		X		X
Gini Index		X		
Age 65 and Older	X	X	X	X
Age 17 and Younger			X	
Disability	X	X	X	
Single-Parent Household		X	X	X
Minority Status			X	
Percent Asian				X
Percent Black				X
Percent Hispanic				X
Percent Native American	X			X
English Language Proficiency		X	X	X
Owner/Renter		X		X
Mobile Homes	X	X	X	X
Vehicle Access		X	X	X
Multi-Unit Structure			X	
Group Quarters			X	
Crowding			X	
Housing Vacancy		X		

Figure 5: Comparing Social Vulnerability Variable Lists

One note here is that not every variable is measured the exact same way across vulnerability indexes. For instance, the HVRI SoVI does not include single-parent households where the parent is male. This table does not take into account these subtle differences, but it shows the origin of each chosen variable and the major differences between vulnerability metrics.

Descriptive Statistics

Variable Names for Modeling	Precise Variable Description
pov	"Percentage of persons below poverty estimate"
unemp	"Estimate of unemployed population/civilian population age 16+ in labor force"
income	"Per capita income estimate"
nohsdip	"Percentage of persons with no high school diploma (25+) estimate"
nohinsur	"Percentage of the population without health insurance coverage"
gini	Commonly used measure of income inequality, ranging from 0 (low inequality) to 1
age65	"Percentage of persons aged 65 and older estimate"
age17	"Persons aged 17 and younger estimate/ total population estimate"
disab	"Percentage of civilian noninstitutionalized population with a disability estimate"
sph	"Single parent household with children under 18 estimate/households estimate"
minority	"Minority estimate/total population estimate"
asian	"Asian estimate/total population estimate"
black	"Black estimate/total population estimate"
hispanic	"Hispanic estimate/total population estimate"
natam	"Native American estimate/total population estimate"
limeng	"Percentage of limited English-speaking households"
owner	"Percentage of owner-occupied housing units"
mobile	"Percentage of mobile homes estimate"
novehic	"Percentage of households with no vehicle available estimate"
mustruc	"Housing in structures with 10 or more units estimate/housing units estimate"
groupq	"Persons in group quarters estimate/total population estimate"
crowd	"Occupied housing units with more people than rooms estimate/occupied housing units estimate"
vacant	"Percentage of vacant rental housing units"
RentOwn	Homeowner or renter (binary)
Age	Age of individual registrant (continuous)
GrossIncome	Gross income of individual registrant (continuous)
HOI	Registrant has homeowner's insurance or does not (binary)
HHComp	Household composition, or number of people in household (continuous)

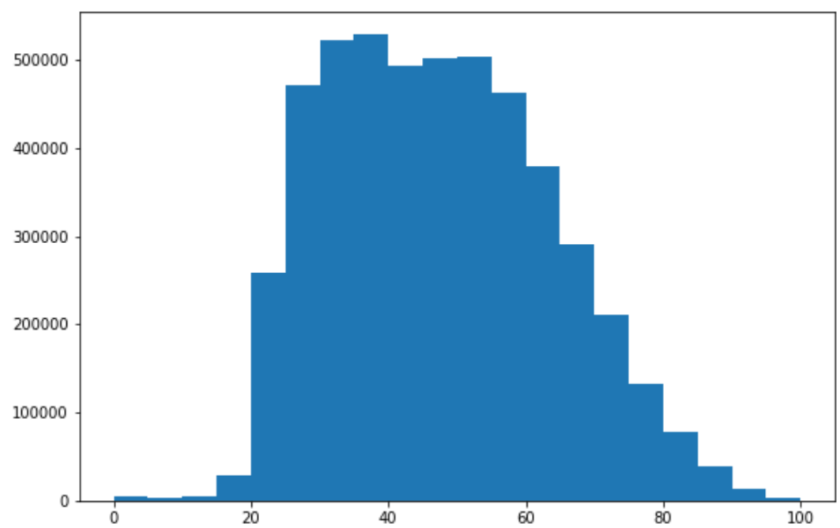
Figure 6: Variable Names and Precise Descriptions

Variable	Class	%	Mean	St. Dev.	Count (N)
Individual Variables					
RentOwn	Owner	53.01			4913452
HOI	Has Insurance	27.97			4930546
Age			46.02	16.08	4930546
HHComp			2.22	1.45	4930546
GrossIncome			48908.58	487405.81	4930546
Community Variables					
pov			0.18	0.11	3511
unemp			0.08	0.05	3511
income			28568.88	15148.56	3511
nohsdip			0.16	0.10	3511
nohinsur			0.14	0.07	3511
gini			0.44	0.07	3511
age65			0.17	0.09	3511
age17			0.22	0.07	3511
disab			0.15	0.06	3511
sph			0.09	0.06	3511
minority			0.45	0.28	3511
black			0.16	0.19	3511
asian			0.03	0.06	3511
hispanic			0.24	0.26	3511
natam			0.01	0.04	3511
limeng			0.07	0.14	3511
owner			0.66	0.18	3511
mobile			0.13	0.15	3511
novehic			0.11	0.05	3511
mustruc			0.10	0.15	3511
groupq			0.03	0.07	3511
crowd			0.04	0.04	3511
vacant			0.02	0.03	3511

