The Socioeconomic Determinants of Climate-related Deaths and Injuries: A County-level Analysis of the U.S.

Kyuin LEE
Georgetown University,
Master of Science in Data Science for Public Policy

EXECUTIVE SUMMARY

Using a combination of parametric (LASSO) and non-parametric (random forest) techniques, this study aims to 1) predict the number of mortality and injuries caused by extreme weather events and 2) identify the most vulnerable communities that are at higher risk of losing lives. In general, the models failed to predict casualty with high accuracy, but results pointed to an important finding that wealthy communities are just as likely to be hit by natural disasters as the socioeconomically disadvantaged communities, if not more. As a result, wealthy communities were exposed to high risk of death and injuries, which was contrary to our initial hypothesis. The current disaster response planning framework that takes into account the housing conditions and the age makeup of the population seems to be relevant for effective disaster prevention and recovery. However, additional factors, such as the lack of emergency health care in the vicinity, could also guide the strategic deployment of resources.

INTRODUCTION

A global warming of 1°C above pre-industrial levels seems like a minuscule increase in Earth's temperature. Over the last 40 years, however, the U.S. has witnessed a surge in extreme weather events, whose damage exceeds $1 billion (inflation-adjusted) in direct losses (National Oceanic and Atmospheric Administration [NOAA], 2023). As high frequency and high intensity weather events become the 'New Normal,' the impacts of climate change are starting to sink in. Notably, 2020 and 2021 was characterized by a diversity of natural disasters that have swept all across America: wildfires across the West, hurricanes along the East coast, drought and heat waves in Western and Central states, and historic cold waves throughout much of the country (Smith, 2022, 2023).

Given the increased frequency and magnitude of extreme weather events, I sought to examine their impacts on human well-being and health, from a justice and equity perspective. The purpose of this research was two-fold: 1) to **investigate the socio-economic predictors of climate-induced deaths and injuries (prediction)** and 2) to **identify the at-risk population that are the least capable of preparing for and recovering from the impacts of climate catastrophes (inference).**

As part of disaster risk management, public health agencies assess the vulnerability of a community by ranking states/census-tracts along 16 socio-economic dimensions, which fall under the following themes: *1) socio-economic status, 2) household characteristics, 3) racial/ethnic minority status*, and *4)* *housing type/transportation* (Centers for Disease Control and Prevention [CDC], 2022a)[1]. However, since social vulnerability scores are determined by percentile ranks, **vulnerability can only be understood in relation to another geography**. **Another limitation of this approach is that each component is given equal weight** in determining the degree of vulnerability (Cutter & Emrich, 2017), regardless of its relevance to the climate disaster context. Therefore, by testing each of the 16 social vulnerability measures as predictors of casualty, I sought to **tease out the most relevant set of structural determinants,** which we have limited understanding of, heretofore. In light of the unprecedented climate crisis, current research not only serves as a timely line of inquiry, but it is also of relevance to emergency response and recovery workers, as it would inform the effective and efficient deployment of resources, ultimately minimizing the loss of lives.

DATA

Research was conducted at the **county-level,** as *counties* were the smallest unit of analysis at which climate disaster-related outcomes were measured. The scope of research spanned the last four years for which data was available (**2016-2020**) and I chose a *<u>multi-year</u>*, rather than a *single-year* timeframe, for the following reasons:

1. *<u>To ensure the representation of every geography in the US</u>*, as communities with populations of less than 65,000 are omitted from 1-year estimates of social, economic, housing, and demographic characteristics (U.S. Census Bureau, 2022); and
2. *<u>To avoid data imbalance for climate-induced casualties</u>*, given that natural disasters are conceptualized as ‘exceptional’ weather events.

The analysis drew upon 2 sources of data:

1. **Social Vulnerability Index (SVI) 2016-2020** (CDC, 2022a, 2022b), for data on *socio-economic characteristics of the U.S. population;* and
2. **Social Determinants of Health (SDOH) 2016-2020** (Agency for Healthcare Research and Quality [AHRQ], 2022a, 2022b), for data on *weather-related casualties, the frequency of extreme weather events, and healthcare access.*

Since data had been aggregated to the *county* level and counties were identified by a common set of geographic identifiers (FIPS), the selected datasets were ideal for combining multiple sources of data.

[1] The 16 vulnerability indicators refer to the number of population/households/housing units: **1) *Below 150% Poverty,*** **2) *Unemployed*, 3) *Housing Cost Burden*, 4) *No High School Diploma*,** and **5) *No Health Insurance*** (for <u>Theme 1</u>: socioeconomic status); **6) *Aged 65 & Older,* 7) *Aged 17 & Younger,* 8*) Civilians with a Disability,* 9) *Single-Parent Households,*** and **10) *Limited English Language Proficiency*** (for <u>Theme 2</u>: household characteristics); **11) *Racial/Ethnic Minority*** (for <u>Theme 3</u>: racial and ethnic minority status); and **12) *Multi-Unit Structures,* 13) *Mobile Homes,* 14) *Crowded Housing,* 15) *No Vehicle,*** and **16) *Group Quarters*** (for <u>Theme 4</u>: housing type and transportation).

Variable selection

The model consisted of 25 features, 16 of which were the *CDC's social vulnerability indicators,* 8 of which were added as supplementary measures of vulnerability—*the population of* *__individual__* *ethnic/racial minority categories (%), households without internet access (%), proximity to the nearest emergency ward (miles),* and *the number of hospitals with emergency wards (per 100,000 population)*—and the remaining variable, which was a proxy for *the frequency of climate disaster events*[2]. CDC's vulnerability measures were available either in *absolute terms* or in *percentages,* but the latter was preferred over the former, to adjust for the size of the population/household within each county.

Besides the *__overall__* percentage of the ethnic/racial minority population, I examined the proportion of *__each__* *ethnic/racial minority status* as a measure of social vulnerability, because evidence pointed to higher mortality rates among certain minority groups, namely among African Americans than Hispanic/Latinos (Klinenberg, 1999; Shonkoff et al., 2009). The *percentage of households without internet connectivity* was also used for prediction, because the internet was an important source of gaining timely and credible information amidst the latest public health crises (Benda et al., 2020). Moreover, *distance to the nearest emergency ward* and *the number of hospitals with an emergency department* were added as supplementary indicators, to gauge the availability of timely treatment.

Feature engineering

Data was aggregated across 2016-2020. *Distance to the nearest emergency ward* was calculated as the *__average__* of the mean distances for 2016-2020, using population weighted tract centroids in the county. The *number of hospitals with emergency wards* was adjusted for the population and then *__averaged__* across the 4 years. The outcome variable, *casualty caused by climate events*, was adjusted for the population and was operationalized as the *sum* of both *direct* and *indirect* deaths and injuries for each year, expressed as *__per 100,000 capita__*.

Descriptive statistics

Ahead of the analysis, 1 obvious outlier and 6 counties that have never been struck by a natural disaster were removed. The final dataset consisted of 3,136 observations (counties), which excluded the

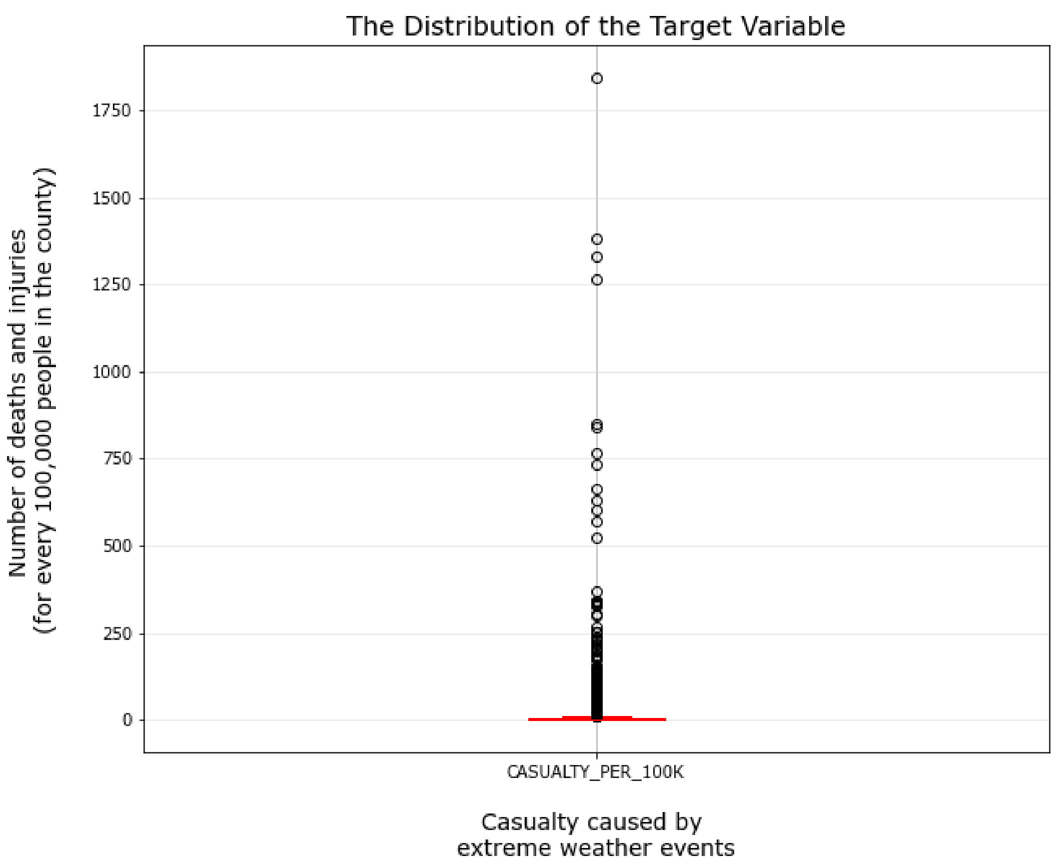
---

[2] Climate disasters are defined as 48 different types of storm events and unusual weather events that have significant potential to cause death, injuries, and damage to property and businesses: Astronomical Low Tide, Avalanche, Blizzard, Coastal Flood, Cold/Wind Chill, Debris Flow, Dense Fog, Dense Smoke, Drought, Dust Devil, Dust Storm, Excessive Heat, Extreme Cold/Wind Chill, Flash Flood, Flood, Frost/Freeze, Funnel Cloud, Freezing Fog, Hail, Heat, Heavy Rain, Heavy Snow, High Surf, High Wind, Hurricane (Typhoon), Ice Storm, Lake-Effect Snow, Lakeshore Flood, Lightning, Rip Current, Seiche, Sleet, Sneaker Wave, Storm Surge/Tide, Strong Wind, Thunderstorm Wind, Tornado, Tropical Depression, Tropical Storm, Tsunami, Volcanic Ash, Waterspout, Wildfire, Winter Storm, and Winter Weather (NOAA, 2021).

5 territories of the U.S[3]. On average, there were 84 climate catastrophes in the U.S, throughout 2016-2020 and 13 deaths and injuries per 100,000 population (**Table 1**). There were large differences in casualty as well as incidence of climate disasters across counties. Counties also differed in terms of minority group populations: in particular, relatively large differences were observed in the make-up of African American and Hispanic/Latino race/ethnicities.

To be able to discern meaningful patterns in our data, it is ideal to have a wide distribution of casualty caused by climate disasters. According to **Figure 1**, casualty does indeed vary widely in range, but the large county-level differences in the number of deaths and injuries are driven by the existence of outliers. The actual distribution is heavily skewed towards the lower bounds of casualty, indicating that an overwhelming majority of U.S. counties have recorded hardly any mortality or injuries, following extreme weather events.

Figure 1. A boxplot of casualty caused by extreme weather events



[3] The 5 territories of the U.S. — American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands — which are referred to as county equivalents, were excluded from the analysis as the SDOH does not keep track of healthcare-related data for these regions.

