2022 COMMUNITY RESILIENCE ESTIMATES FOR HEAT

Quick Guide

Updated July 2024

Small Area Estimates Program

Social, Economic, and Housing Statistics Division

U.S. Census Bureau, Department of Commerce

Estimates have been reviewed to ensure that no confidential information is disclosed.
DMS P-7517412; CBDRB-FY24-0362.

Contents

Overview

Community resilience describes the capacity of individuals and households within a community to absorb a disaster's external stressors. The standard [Community Resilience Estimates (CRE)] measures a community's social vulnerability to natural disasters. However, the social vulnerabilities to *extreme heat exposure* differ from other natural disasters. As a result, the CRE Team created a new set of estimates called the Community Resilience Estimates for Heat (CRE for Heat).

In collaboration with Arizona State University's Knowledge Exchange for Resilience (KER), the CRE Team produced the 2022 CRE for Heat using data on individuals and households. The data sources include the 2022 American Community Survey (ACS), the 2021 American Housing Survey, the Census Bureau's Population Estimates Program (PEP), and the 2020 Census.

Local planners, policymakers, public health officials, and community stakeholders can use the CRE for Heat to assess their community's vulnerability to extreme heat.

Background

The CRE provides an easily understood social vulnerability metric to disasters including hurricanes, floods, and pandemics. As an experimental data product, it was published to collect feedback from data users and stakeholders on the quality and usefulness of the new product.

After the publication of the CRE, Arizona State University's (ASU) Knowledge Exchange for Resilience (KER) reached out to the CRE team to further discuss the product. KER was particularly interested in how the CRE could measure social vulnerability to extreme heat. KER and ASU are well known for their research on heat exposure and emergency management. The CRE Team consulted with the KER to adjust the CRE's social vulnerability components relevant to heat exposure.

The 2019 CRE for Heat was released in April 2023 as an experimental data product to collect feedback. KER and other stakeholders provided recommendations for improvement which were incorporated into the current data release. The 2022 CRE for Heat was released in July 2024 as an experimental dataset to gather further feedback for potential enhancements.

What's New?

Based on feedback on the experimental 2019 CRE for Heat, two changes were incorporated into the current experimental data product, the 2022 CRE for Heat.

One, "households potentially lacking air conditioning" was added as a new component of social vulnerability. Since the American Community Survey does not contain detailed data on this topic, we modeled these estimates using the American Housing Survey and machine learning techniques. See the section "Modeling Households Potentially Lacking Air Conditioning" for more details.

Two, a measure of exposure to extreme heat alongside the CRE for Heat estimates was added. Not all socially vulnerable communities are equally exposed to extreme heat. Pairing the CRE for Heat estimates with heat exposure data provides a more comprehensive look at social vulnerability to heat. In the 2022 CRE for Heat dataset, an area is considered exposed to extreme heat if it meets one of two criteria:

- Areas where the maximum air temperature has reached or exceeded 90 degrees Fahrenheit for two or more days in a row during 2022.[1]
- Areas where estimated wet bulb globe temperature has reached or exceeded 80 degrees at any time during 2022.[2]

See the "Exposure Data" section for more details about the extreme heat data.

Data

The American Community Survey (ACS) is a nationally representative survey with data on the characteristics of the population of the United States. The sample is selected from all counties and county equivalents and has a sample size of about 3.5 million housing units yearly. It is the best source for detailed population and housing data about our nation and its communities. We use individual and household-level ACS data to determine the population estimate of the components of social vulnerability for individuals.

The American Housing Survey (AHS) is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The survey is the most comprehensive national housing survey in the United States. Using data from the 2021 AHS, we created a machine learning model that determines if a household in the ACS potentially lacks an air conditioning unit. Additional data was also used as predictors in the modeling process. The "Modeling Households Potentially Lacking Air Conditioning" section explains the data and our machine learning methodology in more detail.

We also use auxiliary data from the Population Estimates Program (PEP), which is the Census Bureau's program that produces and publishes estimates of the population living at a given time within a geographic entity in the U.S. and Puerto Rico. We use population data from the PEP by age group, race and ethnicity, and sex. Since the PEP data does not go down to the census tract level, the CRE also uses the Public Law 94-171 summary files (PL94) and Demographic Housing Characteristics (DHC) tables from the 2020 Decennial Census to produce the base estimates.

Once the weighted estimates are tabulated, small area modeling techniques are used to create the CRE estimates.