Figure 7: Descriptive Statistics Table

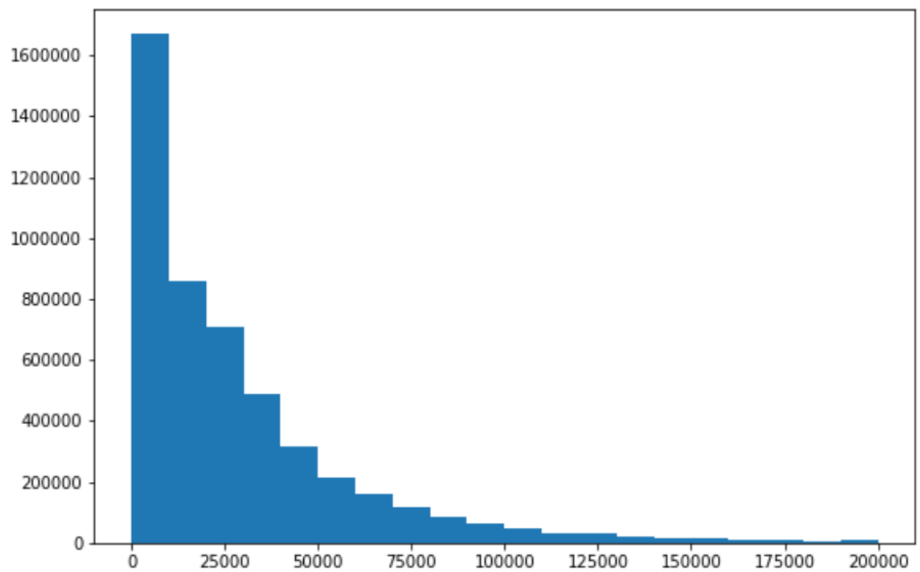
This summary table covers the key descriptive statistics for each independent variable, both at the individual and at the community level. DR 1791, Hurricane Ike, was removed from the data set as the disaster occurred in 2008, out of the range of the timeframe for the community-level variables. For binary variables, the class and proportion are displayed. For non-binary variables, the mean and standard deviation are displayed. Each variable has a count as well. The complete data set for the seven disasters has 4930546 registrants. Some registrants did not report whether they were a homeowner or renter, hence the disparity in count for that variable. Across disasters, 3511 ZIP codes are represented. One notable aspect of this summary table is the high standard deviation of the self-reported gross income, indicating that there are many outliers for this data point. Additionally, individual and community-level data points for homeownership and income can be directly compared. At the community-level, there is a higher proportion of homeowners than those who applied for assistance. Conversely, the community-level per-capita income is much lower than the average income reported by registrants, although this number is highly skewed by outliers, based on the high standard deviation.

Age



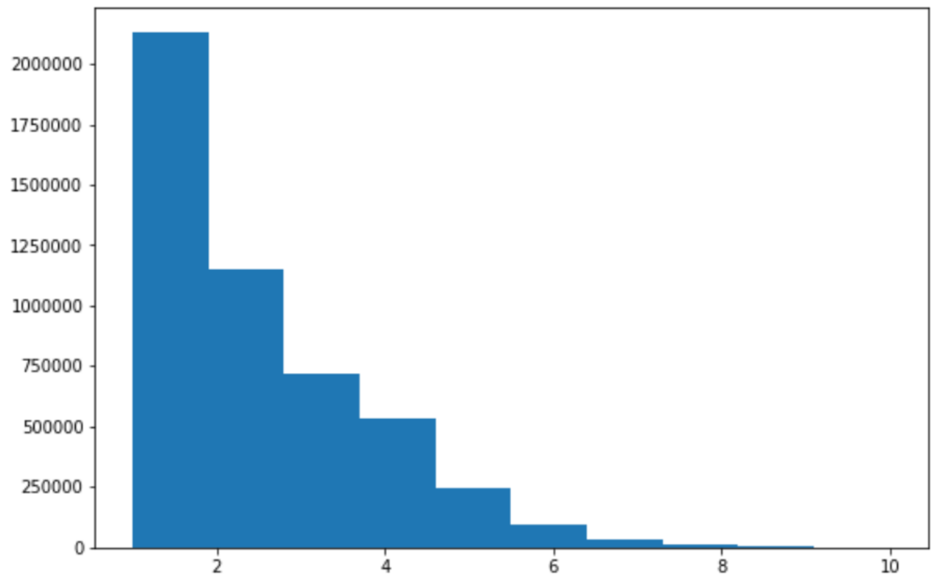
This histogram displays the distribution of ages across all registrants. Around 16% of the registrants reported an age below 18. These ages are uniformly distributed. The age data point will be removed from the dataset because of the high number of anomalies, and the difficulty in interpreting the model coefficient.

Gross Income



The self-reported gross income histogram shows that most of the registrants make less than $100,000 per year. There is a long tail on this histogram which extends much greater than $200,000. Again, the standard deviation for self-reported income is in the hundreds of thousands, much greater than the average income, meaning that there are many outliers which need to be dealt with. Moving forward, I chose $200,000 as the cutoff for individual income for registrants.

Household Composition



The histogram for household composition resembles that of self-reported gross income, in that the majority of households have a low number of residents, with a long tail at the far end of the spectrum for the number of people in the registrant's household. In order to improve computation time and likelihood of model convergence, registrants with a household composition higher than 10 were removed from the data set.

Chapter 3: Methodology

Data Preparation

The first step in model development is the preparation of the data. Because one of the eight disasters, Hurricane Ike, occurred in 2008, much earlier than the other disasters and out of the range of the 2013-2017 American Community Survey 5-Year Estimate data tables, data points from this disaster were removed. After this removal, several dummy variables were added to the models corresponding the specific disaster for which each registrant or community was applying for assistance. The main reason for this was that there were significant disparities in outcomes across disasters, and we want to account for those broad differences in the analysis. Only six dummy variables are added as the seventh is redundant. Roughly 0.3% of registrants had missing individual-level ownership status data and 0.4% of registrants had missing ACS data, so these rows were removed.

Model Development Process

Multilevel Modeling

The modeling approach selected for this analysis is multilevel modeling. There are many advantages to taking a multilevel modeling approach over other methods such as ordinary least-squares regression, but the simplest explanation for its use in this analysis is that the data we are studying has two levels—the individual (or household) level and the community level. In this study, we want to not only explore the relationship between individual-level characteristics and FEMA IA outcomes, but also the relationship between community-level characteristics and FEMA IA outcomes. Multilevel modeling provides us a framework to explore multilevel relationships. When the data being analyzed is in a multilevel structure, a key assumption of traditional regression modeling approaches is violated. The observations can no longer be assumed to be independent of one another, because the data is now clustered within the higher level, in our case, the community level. Disregarding this clustering effect will result in the underestimation of standard errors. Another advantage is the ability to describe the distribution in variance between levels.