Table 1. Descriptive statistics (N = 3,136)

Variable	Mean	SD	Minimum	Median	Maximum
Socioeconomic status					
Below poverty (%)	24.51	8.50	0.00	23.75	71.00
Unemployed (%)	5.20	2.58	0.00	4.90	30.40
Burdened with housing costs (%)	22.28	5.26	0.00	22.00	49.40
No high school (%)	12.41	6.04	1.40	11.20	78.10
No health insurance (%)	9.54	5.10	0.50	8.50	42.60
Household characteristics					
Aged 65 and older (%)	19.23	4.78	3.00	18.85	57.80
Aged 17 and younger (%)	22.09	3.55	5.20	22.10	42.70
With disability (%)	16.00	4.49	4.30	15.50	47.90
Single parent (%)	5.88	2.38	0.00	5.60	22.70
Limited English proficiency (%)	1.62	2.69	0.00	0.70	32.00
Racial/ethnic minority status					
Minority (%)	24.23	20.21	0.00	16.90	99.00
African American (%)	8.89	14.39	0.00	2.20	87.80
Hispanic/Latino (%)	9.59	13.92	0.00	4.40	98.90
Asian (%)	1.39	2.81	0.00	0.60	41.70
American Indian/Alaska Native (%)	1.79	7.53	0.00	0.20	92.90
Native Hawaiian/Pacific Islander (%)	0.09	0.41	0.00	0.00	11.00
Housing type/Transportation					
Multi-unit housing (%)	4.79	5.83	0.00	2.90	89.60
Mobile homes (%)	12.59	9.54	0.00	10.40	56.90
Crowded housing (%)	2.37	2.39	0.00	1.80	42.20
No vehicle (%)	6.18	4.50	0.00	5.50	89.30
Group quarters[4] (%)	3.49	4.52	0.00	1.90	45.50
Other					
No internet (%)	16.91	7.57	2.20	15.60	62.40
Nearest Emergency Room (ER) (miles)	8.16	13.43	0.32	5.69	468.81
Climate disaster events	83.83	107.22	1.00	60.00	2348.00
Climate-related health outcomes					
Casualty (per 100k capita)	12.85	70.47	0.00	0.00	1844.68

[4] Group quarters are defined as living arrangements other than housing units. Typically, occupants of group quarters are not related by family ties. Examples of such include correctional facilities, nursing homes, and mental hospitals (institutional group quarters) and college dormitories, military barracks, group homes, and shelters (institutional group quarters) (U.S. Census Bureau, 2021).

Table 2 represents the correlation between key variables. While vulnerability measures were low-to-moderately correlated with one another, they exhibited low degrees of correlation with the number of casualty. Since correlation detects only linear relationships, I visualized a scatterplot of each social vulnerability measure against the number of casualty (**Figure 2**). Although non-linear, some dimensions of socioeconomic vulnerability demonstrated a quadratic pattern with mortality and injuries.

Figure 2. A scatterplot of social vulnerability indices and casualty

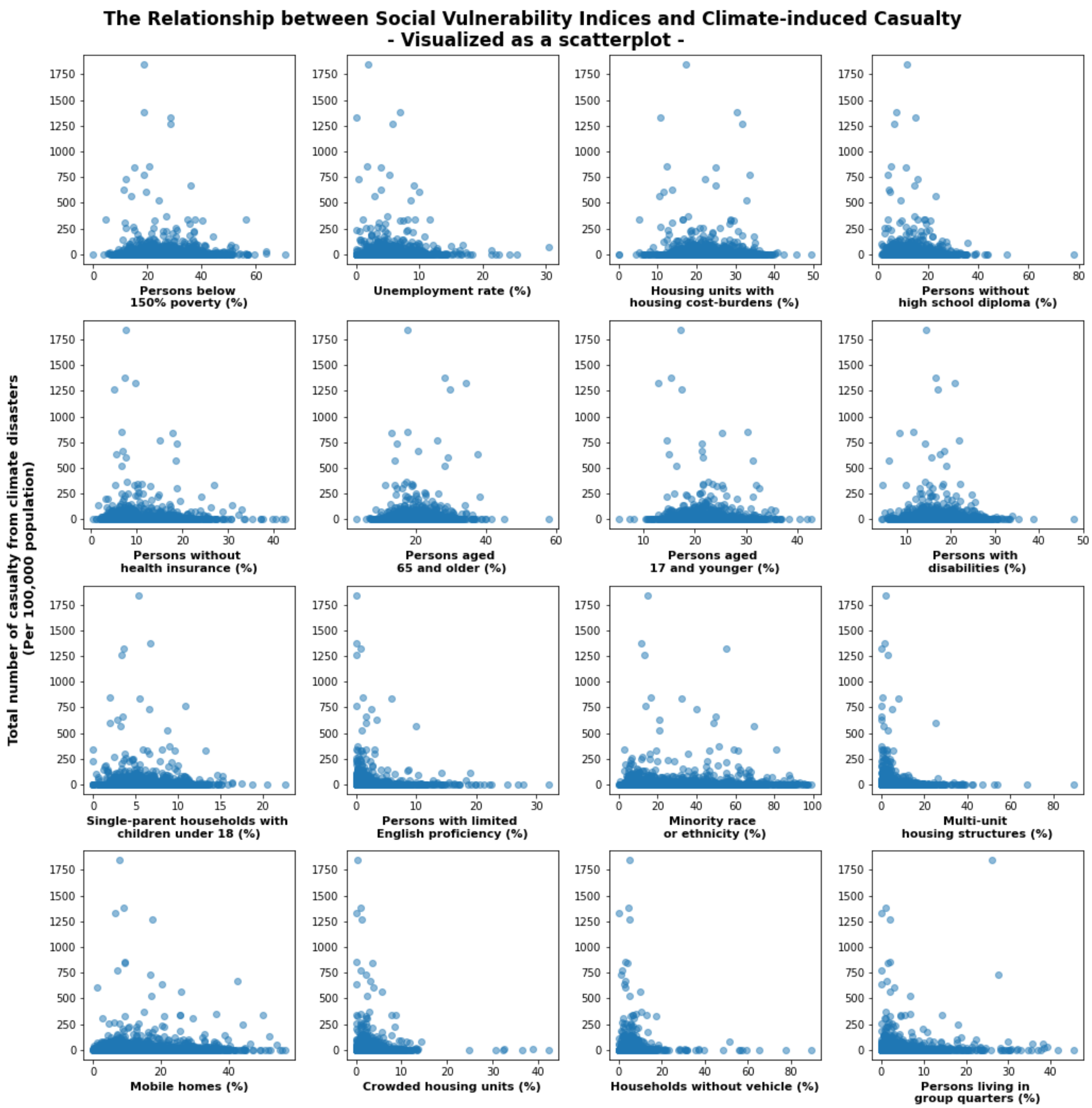


Table 2. Correlation between 16 main social vulnerability measures and climate-induced deaths and injury

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Below poverty	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. Unemployed	0.56	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3. Housing burden	0.25	0.28	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4. No high school	0.65	0.37	0.05	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—
5. Uninsured	0.44	0.25	0.02	0.54	1.00	—	—	—	—	—	—	—	—	—	—	—	—
6. 65 and older	-0.04	-0.13	-0.19	-0.15	-0.16	1.00	—	—	—	—	—	—	—	—	—	—	—
7. 17 and younger	0.09	0.05	-0.13	0.22	0.33	-0.57	1.00	—	—	—	—	—	—	—	—	—	—
8. Disability	0.53	0.34	-0.03	0.39	0.10	0.41	-0.27	1.00	—	—	—	—	—	—	—	—	—
9. Single parent	0.47	0.41	0.34	0.38	0.29	-0.44	0.35	0.06	1.00	—	—	—	—	—	—	—	—
10. Language barrier	0.08	0.05	0.18	0.47	0.35	-0.29	0.29	-0.22	0.18	1.00	—	—	—	—	—	—	—
11. Minority	0.40	0.42	0.36	0.48	0.50	-0.39	0.30	-0.09	0.58	0.55	1.00	—	—	—	—	—	—
12. Multi-unit housing	-0.22	-0.05	0.43	-0.24	-0.16	-0.34	-0.04	-0.42	0.07	0.25	0.20	1.00	—	—	—	—	—
13. Mobile homes	0.54	0.30	-0.10	0.53	0.33	0.13	-0.06	0.52	0.13	-0.04	0.16	-0.44	1.00	—	—	—	—
14. Crowded housing	0.28	0.34	0.10	0.33	0.43	-0.31	0.41	-0.08	0.29	0.40	0.51	0.10	0.08	1.00	—	—	—
15. No vehicle	0.37	0.42	0.26	0.25	0.14	-0.16	0.05	0.11	0.34	0.08	0.34	0.28	0.01	0.44	1.00	—	—
16. Group quarters	0.15	0.09	0.07	0.16	0.03	-0.15	-0.31	0.06	0.08	0.02	0.15	0.00	0.10	0.02	0.14	1.00	—
17. Casualty	0.00	-0.01	-0.02	-0.01	0.03	0.07	-0.04	0.01	-0.03	-0.01	0.03	-0.05	0.05	-0.01	-0.04	0.05	1.00

Limitations of the dataset

All in all, the limitations of the dataset could be summarized as follows: First of all, **imbalance in the number of climate-related casualties** could undermine the accuracy of our model. Secondly, our data on climate disaster events **fails to account for different *types* and *magnitudes of each episode***. By nature, certain types of natural disasters are more destructive than others and events of the *same* type but of *different* magnitudes would have highly varying degrees of impact on casualty. But due to lack of data on individual climate events, our model would be unable to pick up such nuance. Thirdly, **the current rate of casualty is an underestimation of the actual number of deaths and injuries**, due to missing data for 2018 (AHRQ, 2022b). Hence the social vulnerability measures were expressed as 5-year estimates (2016-2020), whereas medical access, the frequency of extreme weather events, and casualty were presented as 4-year aggregates or estimates.

METHODOLOGY

Since the purpose of this research was not only to predict climate-related casualty, but also to identify the most relevant set of determinants, I employed *variable selection techniques*—such as 1) LASSO and 2) random forest. The analysis drew upon regression techniques, as our measure of casualty was quantitative and continuous in nature.

LASSO regression assumes a linear relationship between the variables of interest. It produces a parsimonious model by removing redundant or less important predictors and reducing their coefficients to zero, while minimizing the loss of information (Muthukrishnan & Rohini, 2016). As part of a study on life satisfaction among the elderly, Shen et al. (2023) tested the accuracy of their models, using 18 predictors that had moderate to strong degrees of correlation with life satisfaction. Out of the 18 predictors that corresponded to either 1) *individual* (dutifulness, openness, extroversion, neuroticism, optimism, pessimism, positive and negative emotion, hopelessness), 2) *family-related* (positive support from spouse and children, negative support from spouse and children, intimacy with spouse), or 3) *social* factors (subjective social status, social participation/activity, neighborhood relations, and environmental identity), all were shown to be relevant, since none of the coefficients were reduced to zero. In terms of relative magnitude, however, LASSO regression revealed that subjective social status as well as both positive and negative emotions were the most critical predictors of life satisfaction.

Random forest is a non-linear ensemble method that relies on 'the wisdom of the crowd.' It averages the results of multiple decision trees that do not look similar to each other. I opted for the random forest model over simple decision trees, as the former tends to produce more accurate and stable predictions. While the random forest model is unable to pick up the direction of the effect, it allows us to rank the relative importance of each feature. For example, Singh et al. (2017) used a random forest model to predict the soil infiltration rate (i.e. velocity at which water enters the soil) based on the *type of impurities in the soil, concentration of impurities, moisture content*, and *cumulative time*. Although all 4 predictors informed the prediction of the outcome, the temporal element was found to be the most important, as model accuracy plummeted when *cumulative time* was removed (i.e. the model was highly dependent on that feature in predicting the rate of soil infiltration).

