

Jonathan Barta

Dec 8th, 2025

Dr. Toups

BIOS 512: Data Science Basics

A FEMA Disaster Analysis

This project analyzes the efficiency and equity of the United States federal disaster response system. By merging historical disaster data with county-level socioeconomic indicators, I explore whether a county's wealth or disaster experience influences how quickly the US Federal Emergency Management Agency (FEMA) responds to emergencies.

Data Sources and Structure

The analysis utilizes two primary datasets merged on 5-digit FIPS county codes:

- **FEMA Disaster Declarations:** A historical dataset of all federally declared disasters (1953–2021) sourced from FEMA. It includes incident types, declaration dates, and aid program flags.
- **US Census Data (ACS):** Demographic data sourced via the tidycensus API (2021 ACS 5-Year Estimates), providing median household income, poverty counts, and population for every US county.

Preprocessing and Environment

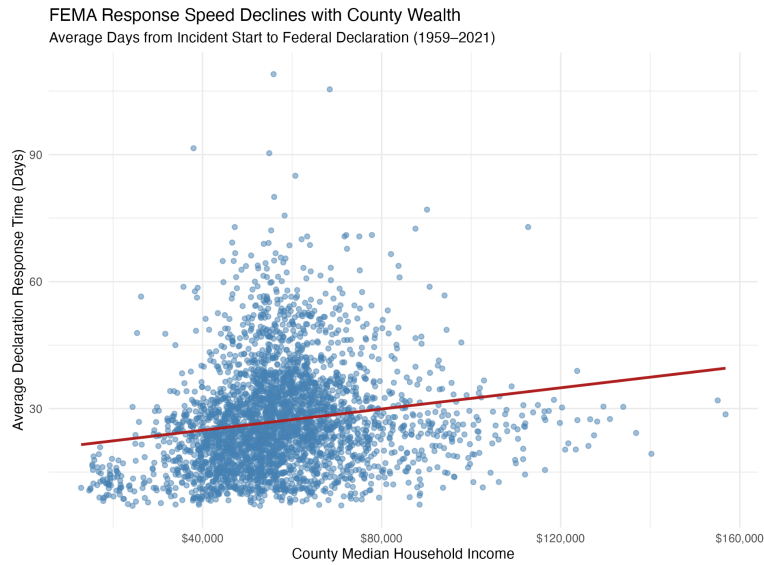
The FEMA dataset was pre-cleaned by the provider to remove duplicate entries, requiring minimal manual cleaning. My preprocessing pipeline focused on feature engineering: calculating the `response_time_days` (declaration date minus incident begin date), log-transforming income distributions, and standardizing FIPS codes with leading zeros to ensure a successful merge with Census data.

The project was developed in a JupyterLab environment using an R kernel. The source code and raw data are version-controlled and hosted in a public GitHub repository, while the final reproducible report was generated using RMarkdown.

In this analysis, I sought to answer four primary questions regarding the equity and efficiency of federal disaster response:

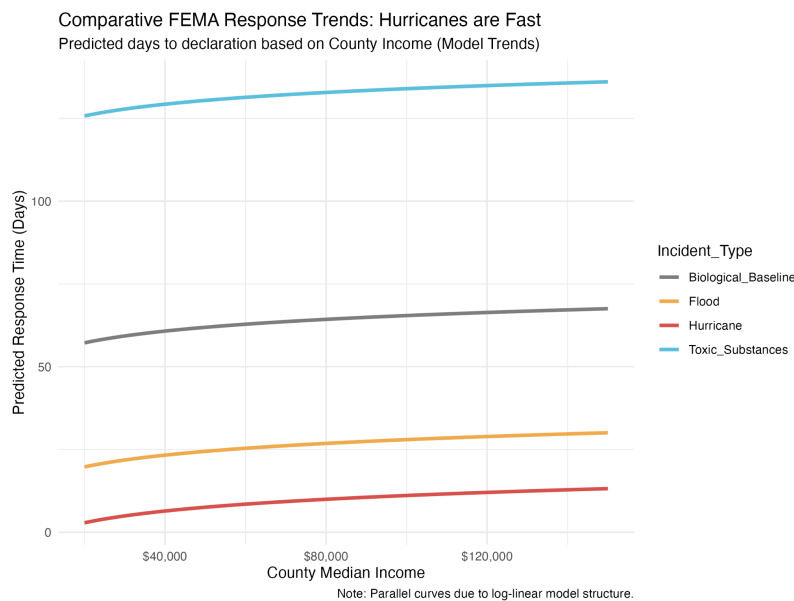
1. **Response Time by Socioeconomic Status:** Do lower socioeconomic status (SES) counties experience slower federal response times compared to wealthier counties?
2. **Response Time by Incident Type:** How does the type of disaster (e.g., Hurricane vs. Biological) impact the speed of the declaration process?
3. **Patterns of Natural Disasters:** Are there distinct seasonal or operational patterns in when specific disasters occur?
4. **The Aid Gap:** Is there an equitable distribution of Individual Assistance (cash aid) across different income levels, or does an aid gap exist?