The main idea behind multilevel models is to add random effects to the intercept, slope, or both for the model, at the higher level. In multilevel modeling, fixed effects are the part of the model for which the relationship with the dependent variable is fixed for all clusters. The random effects are allowed to vary between clusters. In the simplest example for this analysis, a random effect in the form of a random intercept can be added at the community level, which keeps the model coefficients fixed but allows the intercept to vary from community to community. The added random effects are determined by the modeler based on the theory the model is based on (Finch, et. al., 2014; Hox and Maas, 2005).

In the next few subsections, I will describe the modeling process for multilevel modeling. Many of the assumptions are similar to those in traditional regression modeling approaches, including first checking the collinearity of the independent variables.

Checking Collinearity

Collinearity occurs when two or more independent variables correlate with each other, which makes it difficult to determine which has an effect on the dependent variable. When variables are collinear, variables should be removed from the model until the collinearity reaches an acceptable threshold. Collinearity should be checked before any modeling takes place.

Collinearity may not be critical for a good model in cases where the main goal is prediction. In these cases, statistically significant but collinear variables may be included. However, in this case a major goal is testing the specific variables chosen by FEMA against the dependent variable. We want to be able to interpret the relative predictive power of each variable through its coefficient in the model, and collinearity would inhibit our ability to draw specific conclusions. Therefore, we will test for collinearity in our variables before proceeding.

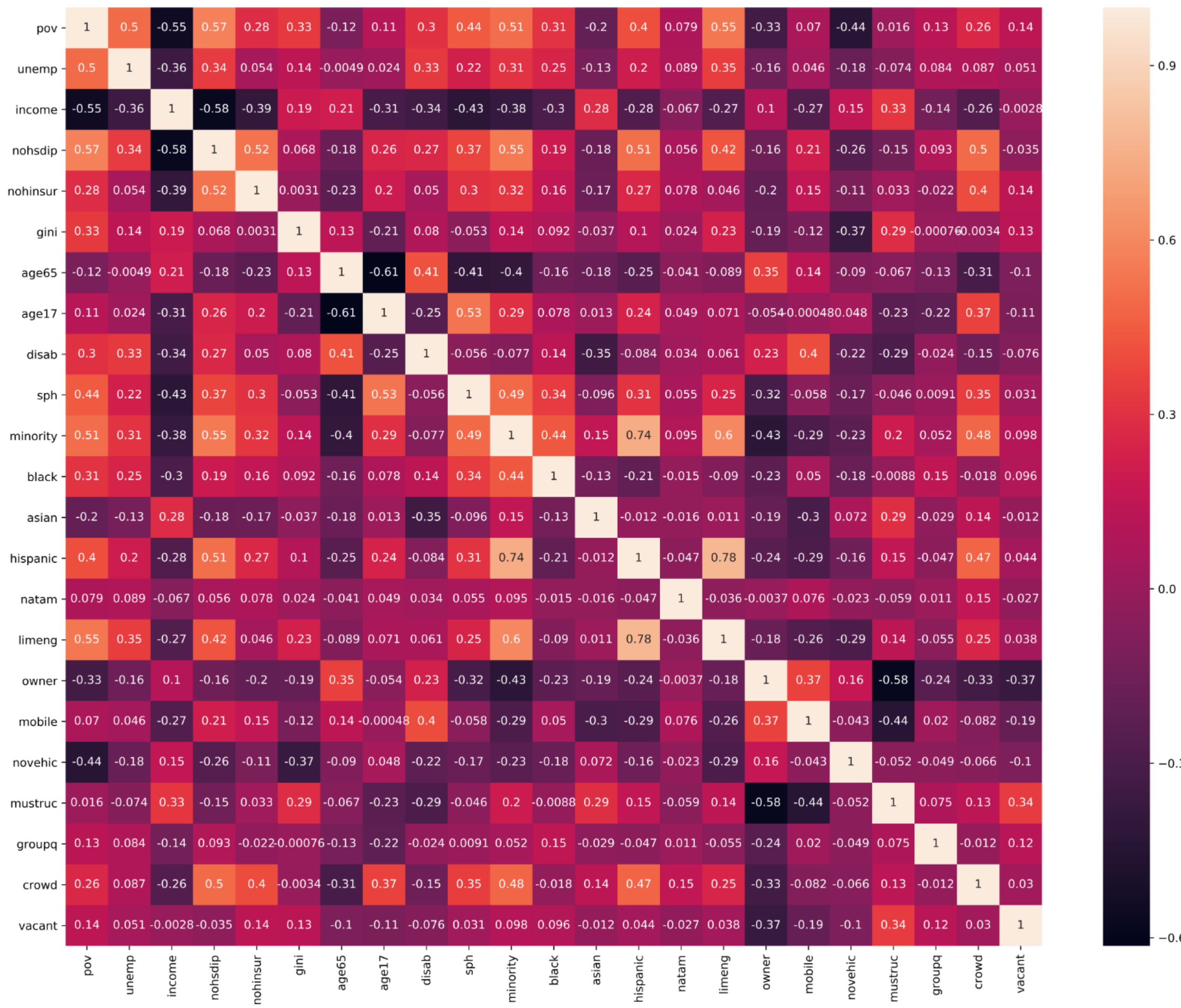


Figure 8: Collinearity Matrix

Here, the only variables with moderately high cross-correlation are the proportion Hispanic population, proportion minority population, and proportion of the population that speaks English less than "well." I removed the proportion Hispanic population variable in order to avoid these cross-correlation issues.

Null Model

In multilevel modeling, often the next step after checking multicollinearity is to create a null model. In this model, the dependent variable is only predicted by a random intercept at the higher level. This random effect allows the intercept of the model to vary by the specific cluster. The output of this model that important to note is the variance components at each level, both within clusters and between clusters. This allows us to calculate the proportion of the variance at each level. If there is a substantial portion of the variance at the between-cluster level, the use of multilevel modeling is warranted (Guo and Zhao, 2000).

Intermediate Variable Selection

The methodology we used for the intermediate variable selection phase is the stepwise regression technique. In this technique, first each independent variable is included in a single linear regression model with the response variable. The variable with the lowest p-value below 0.05 is selected. Then new models with the chosen variable and every other independent variable are run, and the new variable with the lowest p-value is chosen. This process continues, selecting variables and then running with every other variable, until no added variable has a p-value of below 0.05. The variables already chosen in previous steps are selected for the model.