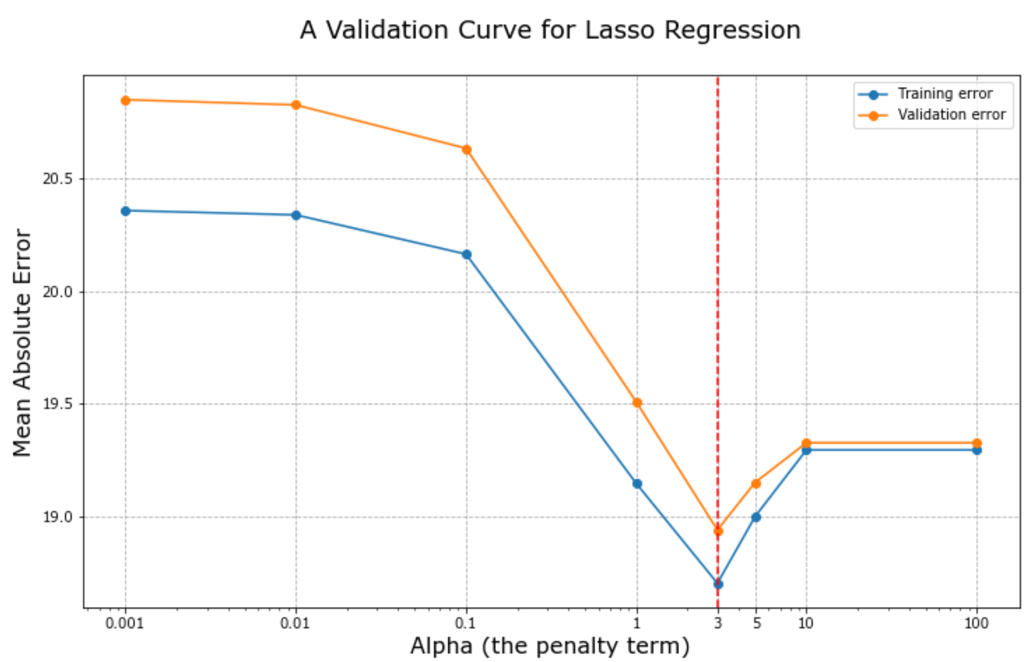
While feature selection and reduced model complexity is a major advantage of the LASSO regression, it is not well-equipped to handle multicollinearity, which could lead to arbitrary selection of variables. Since the 16 main indicators of social vulnerability are grouped under 4 higher-level themes, I observed moderate levels of correlation among the features (**Table 2**). In this regard, random forest was a suitable alternative. Although random forest is difficult to visualize and cannot provide us with individual coefficient estimates, it also had the edge in handling outliers (**Figure 1**) and non-linearity in the data (**Figure 2**), as was the case with our study.

RESULTS

Hyperparameter tuning

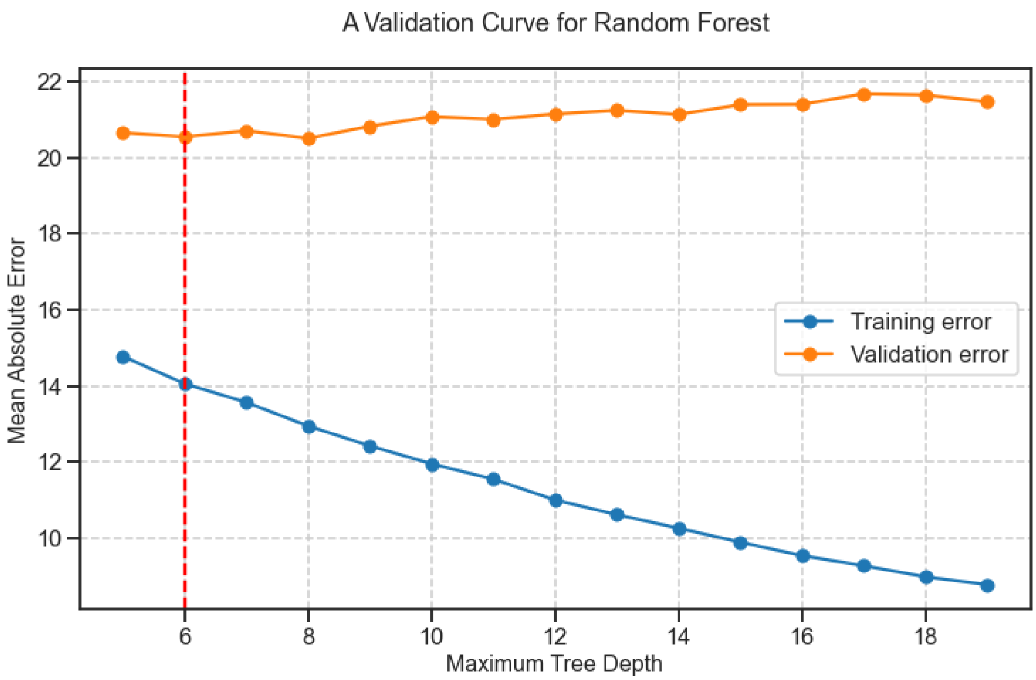
Prior to training the model, I optimized the values of alpha and maximum tree depth for LASSO and random forest, respectively. Mean Absolute Error (MAE) was used as the main metric for evaluation, because of its low sensitivity to outliers, compatibility with non-linearity in the data, and ease of interpretation. **Figure 3** and **Figure 4** show the validation curve of the 2 models, across varying degrees of alpha and tree depth.

Figure 3. Validation curve for the LASSO regression model



For the LASSO regression, an alpha of 3 was selected because it was the point at which the MAE was at its lowest. An alpha value beyond 3 was suboptimal, as the MAE began to increase within this range. With the penalty term (alpha) set to 3, the LASSO regression model removed 18 out of the 25 features available. Thus the final model consisted of 7 features, 4 of which were the CDC's main social vulnerability measures.

Figure 4. Validation curve for the random forest regression model

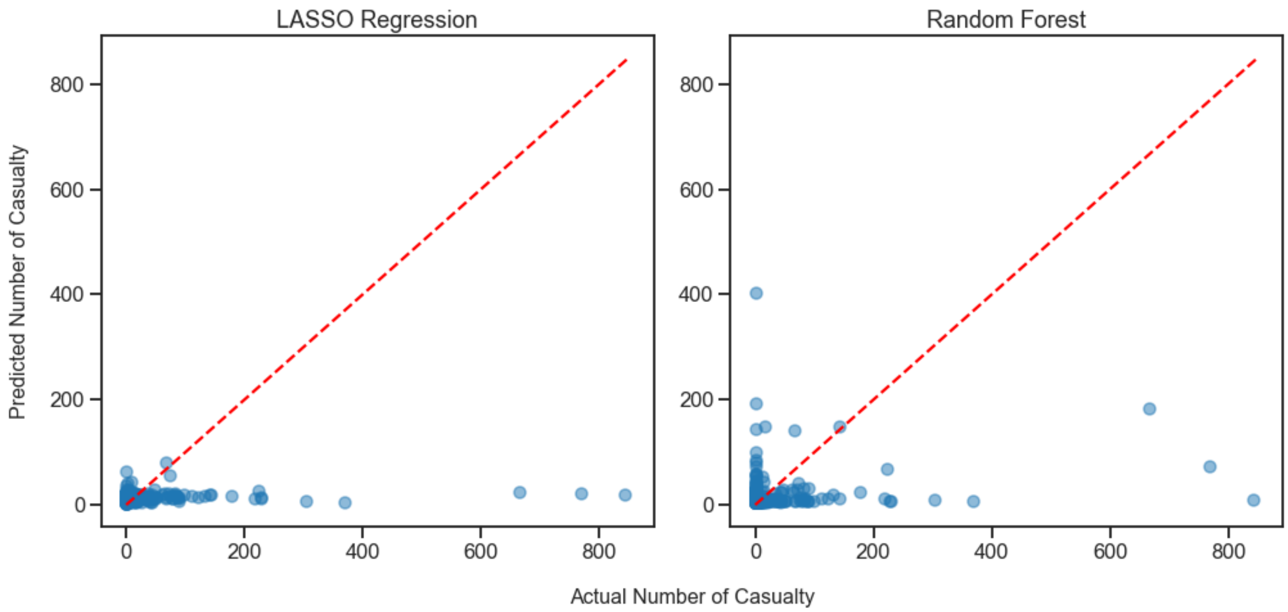


As for the random forest model, there was very little variation in MAE across the range of values. I opted for a maximum tree depth of 6 rather than 8, so as to reduce computation time.

Model performance

Out of the 2 models, random forest (MAE: 18.906, MSE: 3666.207, RMSE: 60.549) outperformed LASSO (MAE: 19.455, MSE: 3656.482, RMSE: 60.469), albeit by a narrow margin. Overall, however, both models failed to capture the relationship between social vulnerability and climate-induced casualties: I found **evidence of underfitting**, where the two models performed poorly on both the training set and the validation/test set (**Figure 3** and **Figure 4**). A scatterplot of the actual number of casualty and the predicted number of deaths and injuries (**Figure 5**) shows poor predictive accuracy of the 2 models. **For counties with lower number of death tolls and injuries, the models tended to overestimate the impacts of climate catastrophes**. On the other hand, **the models tended to underestimate the adverse impacts on human lives, for counties with higher degrees of casualty**.

Figure 5. Deviations from the Actual Number of Deaths and Injuries



<u>Interpretation</u>

Figure 6 denotes the importance of the features, while **Figure 7** shows the strength and direction of the relationships. There are 4 key findings that stand out. Firstly, **the size of the aging population informed the degree of casualty.** For effective prevention and disaster recovery, public health officials could focus more on supporting the elderly.

Secondly, the percentage of the population **living in *<u>non-traditional</u>* forms of housing**–that is, living in units other than a house, apartment, or rented rooms–**was one of the top 3 predictors of climate-related deaths and injuries**. Occupants of *group* quarters 1) live in *<u>crowded facilities</u>* (e.g. students in college dormitories) and/or are populations that are 2) *<u>under care</u>* (e.g. patients) or <u>taken into *custody*</u> (e.g. inmates). It seems that their living conditions and/or health status increases their susceptibility to climate risks.

With respect to race and ethnicity, there were **some evidence that minority status could play a role**. Although there was inconclusive evidence to suggest that counties with a larger African American population were more at risk than others, **counties with a larger Asian population were by far the least affected**, a predictor that was found to be important under both models.

Lastly, ***distance to the nearest emergency ward***–which does <u>not</u> belong to CDC's measure of social vulnerability, but was nonetheless considered relevant–turned out to be highly predictive of casualty.

However, an important caveat is that these results are far from conclusive, due to low model performance. Therefore care must be taken in interpreting such findings.