Figure 1: Response Time by County Wealth (Scatter Plot)



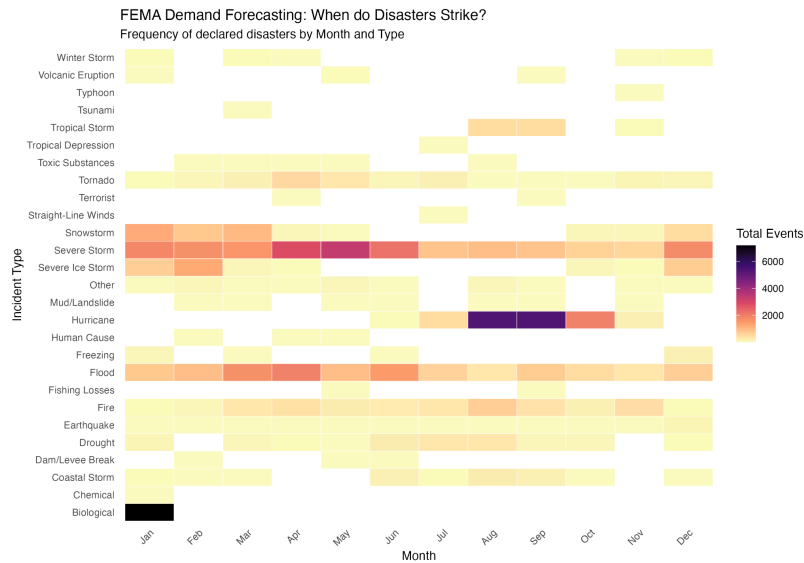
This regression reveals a statistically significant positive correlation between income and response time. Contrary to the assumption that wealthy areas receive preferential speed, higher median income is associated with a slightly slower federal declaration, suggesting these counties may rely on local resources during the initial phase of a disaster.

Figure 2: Response Trends by Incident Type (Line Chart)



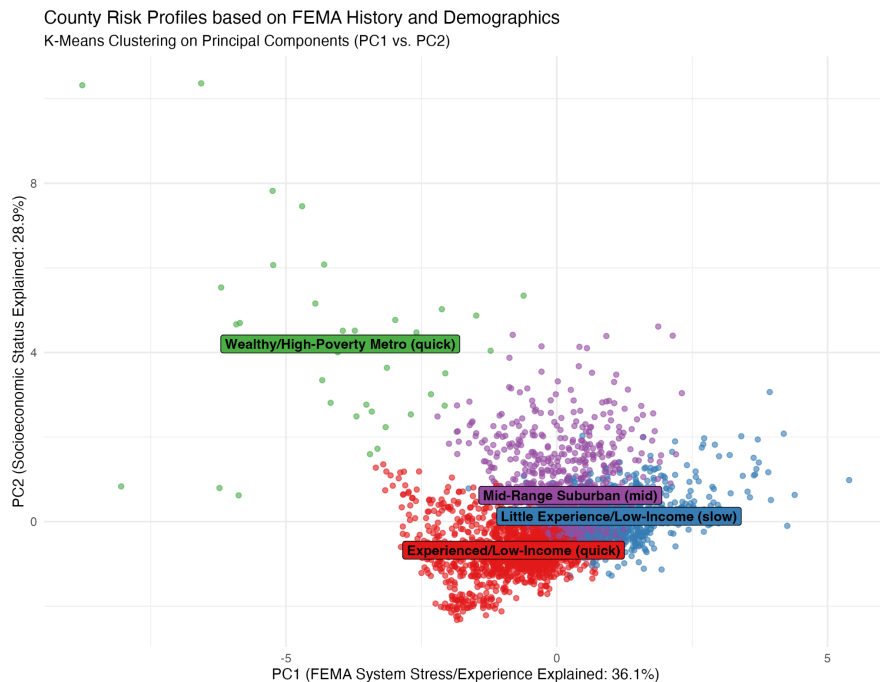
While the type of disaster is the strongest predictor of speed, with Hurricanes (red) triggering the fastest declarations and Toxic Substances (blue) the slowest, the ‘wealth penalty’ remains consistent. The upward slope across all disaster types confirms that as county income rises, federal response time increases regardless of the specific incident.