Testing Residuals

In traditional regression modeling, the residuals are assumed to be normally distributed, centered at 0, and with constant variance. With linear modeling, the residuals are by default centered at 0, so this assumption does not need to be checked. What we want is homoscedasticity, where the variance is constant. This can be checked by plotting the residuals with the explanatory variables and with the fitted values. We are looking for an oval shape in the plots, where the scatter points are evenly distributed, with no obvious clusters. The last assumption to check with residuals is that they are normally distributed. This can be checked with a histogram of the residuals or a normal probability plot.

In multilevel modeling, there are some modifications to this list of assumptions. First is that the homoscedasticity is no longer expected due to the random effects. However, the variance should be constant at the between-cluster level. Furthermore, some added assumptions are that the residuals of the random effects in the model are also assumed to have a normal distribution and constant variance (Guo and Zhao, 2000).

Model Validation

Once completing the final model and all assumptions have been met, we should validate the model. One technique is cross-validation. Cross-validation involves randomly splitting the data into 75% and 25% sections, developing a model using the 75% section, and then testing its predictive power on the 25% section.

Chapter 4: Analysis and Results

Amount of Assistance Received

When modeling the amount of assistance received, I reduced the data set to just those registrants who were approved for general assistance. This reduced the number of registrants to 1,435,476. Whether an applicant was approved for assistance is modeled separately. Additionally, self-reported income and number of people in the household were scaled to be between 0 and 1. We can check collinearity for the individual-level demographic variables with a correlation matrix.

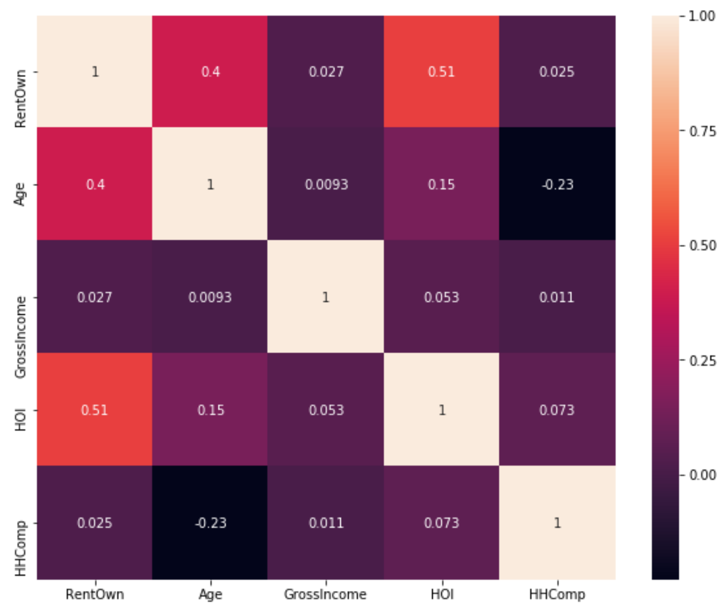


Figure 9:Individual-Level Variable Correlation Matrix

None of the variables are highly correlated with each other, so all individual demographic variables will remain in the model. The next check is the null model. Setting the random effect as an intercept that varies with the ZIP code, the proportion of variance at the community level is 31%. This is substantial and shows a high clustering effect in the amount of assistance received at the ZIP code level, necessitating the use of multilevel modeling.

To start, I run a multilevel model with all independent variables included as fixed effects along with the disaster designation, meaning that I expect these variables to have a fixed relationship with amount of assistance received at the individual level regardless of ZIP code. I add a random effect in which I allow the intercept of the model to vary based on ZIP code. Below is the model output. The models in this section were created using the lmer package in R.

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	Astn_Tot
No. Observations:	1700538	Method:	REML
No. Groups:	1517	Scale:	20618436.0549
Min. group size:	30	Likelihood:	-16735791.6052
Max. group size:	13390	Converged:	Yes
Mean group size:	1121.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	3053.415	853.558	3.577	0.000	1380.472	4726.357
RentOwn	1988.091	8.939	222.404	0.000	1970.571	2005.612
GrossIncome	-258.581	13.332	-19.396	0.000	-284.711	-232.452
HOI	-572.865	10.890	-52.603	0.000	-594.210	-551.520
HHComp	2594.783	24.071	107.799	0.000	2547.606	2641.960
DR_4193	-486.793	249.688	-1.950	0.051	-976.172	2.585
DR_4277	982.409	436.910	2.249	0.025	126.081	1838.736
DR_4285	-1750.462	434.092	-4.032	0.000	-2601.268	-899.657
DR_4332	-1156.428	408.146	-2.833	0.005	-1956.379	-356.478
DR_4339	-3543.445	628.134	-5.641	0.000	-4774.565	-2312.326
DR_4337	-3369.986	403.311	-8.356	0.000	-4160.461	-2579.510
pov	618.476	854.065	0.724	0.469	-1055.461	2292.412
unemp	127.643	1029.394	0.124	0.901	-1889.933	2145.219
income	1826.146	527.707	3.461	0.001	791.860	2860.433
nohsdip	-2365.675	984.625	-2.403	0.016	-4295.505	-435.844
nohinsur	3810.072	1098.178	3.469	0.001	1657.684	5962.461
gini	-2385.855	1019.368	-2.341	0.019	-4383.779	-387.931
age65	-872.666	785.152	-1.111	0.266	-2411.536	666.204
age17	2155.222	1386.455	1.554	0.120	-562.180	4872.625
disab	-686.316	1169.609	-0.587	0.557	-2978.708	1606.076
sph	-2288.765	1395.640	-1.640	0.101	-5024.169	446.640
minority	-2031.831	377.880	-5.377	0.000	-2772.462	-1291.200
asian	-3644.122	1319.976	-2.761	0.006	-6231.227	-1057.016
black	2573.199	401.666	6.406	0.000	1785.949	3360.449
natam	-411.978	1115.955	-0.369	0.712	-2599.211	1775.254
limeng	5339.052	858.324	6.220	0.000	3656.767	7021.336
owner	881.660	524.254	1.682	0.093	-145.860	1909.179
mobile	809.934	393.807	2.057	0.040	38.086	1581.782
novehic	153.203	1404.266	0.109	0.913	-2599.109	2905.514
mustruc	182.845	421.268	0.434	0.664	-642.826	1008.515
groupq	-795.995	869.814	-0.915	0.360	-2500.798	908.809
crowd	-2901.665	2141.173	-1.355	0.175	-7098.287	1294.956
vacant	3225.260	2181.788	1.478	0.139	-1050.966	7501.486
Group Var	2242744.775	19.018				