Figure 6. An overview of feature importance (random forest)

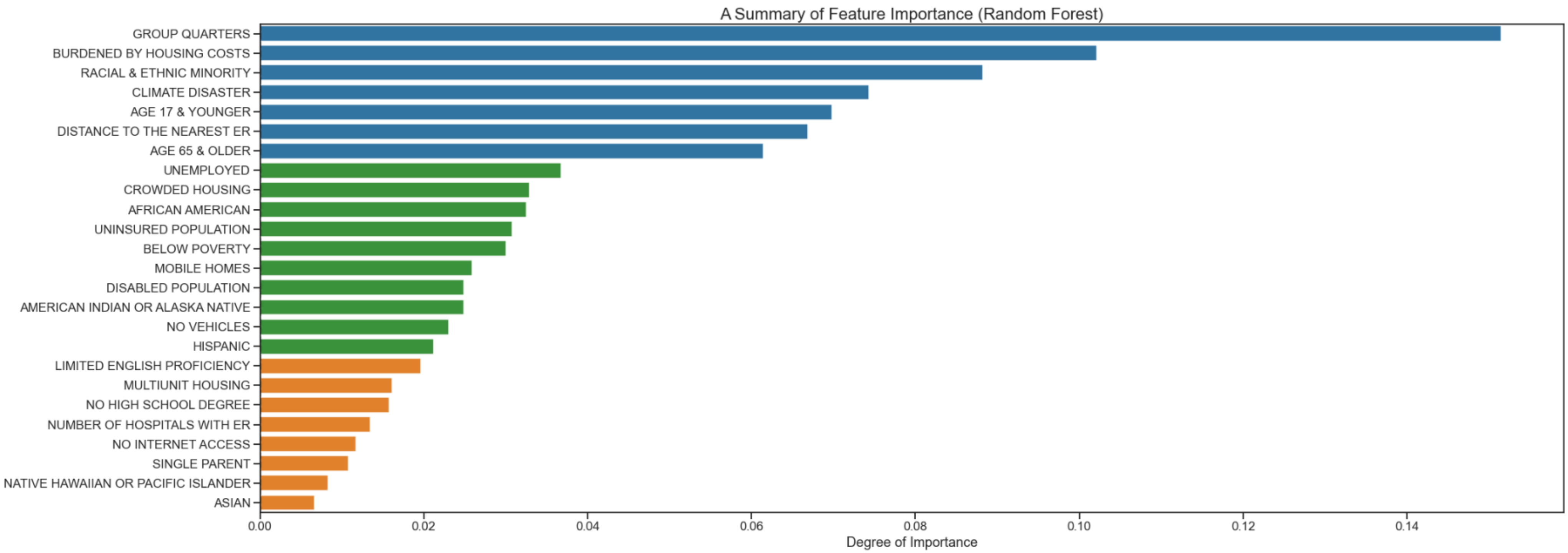
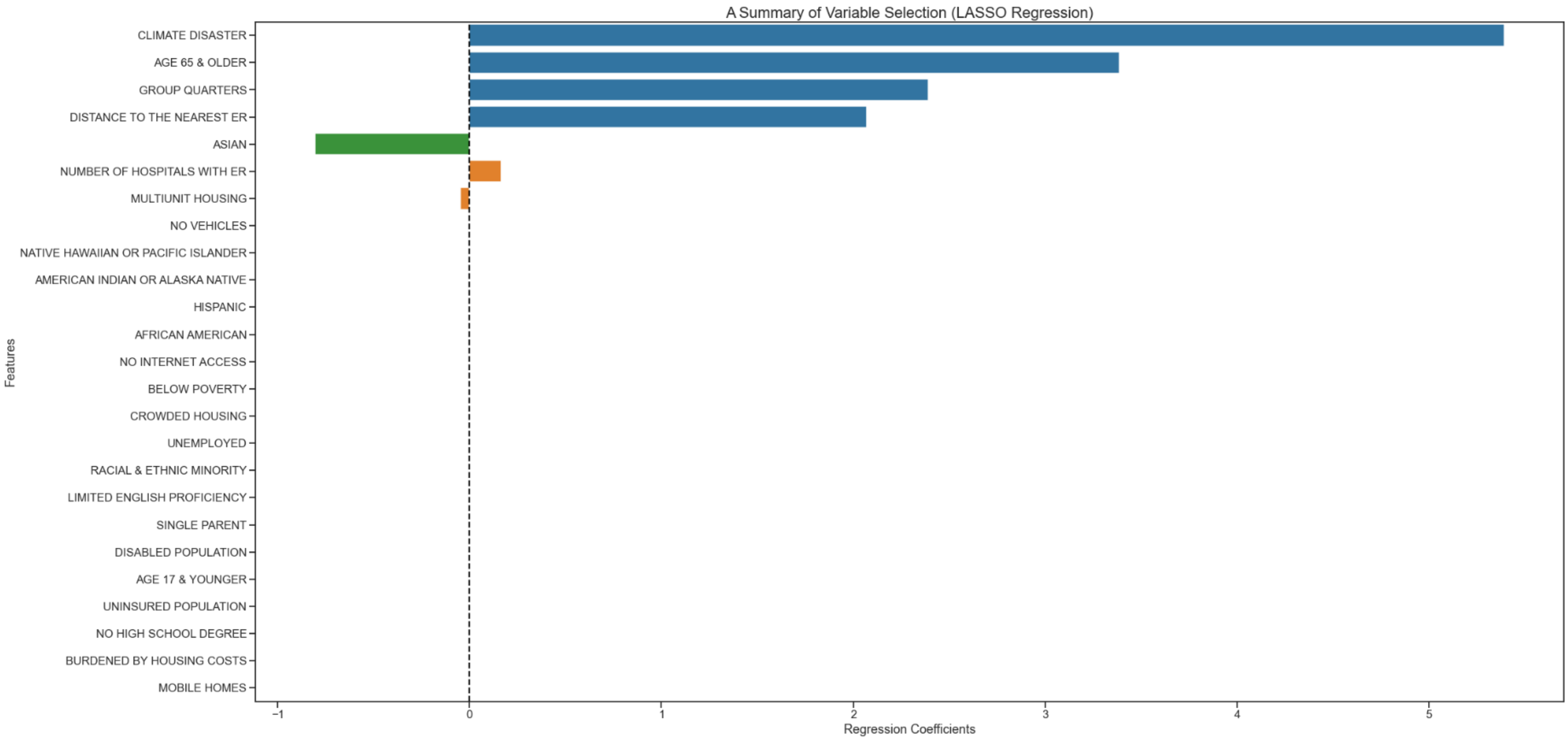


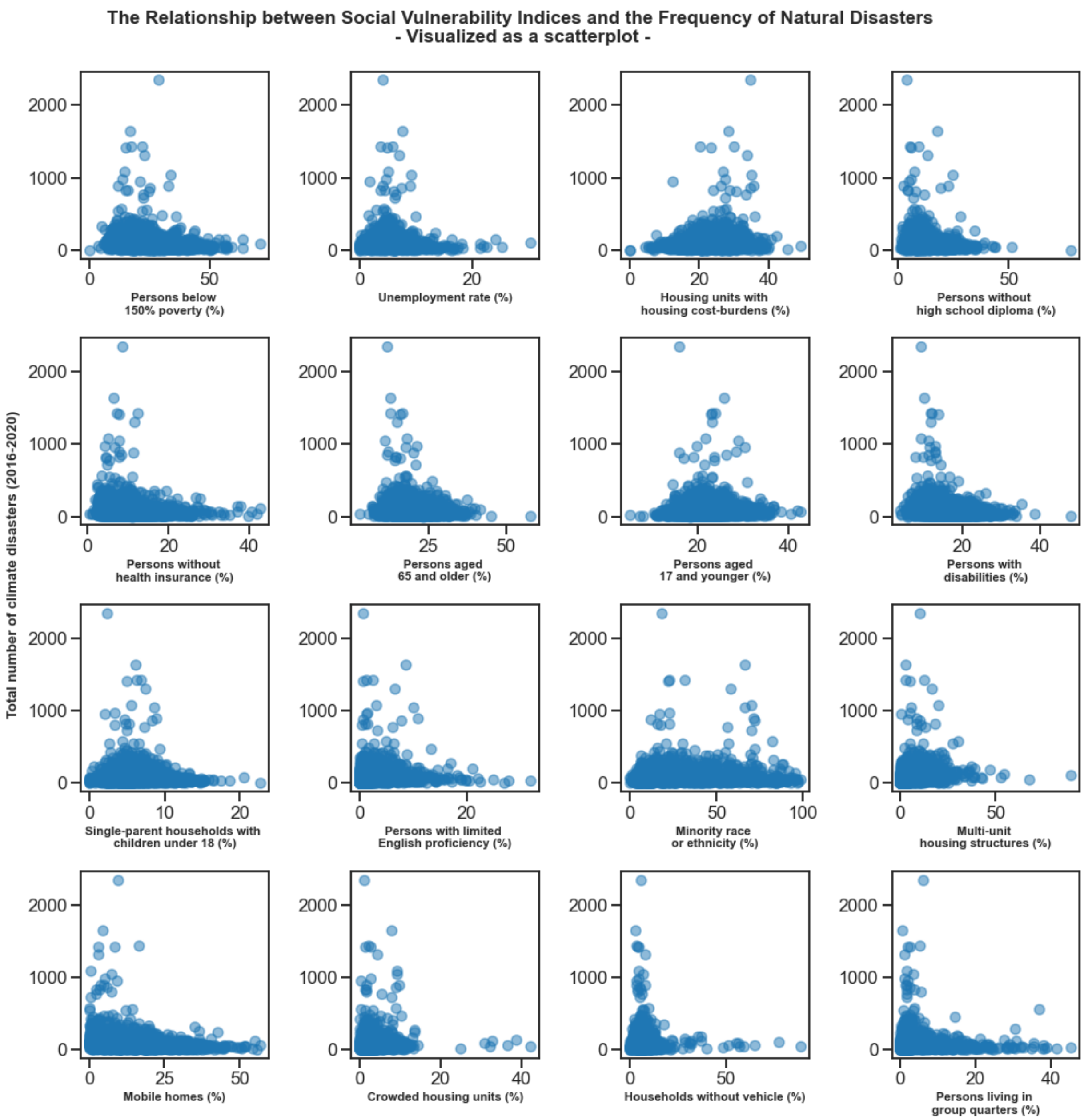
Figure 7. The results of variable selection (LASSO regression)



Post-hoc analysis

To explore the potential causes of low model accuracy, I revisited the relationship between vulnerability and casualty. Contrary to my hypotheses, socioeconomic vulnerability was *inversely* related to climate-induced deaths and injuries (**Figure 2**). A closer examination revealed that **communities with a larger percentage of the vulnerable population were struck by fewer incidence of climate catastrophes, which most likely explains the low fatality rates in these regions** (**Figure 8**).

Figure 8. The degree of exposure to climate disasters, across different levels of social vulnerability



As the subset of the population that we are most interested in–the socially vulnerable–had limited exposure to climate events to begin with, it precludes us from making any generalizations about the disproportionate impacts that fall on such communities.

As part of a post-hoc, exploratory spatial analysis, I constructed an *interactive* heatmap of U.S. counties that fell within the top 10% of social vulnerability. According to **Figure 11** (see Appendix), regions that were consistently mapped as the nations' most socioeconomically vulnerable were the Central Valley areas of California, the landlocked counties at the Southern tip of Florida, and the Texas Colonias (i.e. along the U.S.-Mexico border). While the most impoverished regions of California were hit by a fair number of extreme weather events, it did not necessarily translate into large number of casualties, relative to other parts of California (**Figure 9** and **Figure 10**, see Appendix). An in-depth look into Texas also showed that regardless of the frequency of weather events, most counties recorded either zero or near-zero fatality rates (**Figure 9** and **Figure 10**). Although it would require more data about individual climate events, it may be helpful to examine the *types* and *magnitudes* of natural disasters that had a blow on these regions, in tandem with the *frequency* of such incidents. Lastly, consistent with our previous finding vis-a-vis limited exposure to climate disasters among the socioeconomically vulnerable, the inland areas of South Florida experienced fewer climate catastrophes than the affluent parts of coastal Florida (**Figure 9** and **Figure 11**). Reasons why the wealthier population settle down in more disaster-prone areas cannot be determined from the data alone. However, one possible explanation could be that beachfront properties overlooking the Gulf or the Atlantic are top destinations for the rich because of their natural wonders, but that coastal areas are simultaneously more susceptible to flood, tropical storms, and hurricanes.

CONCLUSION

Limitations of the study

Taken together, the greatest limitation of this study lies in its poor predictive power. To improve the accuracy of the model, I could have employed more flexible machine learning techniques (e.g. neural networks) to better account for the complexity of the relationship. However, a post-hoc analysis revealed that such alternatives are unlikely to work. Though somewhat counterintuitive, **areas with the highest rates of casualty were *not* the communities that have been historically underserved**. In fact, those that **fared *worse*** in the wake of climate disasters turned out to be **the communities that were *the least socioeconomically vulnerable,*** a result of which was attributable to <u>more frequent episodes of extreme weather events</u>.

