Capstone 2 Project Proposal

By Kealie Pretzlav

Problem Statement

Floods occur every year all over the United States, impacting human life, local economics, and infrastructure. Climate change projections indicate that flood will likely increase in frequency and magnitude in the coming years. Flood infrastructure updates can typically only occur in areas where resources are available. Areas with higher poverty rates are typically overlooked or under-resourced, disproportionately impacting underserved populations further perpetuating poverty cycles. In this analysis, I will combine a national flood dataset with US Census data on household income and public assistance amounts by zip code tabulation areas (ZTCAs) to answer the following questions:

1. Is flood occurrence a good predictor of poverty or household income?
2. Are accelerating or increasing flood impacts (i.e. more recent or large floods) a good predictor of poverty?
3. How sensitive is poverty to changes in flood occurrence? Will increases in flood occurrence impact poverty?
4. Do the answers to the above questions vary considerably by state?

The goal of this analysis is to use a sensitivity analysis to quantify the impacts of changing climate on poverty status, predicting which communities will be most vulnerable to climate change-induced flood impacts, so that flood infrastructure resources can be more efficiently allocated to help the people it will benefit the most.

Datasets

To complete this project, I used two primary data sources, data from the US Census and storm event data from the National Oceanic and Atmospheric Administration (NOAA).

US Census Data

The poverty data are from the American Community Survey 5-year dataset ending in 2022 which represents 5-year average estimates for all variables from 2018 to 2022. These data are downloaded for the Zip Code Tabulation Area (ZCTA) geography. I'll begin the analysis with median household income, each quintile of household income, and the household income of the top 5 percent, and household public assistance amount (social security, food stamps/SNAP, etc.).

I will use the US Census ZCTA shapefiles to map the flood locations to ZCTAs.

US Census American Community Survey 5-year dataset ending 2022:
https://www.census.gov/data/developers/data-sets/acs-5year.html

US Census Zip Code Tabulation Area shapefiles
https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

NOAA StormEvent

NOAA maintains a national storm event database for storm events such as tornados, wildfires, floods, hail, thunderstorms, etc. For this project, I will use the “Flood” and “Flash Flood” event types to relate poverty status to flood occurrence by zip code. The data includes location (latitude and longitude, which will be mapped to ZCTAs), state, date, injuries, deaths, and damage to property and crops. Additional features may be calculated such as event length (days) or other potentially relevant features as needed.

NOAA StormEvents Dataset:
https://www.ncdc.noaa.gov/stormevents/ftp.jsp

Modeling Methods and Final Products

I will first attempt a linear regression model to establish initial feature correlations, and then try Random Forest to evaluate feature importance. Modeling will be limited to the 48 continental states, and excluding the District of Columbia. The final products for this project will be a report and slide presentation.

Possible Constraints

Some of the analysis may be conducted at the state level, but it is important to recognize that ZCTAs do not map perfectly to states in a many-to-one hierarchy. So, state-level and ZCTA-level analyses should be carefully compared.