Figure 10: Initial Model for Assistance Amount

There are several notable aspects of this initial model. One finding is that each individual-level variable—homeownership, homeowner's insurance status, income, and household composition, all have a significant relationship with total assistance. Just being a homeowner versus being a renter meant registrants were receiving on average close to $1988 more than their counterparts. However, insurance status had a strong, negative correlation with assistance level. Household income level had a negative, but lesser relationship with assistance level. The number of people in the household had a positive relationship with assistance level. Holding all else equal each extra person in the household, registrants could expect to receive an extra $259.

Of the community-level variables, several were non-significant. This included the proportion of households below poverty, the unemployment rate, proportion of the population at least 65, proportion of the population below 18, proportion of population with a disability, proportion of households that are single-parent, Native American population, percentage of households without access to vehicle, proportion of housing in multi-unit structures, percentage of population in group quarters, crowding rate, and vacancy rate. Other variables, including average income, education level, health insurance status, income inequality, minority status, proportion of population Asian and Black, English-language proficiency, and proportion of housing that are mobile homes, were significant in this model. Of the significant variables, many of the highest-magnitude coefficients are in the category of minority status and language proficiency. Registrants living in areas with higher proportion Black and low English-proficiency were likely to receive more assistance, while high proportion Asian communities and Native American communities received less. This could be mainly due to which areas were affected by natural disasters in this dataset. Interestingly, community minority status taken as an aggregate had a negative relationship with assistance level. Except for poverty level, each socioeconomic status variable was significant in this model. Registrants living in areas with lower education level were likely to receive less assistance than their counterparts. Furthermore, registrants living in areas with higher income inequality were likely to receive less assistance than their counterparts. Conversely, registrants living in areas where a higher proportion of the population had no health insurance coverage received more assistance.

Two interesting comparison points are the individual and community-level income and homeownership variables. Individual-level self-reported income had a significant but miniscule relationship with assistance level. However, the community-level per capita income had a much stronger relationship with assistance level. Due to the scaling of this independent variable, registrants living in areas with a per capita income of $10,000 more than other areas could expect to receive an extra $200 in assistance. Here, the community-level variable had a much stronger effect than the individual-level variable, indicating that the wealth of the area a registrant lives is more important in the assistance amount than the wealth reported by the individual registrant.

Conversely, homeownership status at the community level seems to have a diminished relationship with assistance level in comparison to individual homeownership status. The coefficient for the proportion of homes owned by their resident in an area was less than half of the coefficient for individual homeownership status, and the variable was less significant in the model.

Finally, the random effect in this model is only the intercept, which varies at the community level. The relatively high standard deviation indicates that there is high variance in assistance level from community to community.

```
            Mixed Linear Model Regression Results
==================================================================
Model:              MixedLM   Dependent Variable:   Astn_Tot
No. Observations:   1700538   Method:               REML
No. Groups:         1517      Scale:                20618426.6198
Min. group size:    30        Likelihood:           -16735881.1358
Max. group size:    13390     Converged:            Yes
Mean group size:    1121.0
------------------------------------------------------------------
               Coef.    Std.Err.    z    P>|z|   [0.025    0.975]
------------------------------------------------------------------
Intercept      3278.037  577.430   5.677 0.000  2146.294  4409.779
RentOwn        1988.176    8.938 222.431 0.000  1970.657  2005.695
GrossIncome    -258.578   13.331 -19.396 0.000  -284.707  -232.449
HOI            -572.874   10.890 -52.605 0.000  -594.219  -551.530
HHComp         2594.867   24.070 107.804 0.000  2547.690  2642.044
DR_4193        -484.354  249.590  -1.941 0.052  -973.543     4.834
DR_4277        1217.109  421.347   2.889 0.004   391.284  2042.935
DR_4285       -1546.513  418.597  -3.695 0.000 -2366.949  -726.077
DR_4332        -979.863  394.115  -2.486 0.013 -1752.314  -207.412
DR_4339       -3134.641  584.900  -5.359 0.000 -4281.023 -1988.259
DR_4337       -3194.015  390.048  -8.189 0.000 -3958.495 -2429.535
income         1910.633  455.336   4.196 0.000  1018.190  2803.076
nohsdip       -2965.254  826.689  -3.587 0.000 -4585.535 -1344.973
nohinsur       3693.977  957.632   3.857 0.000  1817.053  5570.900
gini          -2536.716  838.468  -3.025 0.002 -4180.084  -893.349
age17          3263.130  950.668   3.432 0.001  1399.855  5126.404
sph           -2868.265 1327.202  -2.161 0.031 -5469.533  -266.996
minority      -2117.591  341.767  -6.196 0.000 -2787.442 -1447.739
asian         -3379.920 1274.465  -2.652 0.008 -5877.825  -882.016
black          2662.854  365.740   7.281 0.000  1946.016  3379.692
limeng         5502.728  773.955   7.110 0.000  3985.805  7019.652
mobile          883.979  370.668   2.385 0.017   157.484  1610.474
Group Var   2239091.308   18.927
==================================================================
```