In future studies, we could examine climate-related casualties over a longer time span, given that there were–*on average*–few incidents of casualty throughout the last 4 years. With only a handful of incidents that could be fed as inputs for training, algorithms may have failed to spot an underlying pattern in climate-related deaths and injuries.

Empirical and theoretical implications

Notwithstanding these limitations, the current study offers insights into disaster risk management strategies as well as avenues for future research. Emergency response planners often prioritize communities with higher levels of socioeconomic vulnerability, on the presumption that they are the least prepared for hazards and the least resilient. However, our findings point to the fact that **even the wealthiest communities are not immune to, if not more adversely impacted by climate events**. Hence the practical implication of this study is in the **importance of targeting the at-risk communities, which *may <u>not</u> necessarily <u>conform to the conventional ideas of 'vulnerability</u>'*.**

Another key takeaway, which is more in line with the traditional notions of social vulnerability, is that **resources could be directed at areas with limited access to healthcare**. The CDC's preexisting social vulnerability framework was unable to predict climate-induced health risks in itself. But additional measures of vulnerability which I incorporated into the model–namely *proximity to the nearest emergency ward*–were a consistently strong predictor of casualty for both models that were tested. In minimizing the loss of lives, public health officials would benefit from prioritizing areas with remote access to urgent medical care.

With respect to its theoretical implications, **future research should explore the risks of casualty across different minority groups**, by looking at *<u>individual</u>* racial and ethnic categories, rather than a *composite* measure of minority status. Again, the CDC's preexisting social vulnerability framework was inadequate, in accounting for the disproportionate impacts borne by certain subset of the population. Although there was inconclusive evidence to suggest *which particular race/ethnicity* was more at risk than the other, findings did imply that *<u>not all</u>* racial/ethnic minorities demonstrate comparable rates of mortality and injury. Thus it calls for a more *<u>granular look</u>* into how the immediate health consequences of climate disasters intersect with race and ethnicity.

In summary, the mix of unforeseen and anticipated findings speaks to the complexity of the phenomenon at hand. Although the relationship between socioeconomic factors and climate-related health risks remains open to question, it warrants further investigation given the ever-increasing impacts of climate change. As was highlighted in this study, the devastating impacts of climate change is not specific to minority or the underserved groups; climate hazards will have implications all across the board, for both the rich and the poor.

REFERENCES

Agency for Healthcare Research and Quality. (2022a). *AHRQ Social Determinants of Health Database Data Source Documentation* [Data dictionary]. https://www.ahrq.gov/sites/default/files/wysiwyg/sdoh/SDOH-Data-Sources-Documentation-v1-Final.pdf

Agency for Healthcare Research and Quality. (2022b). *Social Determinants of Health Database* [Data set]. https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html

Benda, N. C., Veinot, T. C., Sieck, C. J., & Ancker, J. S. (2020). Broadband internet access is a social determinant of health!. *American Journal of Public Health*, *110*(8), 1123-1125. https://doi.org/10.2105/AJPH.2020.305784

Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. (2022a). *CDC/ATSDR Social Vulnerability Index 2020 Database US* [Data set]. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. (2022b). *CDC/ATSDR Social Vulnerability Index 2020 Documentation* [Data dictionary]. https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/pdf/SVI2020Documentation_08.05.22.pdf

Cutter, S. L., & Emrich, C. T. (2017). *Social vulnerability index (SoVI®): Methodology and limitations*. Hazards and Vulnerability Research Institute. https://www.vulnerabilitymap.org/Portals/0/Files/Social%20Vulnerability%20Index%20SoVI%20-%20Methodology%20and%20Limitations.pdf?ver=Dtf6UG34OrlIsrehg3sjLQ%3D%3D

Fang, Y., Li, T., & Zhao, H. (2022). Random Forest Model for the House Price Forecasting. *2022 14th International Conference on Computer Research and Development (ICCRD),* 140-143. https://doi.org/10.1109/ICCRD54409.2022.9730190

Klinenberg, E. (1999). Denaturalizing disaster: A social autopsy of the 1995 Chicago heat wave. *Theory and society*, *28*(2), 239-295.

Muthukrishnan, R., & Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 18–20. https://doi.org/10.1109/ICACA.2016.7887916

National Oceanic and Atmospheric Administration. (2021). *National Weather Service Instruction (NWSI) 10-1605*. https://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf

National Oceanic and Atmospheric Administration. (2023). *U.S. Billion-Dollar Weather and Climate Disasters*. https://doi.org/10.25921/STKW-7W73

Shen, X., Yin, F., & Jiao, C. (2023). Predictive Models of Life Satisfaction in Older People: A Machine Learning Approach. *International Journal of Environmental Research and Public Health*, *20*(3), 2445. https://doi.org/10.3390/ijerph20032445

Shonkoff, S., Morello-Frosch, R., Pastor, M., & Sadd, J. (2009). Environmental health and equity impacts from climate change and mitigation policies in California: a review of the literature. *California, USA: California Environmental Protection Agency*.

Singh, B., Sihag, P., & Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment*, *3*(3), 999-1004. https://doi.org/10.1007/s40808-017-0347-3

Smith, A. (2022, January 24). 2021 U.S. billion-dollar weather and climate disasters in historical context. *Beyond the Data*. http://www.climate.gov/news-features/blogs/beyond-data/2021-us-billion-dollar-weather-and-climate-disasters-historical

Smith, A. (2023, January 10). 2022 U.S. billion-dollar weather and climate disasters in historical context. *Beyond the Data*. https://www.climate.gov/news-features/blogs/2022-us-billion-dollar-weather-and-climate-disasters-historical-context

U.S. Census Bureau (2021, November 22). *Group Quarters and Residence Rules for Poverty*. U.S. Department of Commerce. Retrieved April 6, 2023, from https://www.census.gov/topics/income-poverty/poverty/guidance/group-quarters.html

U.S. Census Bureau. (2022, August 25). *When to Use 1-year or 5-year Estimates*. U.S. Department of Commerce. Retrieved March 2, 2023, from https://www.census.gov/programs-surveys/acs/guidance/estimates.html

APPENDIX

Section I. A country-level overview of climate hazards and their impacts on health

Figure 9. The frequency of climate hazards, across the U.S.

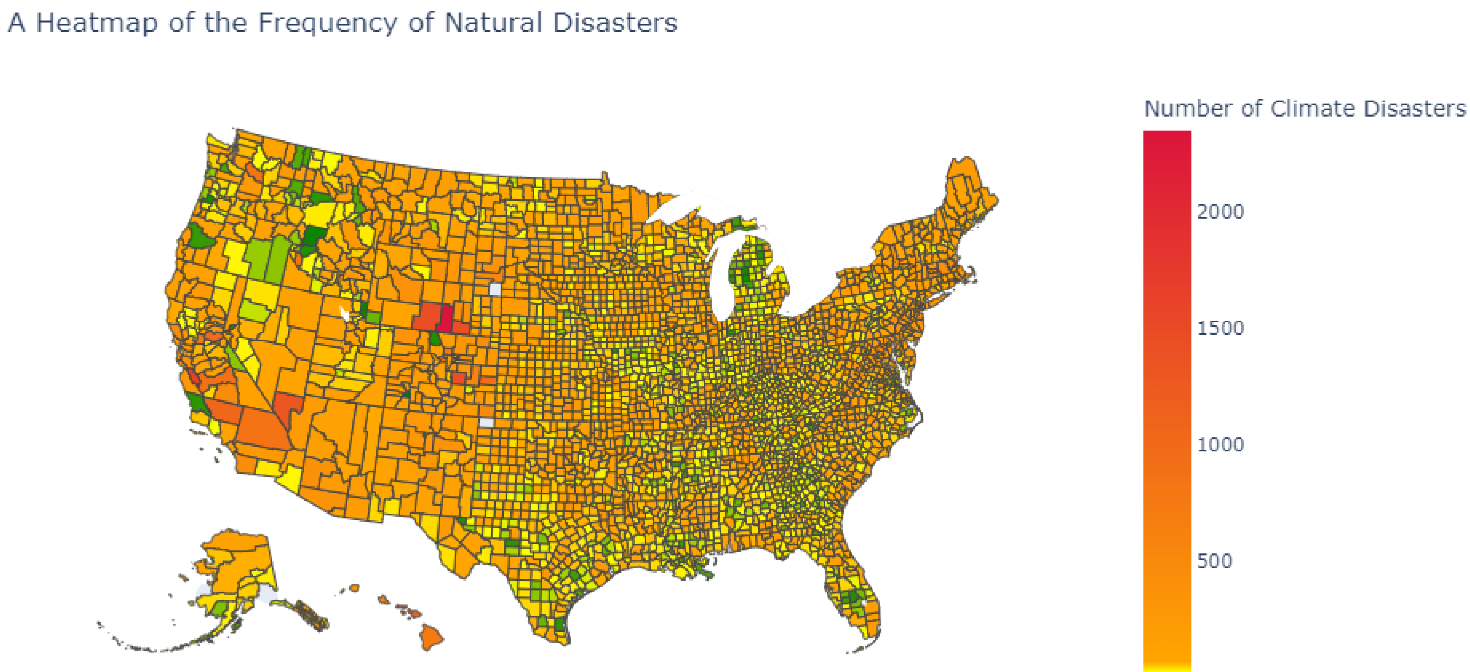
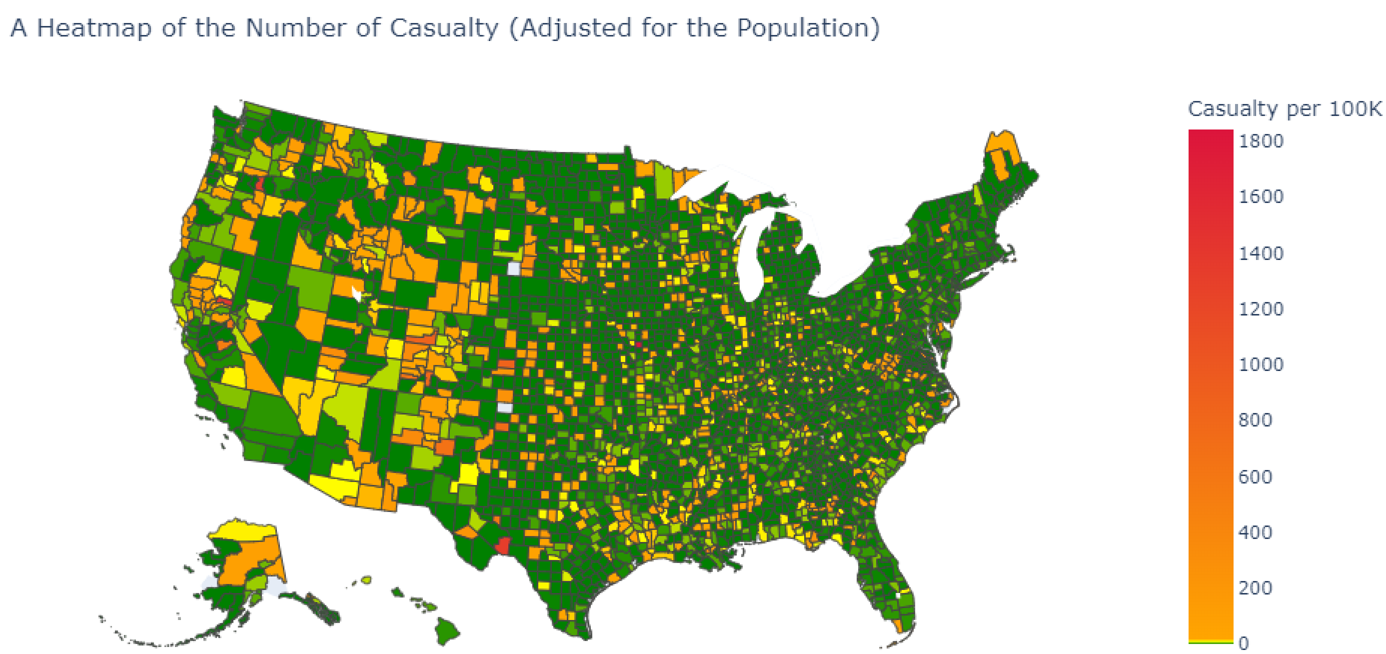
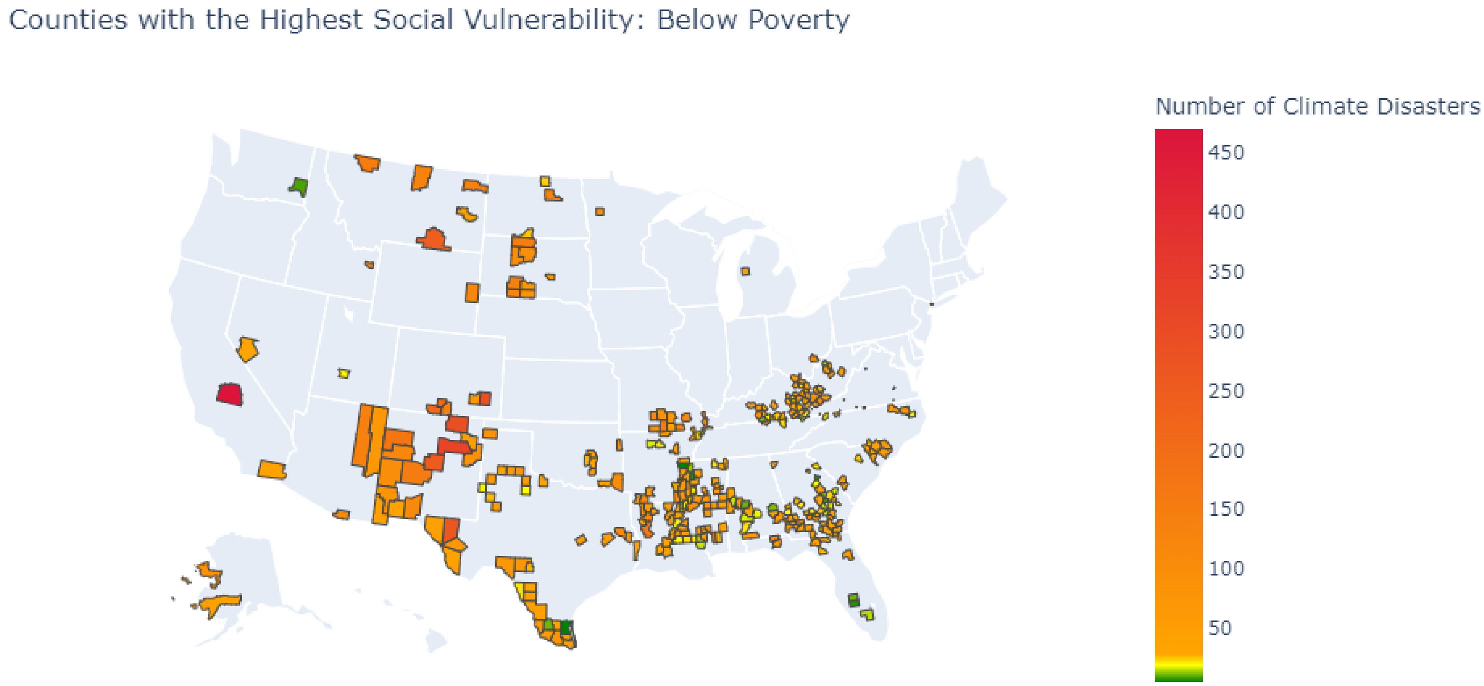