---

[1] Extreme Heat Data. https://www.ready.gov/heat

[2] Wet Bulb Globe Temperature from the National Weather Service. https://www.weather.gov/tsa/wbgt

Components of Social Vulnerability

Resilience to a disaster is partly determined by the social vulnerabilities within a community. To measure these vulnerabilities and construct the CRE, we designed population estimates based on individual and household-level components of social vulnerability based on ACS data. The components are binary and add up to 11 possible indicators.

The specific ACS-defined measures we use are as follows:

<u>Components of Social Vulnerability (SV) for Households (HH) and Individuals (I)</u>

- SV 1 (HH): Financial hardship defined as:
    - Income-to-Poverty Ratio (IPR) < 130 percent or
    - 50% < for housing/rental costs (HH).
- SV 2 (HH): Single or zero caregiver household - only one or no individuals living in the household who are 18-64.
- SV 3 (HH): Housing quality described as:
    - Unit-level crowding with > 0.75 persons per room or
    - Live in mobile home, boat, RV, Van, or other.
- SV 4 (HH): Communication barrier defined as either:
    - Limited English-speaking households[3] or
    - No one in the household has a high school diploma.
- SV 5 (HH): No one in the household is employed full-time, year-round. The flag is not applied if all residents of the household are aged 65 years or older.
- SV 6 (I): Disability posing constraint to significant life activity, defined as:
    - Persons who report having any one of the six disability types: hearing difficulty, vision difficulty, cognitive difficulty, ambulatory difficulty, self-care difficulty, and independent living difficulty.
- SV 7 (I): No health insurance coverage.
- SV 8 (I): Being aged 65 years or older.
- SV 9: Transportation exposure described as:
    - No vehicle access (HH) or
    - Work commuting methods with increased exposure to heat (e.g., public transportation, bicycle, walking) (I).
- SV 10 (HH): Households without broadband Internet access).
- SV 11 (HH): Households that potentially lack air conditioning.

It is important to note that components one, three, four, and nine are <u>not</u> double flagged. If a household is linguistically isolated and no one has attained a high school diploma or more education, the people in that household are only flagged once.

[3] A “Limited English-speaking household” is one in which no member 14 years old and over (1) speaks only English at home or (2) speaks a language other than English at home and speaks English “Very well.”

For the Transportation SV component, which could have a mix of household and individual flags, members of the same household can have different values for the SV component. This happens only under certain conditions. If a household has no access to a vehicle, then everyone in the household is flagged as having the transportation SV component despite anyone's individual commuting choices in the household.

On the other hand, the transportation SV component can differ among individuals in the same household if the household has access to at least one vehicle BUT one or more working persons in the household is exposed to heat through their work commute. For example, let's say a given household has access to a vehicle with two employed individuals. Suppose one person walks to work and the other person drives a car. The person who walks to work would have the transportation SV component, but all other members would not.

The result is a tabulation of aggregate-level (tract, county, and state) small area estimates. The CRE provides an estimate for the number and percentage of people with a specific number of vulnerabilities within a given geographic area. In the current data file layout, the estimates are categorized into three groups: 0 components of social vulnerability, 1-2 components, and 3+ components.

Differences with Standard CRE

Although the methodology is the same as the standard CRE data products, there are some differences in the indicators used to create the CRE for Heat.[4] Three (SV 1, 3 and 9) indicators are modified from the standard CRE to account for vulnerability to heat exposure related to housing and transportation.

For SV 1, the standard CRE had one poverty indicator (income-to-poverty ratio < 130). The CRE for Heat's indicator also includes whether the household's housing costs are more than 50%.

For SV 3, the CRE only used unit-level crowding (> 0.75 persons per room) to measure housing quality. The CRE for Heat also includes housing structure type (living in a mobile home, boat, RV Van, or other) since individuals in these housing types are more exposed to extreme heat.

For SV 9, the CRE has one indicator for transportation – whether the household has access to a vehicle. The CRE for Heat's transportation indicator also captures an individual's commute type (i.e., commuters that use public transit, walking, biking, or other non-personal vehicle methods).

Lastly, SV 11 is a new component for the CRE for Heat, "Households that potentially lack air conditioning", which is not in the standard CRE. The process of adding this component is discussed further in the next section, "Modeling of Households Potentially Lacking Air Conditioning." The remaining seven indicators are the same as the standard CRE.

---

[4] Methodology for standard Community Resilience Estimates: https://www.census.gov/programs-surveys/community-resilience-estimates/technical-documentation/methodology.html

Modeling Households Potentially Lacking Air Conditioning

Experimental data products from the Census Bureau are innovative statistical products created using new data sources or methodologies which can benefit data users in the absence of other relevant products. One goal of experimental products is to seek feedback from data users and stakeholders on the quality and usefulness of the new products. As an experimental data product, the CRE team received feedback on the first version of the CRE for Heat.