Figure 11: Stepwise Model for Assistance Amount

The next step was to remove variables which were insignificant, using a stepwise approach. Excluding the random effect and the dummy variables for each disaster, the independent variable with the highest p-value was removed. Then for each iteration, the independent variable with the next highest p-value was removed until no variable had a p-value above 0.05. The resulting model included each individual variable, and 11 of the 22 community-level variables, in addition to the dummy variables and the random effect. Using an ANOVA test, I found that there was no significant difference in the fit of the intermediate model from the initial model, meaning that removing these variables did not significantly reduce the explanatory power of the model. The new model helps with reducing the computation intensity of running further models with more random effects.

```
                  Mixed Linear Model Regression Results
=================================================================================
Model:                  MixedLM      Dependent Variable:       Astn_Tot
No. Observations:       1700538      Method:                   REML
No. Groups:             1517         Scale:                    19079606.8622
Min. group size:        30           Likelihood:               -16674047.1533
Max. group size:        13390        Converged:                Yes
Mean group size:        1121.0
---------------------------------------------------------------------------------
                               Coef.     Std.Err.    z    P>|z|   [0.025    0.975]
---------------------------------------------------------------------------------
Intercept                    2131.492  229.173   9.301 0.000  1682.321  2580.663
RentOwn                      1657.347   55.994  29.599 0.000  1547.601  1767.093
GrossIncome                  -358.374   30.061 -11.922 0.000  -417.292  -299.456
HOI                          -607.801   45.979 -13.219 0.000  -697.918  -517.684
HHComp                       1895.842   69.834  27.148 0.000  1758.970  2032.714
DR_4193                      -564.163  223.366  -2.526 0.012 -1001.952  -126.374
DR_4277                       790.316  186.924   4.228 0.000   423.952  1156.680
DR_4285                      -199.272  190.825  -1.044 0.296  -573.283   174.739
DR_4332                     -1234.964  176.820  -6.984 0.000 -1581.525  -888.403
DR_4339                     -1680.194  209.303  -8.028 0.000 -2090.421 -1269.967
DR_4337                     -1243.548  175.462  -7.087 0.000 -1587.447  -899.649
income                       1117.691  172.569   6.477 0.000   779.462  1455.919
nohsdip                      -247.445  270.243  -0.916 0.360  -777.112   282.221
nohinsur                     1699.179  317.980   5.344 0.000  1075.950  2322.408
gini                         -813.774  272.905  -2.982 0.003 -1348.658  -278.889
age17                         270.339  353.182   0.765 0.444  -421.885   962.563
sph                          -770.403  469.335  -1.641 0.101 -1690.283   149.477
minority                     -237.891  102.665  -2.317 0.020  -439.109   -36.672
asian                       -1317.333  378.808  -3.478 0.001 -2059.782  -574.883
black                         364.166  103.236   3.528 0.000   161.828   566.504
limeng                        895.693  219.976   4.072 0.000   464.549  1326.837
mobile                       -358.940  134.093  -2.677 0.007  -621.758   -96.122
Group Var                  367201.832    5.051
Group x RentOwn Cov        707160.180   11.459
RentOwn Var               4422126.841   39.338
Group x HOI Cov           -662958.626    9.667
RentOwn x HOI Cov        -1608036.511   25.856
HOI Var                   2702495.754   27.661
Group x HHComp Cov         -90721.268   13.009
RentOwn x HHComp Cov      3018148.835   38.859
HOI x HHComp Cov            68610.222   28.638
HHComp Var                5384589.196   55.614
Group x GrossIncome Cov   -328773.968    5.780
RentOwn x GrossIncome Cov -838474.187   15.009
HOI x GrossIncome Cov      727426.268   12.363
HHComp x GrossIncome Cov  -765858.521   17.183
GrossIncome Var            822386.676   10.101
=================================================================================
```

Figure 12: Modelling Assistance Amount with Random Effects

In this model, random coefficients for each individual variable were added at the community level in addition to the random intercept. The main goal of adding random coefficients is to examine how the relationship between the independent variable and the dependent variable vary based on the community the registrant is located. All of the standard deviations of the random coefficients are higher than the fixed coefficients, indicating that there is high variability in the relationship between homeownership, insurance status, and assistance level at the between-community level.

In the final model, the pseudo-R^2 is 0.20, meaning that 20% of the variance in assistance level is explained by the fixed and random effects included in this model. This is acceptable for a data set like this, where there is a high amount of variability in assistance levels. Furthermore, the main source of variation in assistance level would be expected to come from individual circumstances, such as the level of damages to a disaster survivor's residence. 20% of the variance coming from demographic factors is in fact quite significant.

Number of Contacts

The number of contacts corresponds to the number of times there was some phone interaction between FEMA and the registrant. When modeling this dependent variable, I removed all registrants that had no contacts with FEMA, as this constituted the vast majority of registrants. This reduced the data set to 1,141,388 rows.

Again I ran a null model with just the random intercept added to allow intercepts to vary by ZIP code, and found that around 16% of the variance is found at the community level, warranting a multilevel approach again. I started with a model including all independent variables as fixed effects, and adding a random intercept at the community level.

MODEL INFO:
Observations: 1141388
Dependent Variable: Contact.Count
Type: Mixed effects linear regression

MODEL FIT:
AIC = 7335970.163, *BIC* = 7336412.230
Pseudo-R² (fixed effects) = 0.024
Pseudo-R² (total) = 0.036

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p
(Intercept)	1.937	0.415	4.666	2057.563	0.000
RentOwn	0.048	0.016	3.070	1139630.691	0.002
Age	0.006	0.000	14.357	1141353.000	0.000
GrossIncome	-0.000	0.000	-10.441	1140676.606	0.000
HOI	-0.776	0.016	-47.276	1137160.119	0.000
HHComp	0.271	0.004	68.889	1141226.819	0.000
DR_4193	0.032	0.254	0.125	11006.968	0.900
DR_4277	2.062	0.198	10.399	2934.927	0.000
DR_4285	0.581	0.196	2.966	3431.708	0.003
DR_4332	0.754	0.185	4.081	3480.729	0.000
DR_4339	0.277	0.304	0.910	1910.807	0.363
DR_4337	-0.370	0.181	-2.045	3522.081	0.041
pov	1.122	0.417	2.693	2451.319	0.007
unemp	0.774	0.518	1.495	1929.062	0.135
income	0.000	0.000	3.054	2643.385	0.002
nohsdip	-0.759	0.470	-1.613	2152.243	0.107
nohinsur	0.488	0.537	0.909	2190.435	0.364
gini	-0.808	0.496	-1.631	2216.299	0.103
age65	0.001	0.401	0.003	2005.833	0.998
age17	-1.001	0.696	-1.438	2267.472	0.150
disab	0.175	0.569	0.307	1896.385	0.759
sph	-0.304	0.699	-0.435	2166.852	0.664
minority	-0.210	0.532	-0.395	1021.872	0.693
asian	-1.845	0.856	-2.156	1252.401	0.031
black	1.659	0.506	3.278	991.856	0.001
hispanic	-0.400	0.525	-0.762	1019.671	0.446
natam	0.015	0.723	0.021	1139.906	0.983
limeng	1.257	0.418	3.007	1378.610	0.003
owner	0.391	0.256	1.526	1959.568	0.127
mobile	0.905	0.199	4.538	1831.525	0.000
novehic	-0.476	0.650	-0.733	3176.637	0.464
mustruc	-0.379	0.209	-1.810	1525.725	0.070
groupq	-0.523	0.437	-1.197	2319.242	0.231
crowd	-0.255	0.987	-0.258	2160.674	0.796
vacant	1.298	1.122	1.157	2116.629	0.247