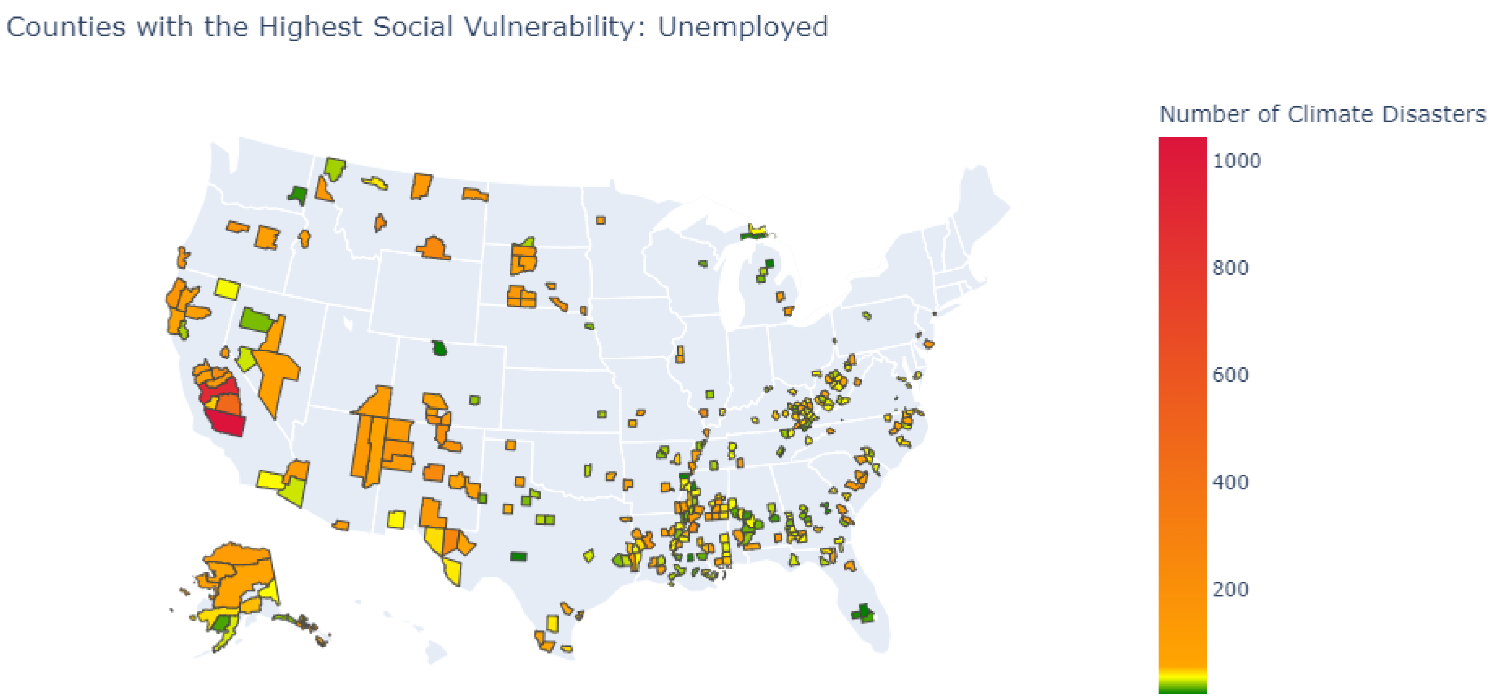
Figure 10. Casualty caused by climate hazards, across the U.S.

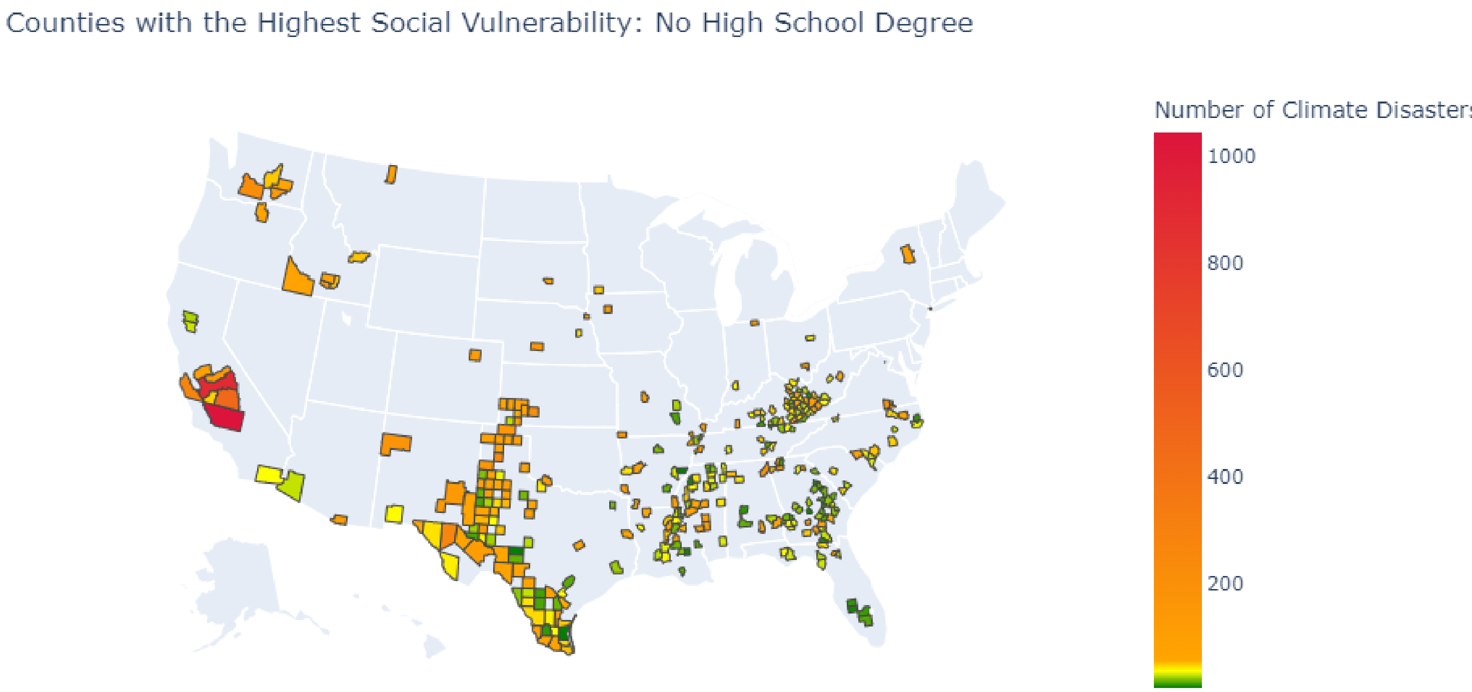


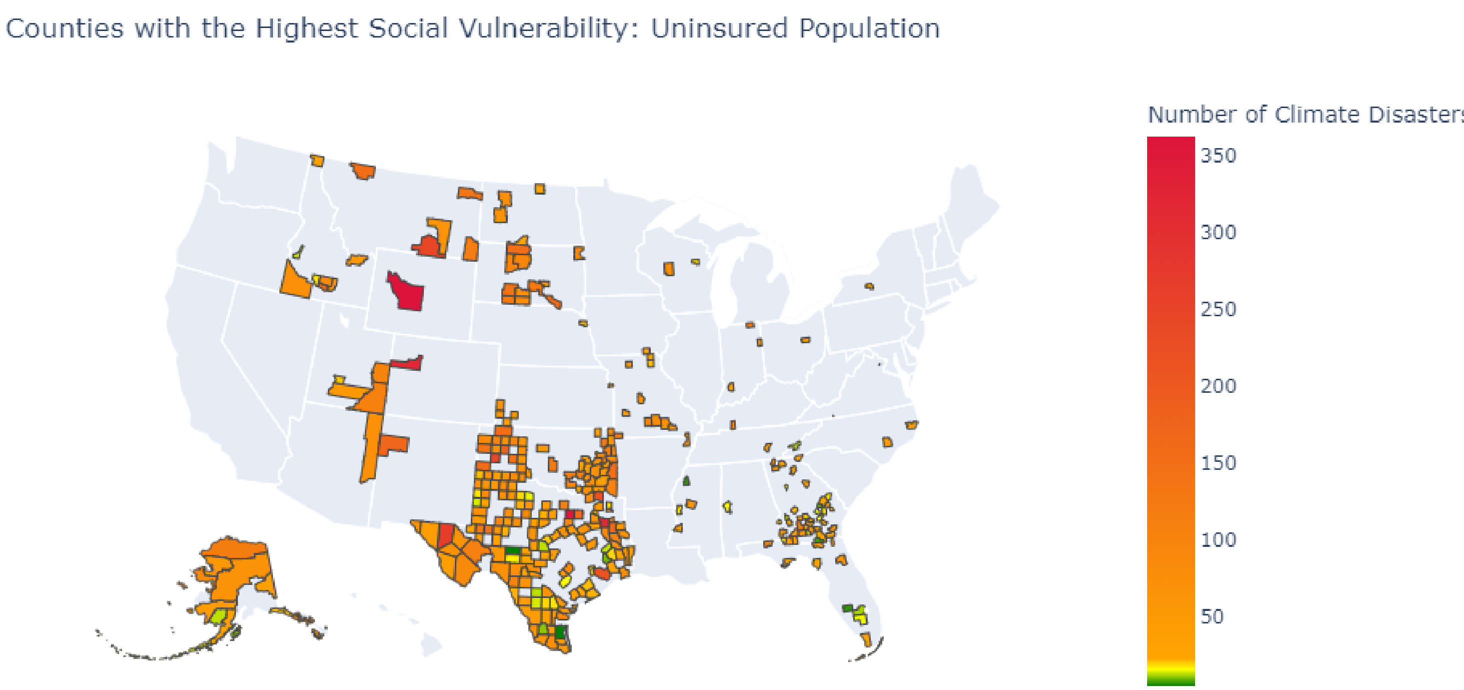
Section II. A closer look at U.S. counties with the highest level of socioeconomic vulnerability