One suggestion received to strengthen our estimates was to have households that did not have air conditioning as a component of social vulnerability. To determine if a household potentially lacked an air conditioning unit, we modeled air conditioning data from the AHS and other sources onto the ACS data using machine learning techniques. We call this method "Cross Survey Modeling."

The methodology entails the harmonization of data between the ACS and the AHS, by collapsing categories for variables in the AHS and ACS data so that they matched one another. We also incorporated geographic data – which are uniform for all housing units in an area – on topics such as weather, utility costs and demographics. The following variables were tested and utilized in the creation of the model:

From the American Housing Survey:

- Year structure built
- Household structure type
- Household income
- Whether the housing unit utilizes electricity as its heating source
- Tenure (collapsed to owner or renter)
- Total rooms in the household
- Education level of householder

Geographic data:

- Wet bulb temperature.[5] July average. Calculated from National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) data as a function of air temperature and relative humidity according to Stull (2011).[6]
- Average electricity cost (state level) from the U.S. Energy Information Administration.
- Percentage of the county that is rural from the CRE for Equity Supplement.
- Coastal county (indicator) from the Census Bureau.

---

[5] Definition of wet bulb temperature from the National Weather Service. "The Wet Bulb temperature is the temperature of adiabatic saturation. This is the temperature indicated by a moistened thermometer bulb exposed to the air flow." https://www.weather.gov/source/zhu/ZHU_Training_Page/definitions/dry_wet_bulb_definition/dry_wet_bulb.html#:~:text=The%20Wet%20Bulb%20temperature%20is,bulb%20wrapped%20in%20wet%20muslin.

[6] Stull, R. (2011). Wet-Bulb Temperature from Relative Humidity and Air Temperature. *Journal of Applied Meteorology and Climatology*, *50*(11), 2267-2269. https://doi.org/10.1175/JAMC-D-11-0143.1

- Climate zones from the U.S. Department of Energy Building America program.

We used extreme gradient-boosted trees (xgboost) to train the model because that class of models performed best in initial testing. We split the data into an 80/20 training and test split and implemented 5-fold cross-validation. We up-sampled the training set to correct for an imbalanced outcome variable (more than 90 percent of households in the AHS have air conditioning). The up-sampling involved random draws from the minority class (households with no AC) with replacement until the number of sample elements matched the majority class (households with an AC).

We then performed hyperparameter tuning to select the model that maximizes accuracy but balances the true positive and true negative rate. The number of trees was fixed at 1,000. The following parameters were tuned using a grid search:

- Tree depth: This parameter controls the maximum number of levels in each decision tree. Deeper trees can model more complex relationships but may also lead to overfitting.
- min_n: This parameter specifies the minimum number of samples required to create a new node in a tree. A higher value can result in more generalized models, whereas a lower value may capture finer details in the data.
- loss_reduction: This parameter sets the minimum reduction in loss required to make a further split in a tree node. It helps to prevent overfitting by avoiding splits that do not significantly improve the model's performance.
- sample_size: This parameter determines the proportion of the total observations used to train each tree. Using a subset for each tree (bootstrapping) increases model diversity and robustness.
- mtry: This parameter defines the number of features to consider when making each split in a tree. It controls the degree of feature randomness in model building, impacting both model accuracy and overfitting.
- learn_rate: This parameter affects the rate at which the model learns by scaling the contribution of each tree. A smaller learning rate requires more trees to converge but typically leads to a more stable model that is less prone to overfitting.

The final model had the following performance metrics:

- Accuracy (the proportion of predictions that match the self-reported value): 0.84
- Sensitivity (the true positive rate): 0.85
- Specificity (the true negative rate): 0.74

We applied the predictive model to the ACS using the parameters from the best-fitting XGBoost model. The model calculated probabilistic predictions for each household record on the ACS. We constrained the predictions to harmonize with state-level data from the Residential Energy Consumption Survey (RECS) from the Energy Information Administration. We did this by ranking housing units by their predicted probability of having an air conditioning unit within each state. We then converted the probability for each household to a binary classification (AC versus no AC) until the state's overall percentage matched the RECS estimates. In this way, the publicly available RECS estimates act as an aggregate threshold for converting probabilistic predictions into binary classifications.