p values calculated using Satterthwaite d.f.

RANDOM EFFECTS:

Group	Parameter	Std. Dev.
DDZip	(Intercept)	0.659
Residual		6.010

Grouping variables:

Group	# groups	ICC
DDZip	2355	0.012

Figure 13: Initial Model for Contact Count

This model is questionable due to the poor pseudo-R^2 value of 0.036. It is likely that when it comes to the number of contacts between the registrant and FEMA, this has much more to do with variables not included in this model, such as the number of programs applied to. Of the individual variables, all are significant but only homeowner's insurance status has a high magnitude coefficient of -0.776, indicating that those with insurance do not make or receive as many calls with FEMA.

MODEL INFO:
Observations: 1141388
Dependent Variable: Contact.Count
Type: Mixed effects linear regression

MODEL FIT:
AIC = 7336139.209, *BIC* = 7336414.007
Pseudo-R² *(fixed effects)* = 0.024
Pseudo-R² *(total)* = 0.036

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p
(Intercept)	2.648	0.197	13.449	3126.090	0.000
RentOwn	0.126	0.015	8.567	1135035.245	0.000
GrossIncome	-0.000	0.000	-10.401	1140690.406	0.000
HOI	-0.777	0.016	-47.333	1136526.771	0.000
HHComp	0.256	0.004	67.498	1141340.657	0.000
DR_4193	0.016	0.254	0.065	11203.514	0.949
DR_4277	2.084	0.193	10.816	3015.994	0.000
DR_4285	0.629	0.190	3.308	3503.148	0.001
DR_4332	0.814	0.177	4.604	3601.762	0.000
DR_4339	0.281	0.273	1.029	1937.192	0.303
DR_4337	-0.364	0.175	-2.083	3630.498	0.037
pov	0.506	0.332	1.522	2040.229	0.128
unemp	0.669	0.500	1.338	1961.629	0.181
nohsdip	-0.929	0.403	-2.306	1881.231	0.021
age17	-1.373	0.420	-3.270	2052.302	0.001
minority	-0.695	0.168	-4.149	1395.836	0.000
asian	-1.280	0.625	-2.047	1463.648	0.041
black	1.960	0.183	10.732	1288.874	0.000
limeng	1.311	0.385	3.408	1337.006	0.001
mobile	0.916	0.189	4.840	1824.727	0.000
mustruc	-0.428	0.156	-2.742	1416.126	0.006

p values calculated using Satterthwaite d.f.

RANDOM EFFECTS:

Group	Parameter	Std. Dev.
DDZip	(Intercept)	0.665
Residual		6.010

Grouping variables:

Group	# groups	ICC
DDZip	2355	0.012

Figure 14: Stepwise Model for Contact Count

Running the original model through the same stepwise variable removal as in the assistance level modeling, results in the above model. Several of the same variables met the significance criteria as before, including poverty level, education level, unemployment rate, and minority status. Once again, all individual level variables were included in the final model. The pseudo-R^2 is low, but again we would expect the variance in the number of contacts to depend on more specific aspects of the survivor journey, such as which programs were applied for. Nevertheless, we can still note that in areas with a high proportion of the population age 17 and below, areas with lower education attainment, and higher Asian population, registrants are likely to have fewer contacts with FEMA. Areas with more mobile homes, a higher Black population, and higher proportion of the population speaking English less than "well," are likely to have more contacts with FEMA.

Applied for and Approved for General Assistance

For the final two models, I tried employing multilevel logistic regression modeling approaches, as the response variable is binary in both cases. The first variable included all registrants, and tests whether or not they applied for general assistance. The final model predicted the likelihood of being approved for general assistance, given that the registrant applied for general assistance. Unfortunately, neither model was able to converge using the GLMER package in R.

In this logistic regression model, almost all variables are highly significant except for the proportion of the population younger than 17. For the individual variables, homeownership, age, income, and household composition all have a positive correlation with the likelihood of approval for individual assistance. The only negative relationship is with homeowner's insurance status.

Chapter 5: Discussion and Future Work

The widespread impacts of natural disasters provide a rich opportunity for sociologists and other researchers to seek to understand potential disparities along lines of race, class, gender, or other demographic distinctions. Disasters are, in essence, social and societal disruption to the extreme degree. Given the increasing ubiquity of disasters and their impacts, those interested in social inequality should be concerned with how disasters intersect with social vulnerability and create or perpetuate inequality. In this work, I have attempted to explore this relationship further in the context of federal aid, specifically FEMA's Individual Assistance programs. Part of the insight of this work is methodological. Multilevel modeling can provide researchers with the ability to examine the relationship between disaster vulnerability factors and assistance outcomes at both the individual and community levels. One indication of my research is that at times, community-level factors have a greater impact on assistance level than individual-level demographics, as is the case in the relative impacts of per-capita income at the community level and individual-level gross income on assistance level. Another insight of this research is in the variable selection stage for the community-level disaster vulnerability indicators. There is much overlap between the social vulnerability indexes I chose to incorporate in this analysis, mainly that they include some measures of socioeconomic status, age and disability, minority and language status, and housing status. Those wishing to develop new measurements of disaster vulnerability, or those hoping to take it into account in policy recommendations, should follow what has been done in the past. My modeling shows that there are strong relationships between several of these variables and outcomes in FEMA Individual Assistance programs. However, significance testing should not be used to remove variables from social vulnerability indexes—there are many contexts, noted in the literature, in which inclusion of more variables is warranted despite their seeming statistical insignificance.