Figure 11. Exposure to climate hazards, among top 10% of America's most socially vulnerable counties

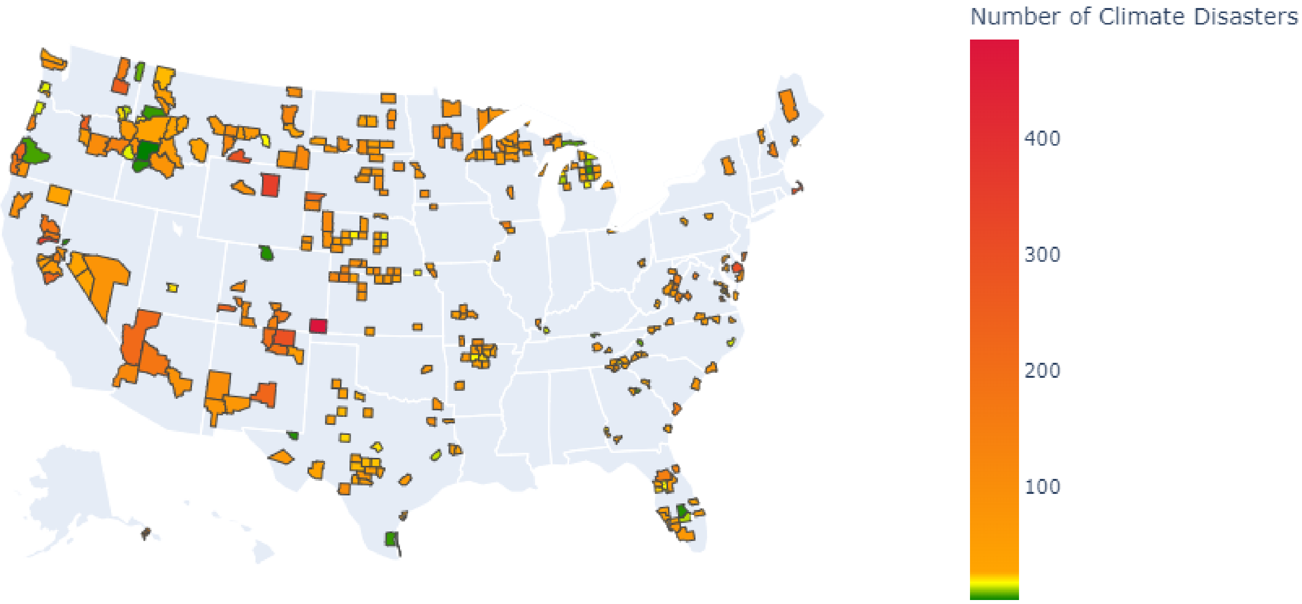




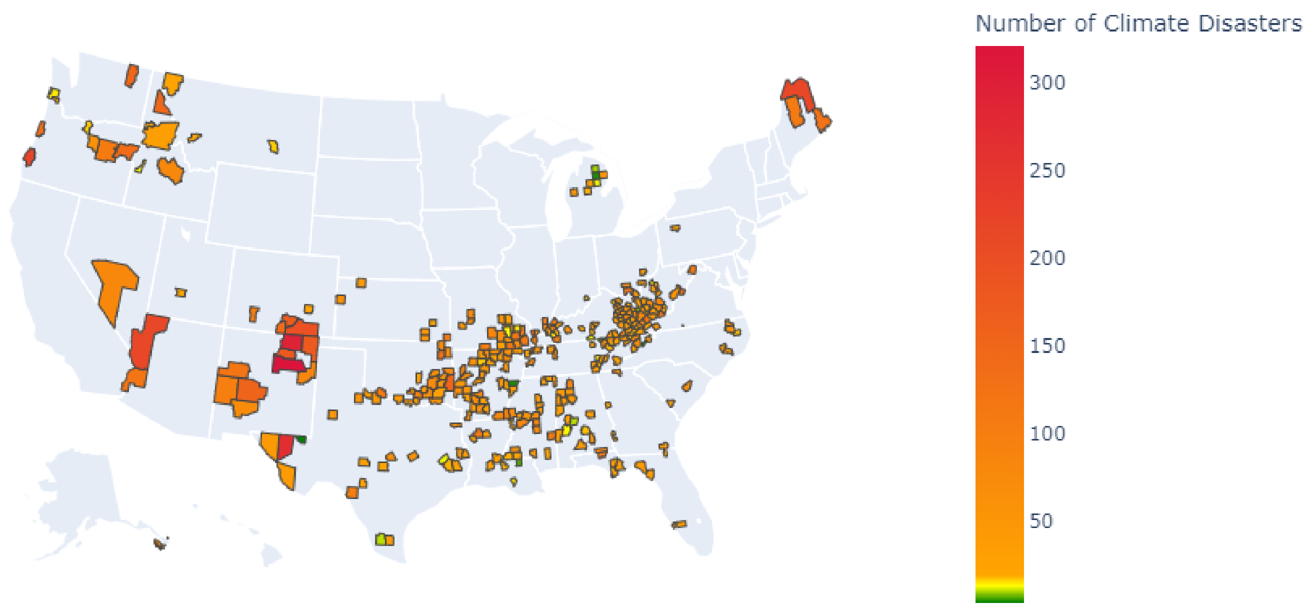

Counties with the Highest Social Vulnerability: No High School Degree
Number of Climate Disasters
1000
800
600
400
200


Counties with the Highest Social Vulnerability: Uninsured Population
Number of Climate Disasters
350
300
250
200
150
100
50

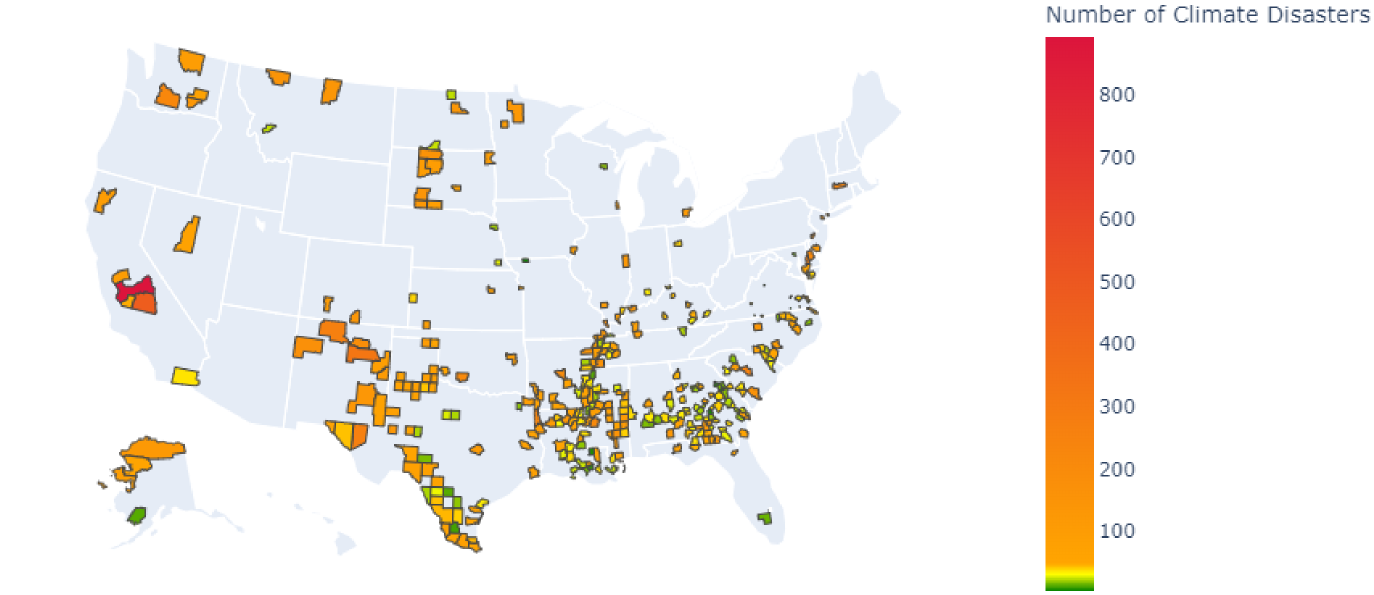
Counties with the Highest Social Vulnerability: Age 65 & Older



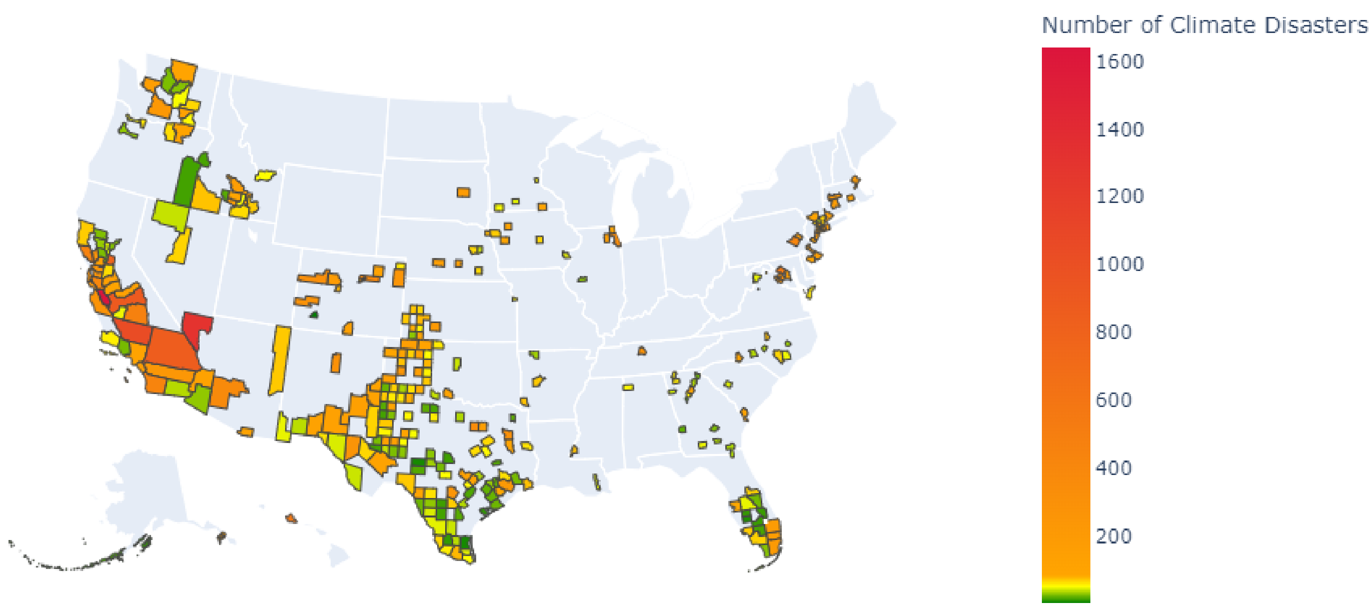
Counties with the Highest Social Vulnerability: Disabled Population