To ensure the reasonableness of our predictions, we applied spatial smoothing at the county level to adjust for local prediction noise. We applied spatial smoothing by fitting a spatial autoregressive at the county level. This model took as input a weighted proportion of households in each county that likely had an AC. These proportions represented a weighted sum (using the ACS housing unit weight) of households with an air conditioner divided by the total number of households in the county. We fed these proportions into a spatial autoregressive model that did not have a predictor but had newly constructed weights to account for two features:

1. Spatial weights from each county's five nearest neighboring counties (within the state).
2. The square root of the population size for each county, so that larger counties contribute more to the weights.

After we obtained new county-level estimates, we repeated the process of "thresholding" predictions that had previously been used to constrain the prediction to match RECS estimates at the state level. This time, rather than thresholding predictions to RECS, we applied the threshold method at the county level and classified households as having an air conditioner until the county-level proportion matched the spatially smoothed aggregate estimates. This step was performed independently for each state.

This methodology has allowed us to add "Households Potentially Lacking Air Conditioning" as a new component of social vulnerability. As part of the experimental data product series, we welcome feedback on this methodology and will explore further uses of cross-survey modeling.

We found that wet bulb temperature, household income, and climate zone were the three most important predictors of whether a household potentially lacks air conditioning.

Exposure Data

The CRE team also received feedback regarding the utility of adding information about heat exposure. Some data users mentioned that they expected the results to combine vulnerability and the likelihood of exposure. The 2019 version of the CRE for Heat only examined social vulnerability to an extreme heat event.

The literature on community resilience states that vulnerability and exposure should be understood together.[7] The Knowledge Exchange for Resilience at ASU defines "The Resilience Equation" as:

$$Risk(Shock) = \frac{Exposure * Vulnerability}{Adaptive\ Capacity * Social\ Cohesion}$$

For the 2022 CRE for Heat, we include exposure data alongside the CRE estimates to connect extreme heat exposure with social vulnerability. Our exposure data in the 2022 CRE for Heat is based on two criteria: air temperature and wet bulb temperature. These criteria were developed after reviewing various definitions of exposure from government agencies and the availability of historical climate data. An area is considered exposed to extreme heat if it meets at least one of the heat exposure criteria:

- Areas where the maximum air temperature has reached or exceeded 90 degrees Fahrenheit for two or more days in a row during 2022.[8]
- Areas where estimated wet bulb temperature has reached or exceeded 80 degrees at any time during 2022.[9]

The estimates for air temperature and wet bulb temperature are derived from the NCEP North American Regional Reanalysis (NARR) data provided by the NOAA Physical Sciences Laboratory (PSL), Boulder, Colorado, USA, from their website at https://psl.noaa.gov.

The NARR data provide eight times daily measurements of temperature and relative humidity for North America on a high-resolution grid. For each day of 2022, we obtained all the eight times daily measurements for both two-meter air temperature and relative humidity on every grid cell, which was approximately 0.3 degrees (32km) resolution at the lowest latitude on a Northern Lambert Conformal Conic grid.

To establish whether a geographic area exceeded the air temperature exposure threshold, we obtained the maximum daily air temperature for each gridded cell and flagged the cells where the maximum temperature exceeded 90 degrees Fahrenheit for two consecutive days. We then performed a spatial join from the gridded data over Census geographies, such as counties, so that each unit of geography inherited the value of the largest grid cell in the spatial overlap.

For wet bulb temperature, we obtained the eight times daily air temperature and relative humidity and then calculated wet bulb temperature.[10] The calculation provided eight times daily wet bulb temperature readings, which we used to determine whether a gridded cell ever

[7] Solis, P. (2023). Extreme Heat, Health, and Housing in Urban Maricopa County Arizona: a story of exchanging knowledge to build community resilience. https://keep.lib.asu.edu/items/193814
[8] Extreme Heat Data. https://www.ready.gov/heat
[9] Wet Bulb Globe Temperature from the National Weather Service. https://www.weather.gov/tsa/wbgt
[10] Stull, R. (2011). Wet-Bulb Temperature from Relative Humidity and Air Temperature. *Journal of Applied Meteorology and Climatology*, *50*(11), 2267-2269. https://doi.org/10.1175/JAMC-D-11-0143.1

reached or exceeded a wet bulb temperature of 80 degrees Fahrenheit at any point during the year. We attached the wet bulb temperature readings to Census geographies using the same spatial join method as in the air temperature threshold calculation.

Further Information

Community Resilience Estimates Program Website
<https://www.census.gov/programs-surveys/community-resilience-estimates.html>

Technical Help
<SEHSD.CRE@census.gov>