Future work should focus on expansion of these models in broader contexts. With this data, further exploration can include other metrics of disaster vulnerability, and interaction terms to test more complex relationships between variables. However, much of this research was limited by the amount of data available. The data set received from FEMA was large and diverse in many ways, in the number of disasters, types of disasters, locations, and in the number of registrants. However, there was little individual demographic data available to fully explore the relationship between demographic variables associated with disaster vulnerability and assistance outcomes. Also, the community-level variables are limited by the measurement techniques and broad timeframes supported by the US Census. It is important to test the relationship between characteristics of where a registrant lives with the outcomes of their assistance, but this analysis provides a more limited picture overall into the vulnerability of an individual disaster survivor than could be achieved with more individual-level demographic data.

Bibliography

Capps, Kriston. "Why Are These Tiny Towns Getting So Much Hurricane Harvey Aid?" *Citylab*, 3 October 2018, https://www.bloomberg.com/news/articles/2018-10-03/hurricane-harvey-recovery-aid-shows-racial-disparities.

CDC SVI 2018 Documentation, 31 January 2020.

Cutter, Susan L., et al. "Social Vulnerability to Environmental Hazards*." *Social Science Quarterly*, vol. 84, no. 2, 2003, pp. 242–61. *Wiley Online Library*, doi:10.1111/1540-6237.8402002.

Dahlhamer, James M. *Loan Request Outcomes in the U.S. Small Business Administration Business Disaster Loan Program. Disaster Research Center*, 1994. *udspace.udel.edu*, https://udspace.udel.edu/handle/19716/621.

Edgemon, Lesley, et al., "Community Resilience Indicator Analysis: County-Level Analysis of Commonly Used Indicators from Peer-Reviewed Research, 2019 Update." *Argonne National Laboratory*, 2019.

Elliott, James R., and Junia Howell. "Beyond Disasters: A Longitudinal Analysis of Natural Hazards' Unequal Impacts on Residential Instability." *Social Forces*, vol. 95, no. 3, Oxford Academic, Mar. 2017, pp. 1181–207. academic.oup.com, doi:10.1093/sf/sow086.

Finch, W. Holmes, Jocelyn E. Bolin, and Ken Kelley. *Multilevel modeling using R*. Crc Press, 2019.

Flanagan, Barry E., et al. "A Social Vulnerability Index for Disaster Management." *Journal of Homeland Security and Emergency Management*, vol. 8, no. 1, De Gruyter, Jan. 2011. *www.degruyter.com*, doi:10.2202/1547-7355.1792.

Fussell, Elizabeth. "The Long Term Recovery of New Orleans' Population after Hurricane Katrina." *The American Behavioral Scientist*, vol. 59, no. 10, Sept. 2015, pp. 1231–45. *PubMed Central*, doi:10.1177/0002764215591181.

Goldblatt, Ran. "POST for Lifeline Impact Assessment: A Preliminary Conceptual Note." 22 July 2019.

Guo, Guang, and Hongxin Zhao. "Multilevel Modeling for Binary Data." *Annual Review of Sociology*, vol. 26, no. 1, 2000, pp. 441–62. *Annual Reviews*, doi:10.1146/annurev.soc.26.1.441.

Howell, Junia, and James R. Elliott. "Damages Done: The Longitudinal Impacts of Natural Hazards on Wealth Inequality in the United States." *Social Problems*, vol. 66, no. 3, Oxford Academic, Aug. 2019, pp. 448–67. *academic.oup.com*, doi:10.1093/socpro/spy016.

Hox, Joop J. and Cora J. M. Maas. "Multilevel Analysis." *Encyclopedia of Social Measurement*, vol. 2, 2005.

Hurricanes and Tropical Storms - Annual 2017 | State of the Climate | National Centers for Environmental Information (NCEI). https://www.ncdc.noaa.gov/sotc/tropical-cyclones/201713. Accessed 28 Aug. 2020.

Pais, Jeremy F., and James R. Elliott. “Places as Recovery Machines: Vulnerability and Neighborhood Change after Major Hurricanes.” *Social Forces*, vol. 86, no. 4, Oxford University Press, 2008, pp. 1415–53. JSTOR.

Preston, Benjamin L. “Local Path Dependence of U.S. Socioeconomic Exposure to Climate Extremes and the Vulnerability Commitment.” *Global Environmental Change*, vol. 23, no. 4, Aug. 2013, pp. 719–32. *ScienceDirect*, doi:10.1016/j.gloenvcha.2013.02.009.

Schiermeier, Quirin. “Droughts, Heatwaves and Floods: How to Tell When Climate Change Is to Blame.” *Nature*, vol. 560, no. 7716, 7716, Nature Publishing Group, July 2018, pp. 20–22. *www.nature.com*, doi:10.1038/d41586-018-05849-9.

SoVI® Evolution | Hazards & Vulnerability Research Institute | University of South Carolina. http://artsandsciences.sc.edu/geog/hvri/sovi%C2%AE-evolution. Accessed 28 Aug. 2020.

Tierney, Kathleen. "Hurricane Katrina: catastrophic impacts and alarming lessons." *Risking house and home: Disasters, cities, public policy* (2008): 119-136.

Tierney, Kathleen. *Disasters: A Sociological Approach*. Medford: *Polity Press*, 2019.

Tierney, Kathleen, et al. “Metaphors Matter: Disaster Myths, Media Frames, and Their Consequences in Hurricane Katrina.” *The ANNALS of the American Academy of Political and Social Science*, vol. 604, no. 1, SAGE Publications Inc, Mar. 2006, pp. 57–81. *SAGE Journals*, doi:10.1177/0002716205285589.

Walker, Mark, et al. “FEMA’s Hurricane Aid to Puerto Rico and the Virgin Islands Has Stalled.” *The New York Times*, 27 Nov. 2019. *NYTimes.com*, https://www.nytimes.com/2019/11/27/us/politics/fema-hurricane-aid-puerto-rico-virgin-islands.html.