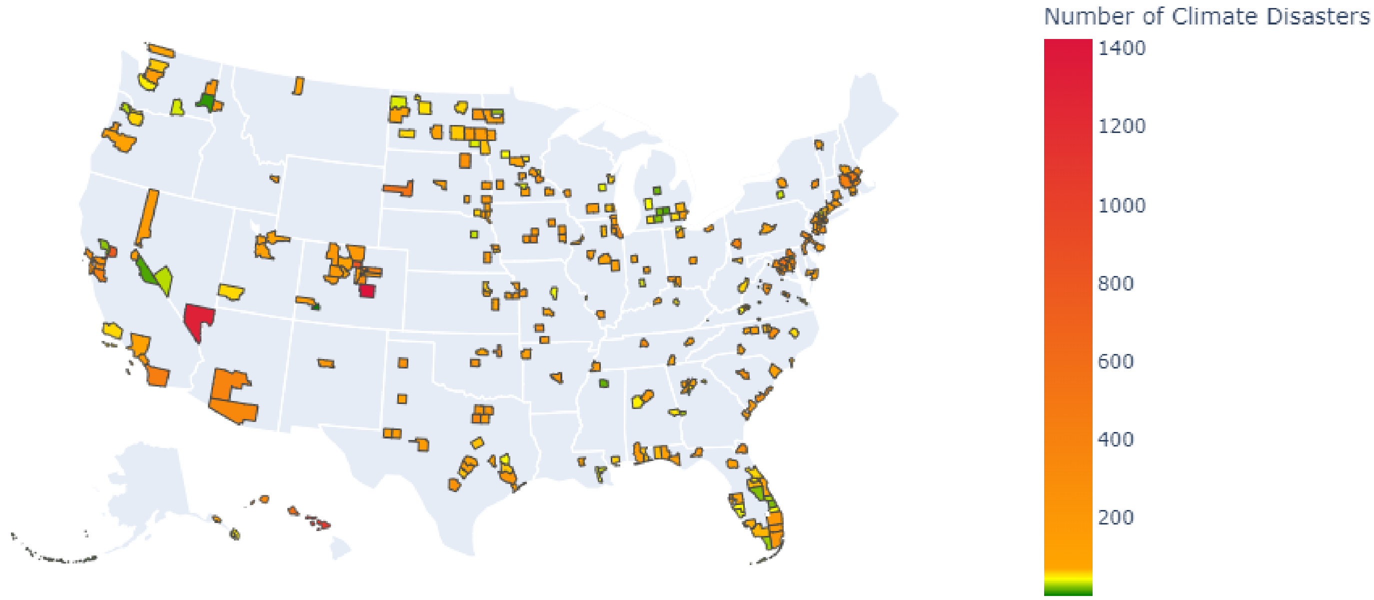
Counties with the Highest Social Vulnerability: Single Parent



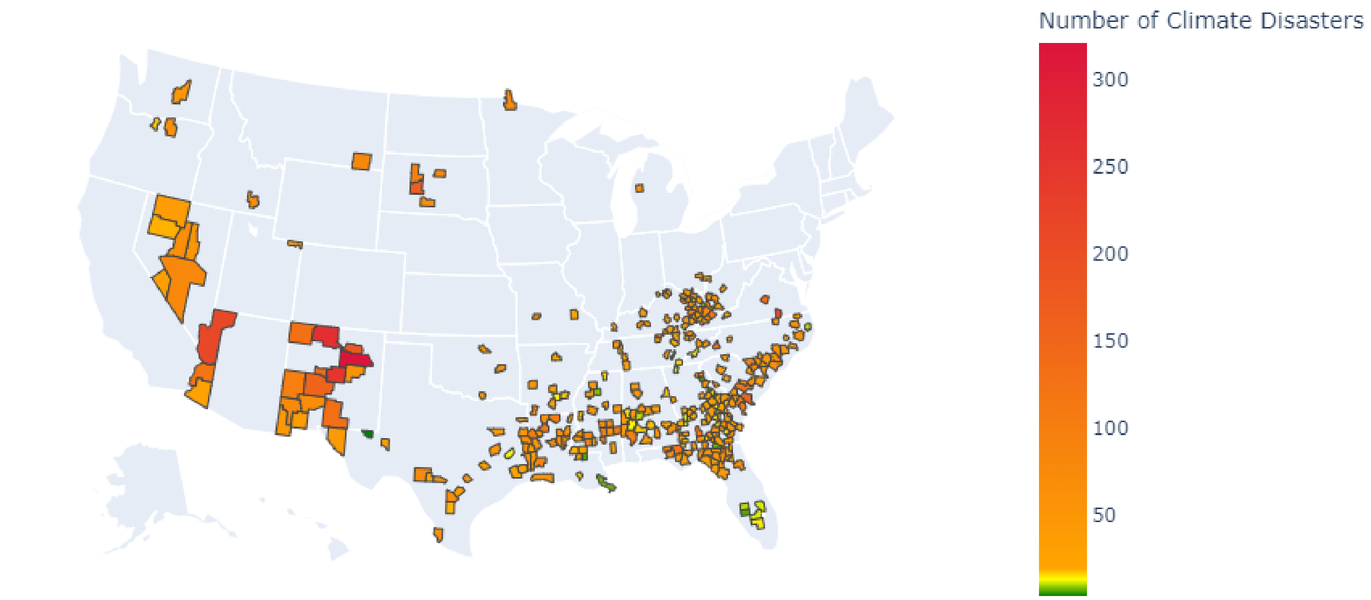
Counties with the Highest Social Vulnerability: Limited English Proficiency



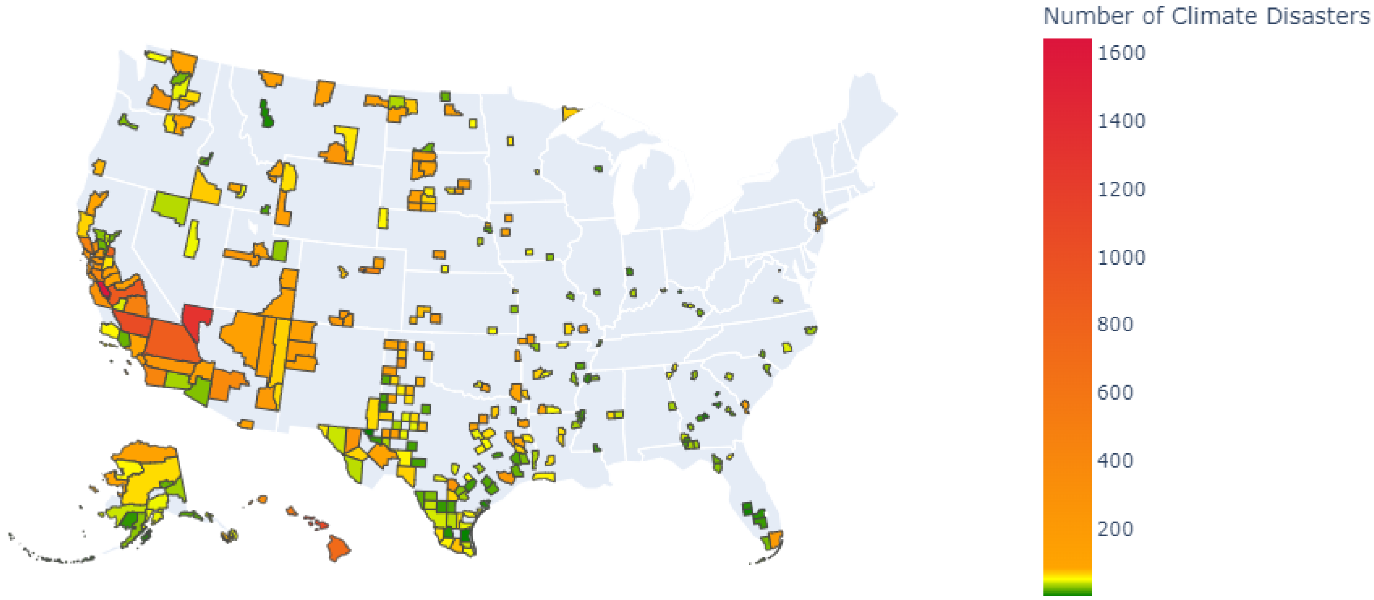
Counties with the Highest Social Vulnerability: Multiunit Housing



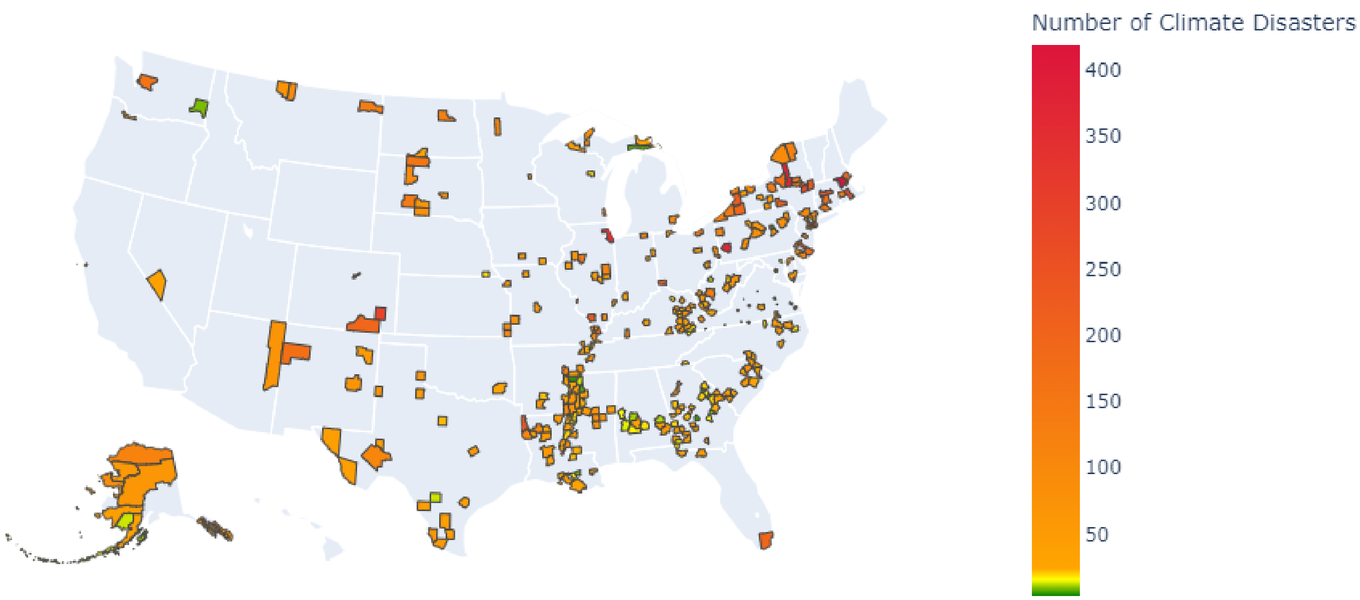
Counties with the Highest Social Vulnerability: Mobile Homes



Counties with the Highest Social Vulnerability: Crowded Housing



Counties with the Highest Social Vulnerability: No Vehicles



Counties with the Highest Social Vulnerability: Group Quarters