Figure 3: Disaster Seasonality (Heat Map)



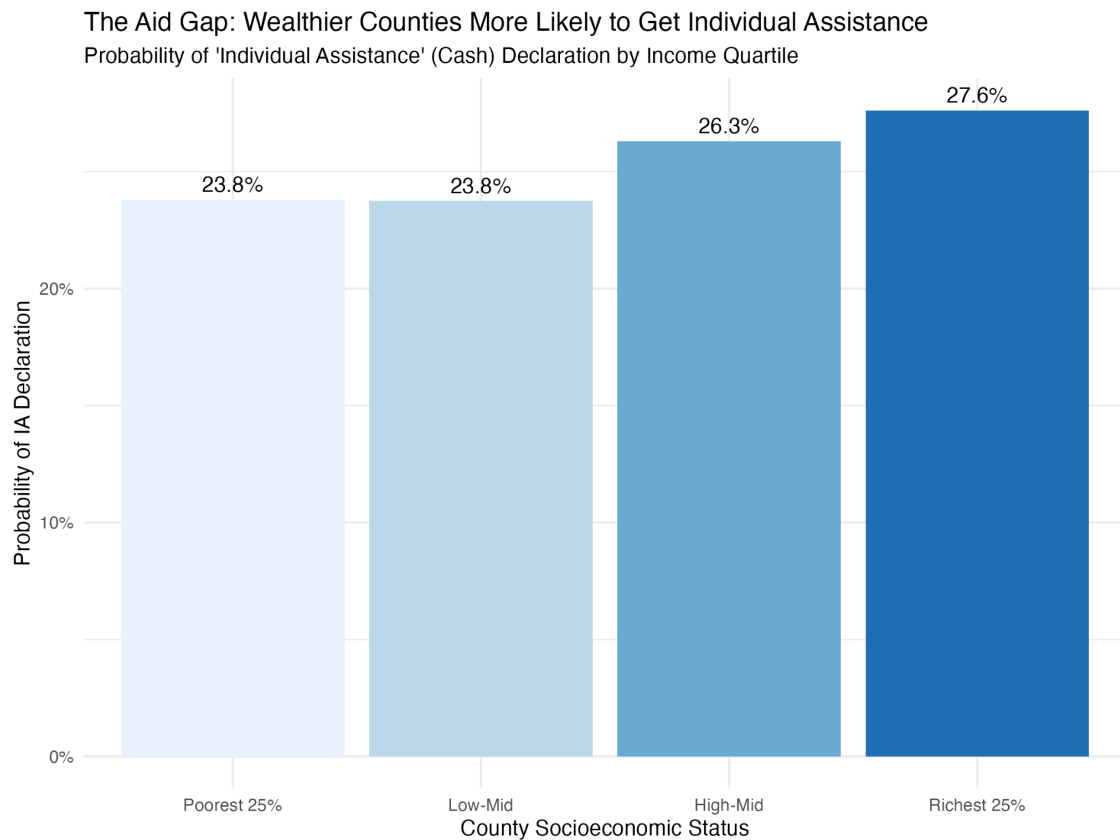
FEMA faces distinct operational peaks. The heat map shows that Flood events (yellow/orange) create consistent demand throughout the spring, while Hurricane events create extreme, concentrated demand in August and September. This seasonality dictates when federal resources are most stretched.

Figure 4: County Risk Profiles (PCA Clusters)



PCA and K-Means clustering identified four distinct county profiles. The most distinct contrast exists between Experienced/Low-Income counties (typically rural, high-frequency disaster zones with fast systems) and Wealthy/Metro counties (high asset value, complex systems).

Figure 5: The Aid Gap (Bar Chart)



This chart highlights a systemic paradox: while wealthier counties experience slower initial response times, they are 3.8% more likely to eventually receive Individual Assistance (cash aid) than the poorest counties. This suggests that administrative capacity to navigate federal bureaucracy may be a stronger driver of securing aid than raw humanitarian need.

Conclusions and Future Work

1. Summary of Findings

This analysis examined over 60 years of FEMA disaster declarations to understand the relationship between socioeconomic status, disaster experience, and federal response. The results uncover a paradox of speed versus access: while lower-income areas receive faster declarations, wealthier areas are more successful at securing financial aid.

- **The Speed Findings (Regression):** The type of incident is the primary driver of response time (e.g., Hurricanes are processed ~54 days faster than Biological events). However, controlling for disaster type, wealthier counties experience slower response times. This suggests that affluent areas may have the local resources to manage the initial phase of a disaster without immediate federal intervention.
- **The Risk Archetypes (Clustering):** PCA and K-Means clustering identified four distinct county profiles. The most distinct contrast exists between Experienced/Low-Income counties (typically rural, high-frequency disaster zones with fast systems) and Wealthy/Metro counties (high asset value, complex systems).
- **The Access Paradox (Bar Chart):** Despite slower response times, wealthier counties are 3.8% more likely to receive Individual Assistance (cash aid) than the poorest counties. This contradicts the assumption that aid favors the most vulnerable. It suggests that administrative capacity (the ability to fill out forms and navigate bureaucracy) or higher property damage values in wealthy areas may play a larger role in securing aid than raw humanitarian need.

2. Policy Implications

The data suggests an operational efficiency gap. While the system is fast at declaring disasters in experienced, lower-income areas, the actual allocation of Individual Assistance skews toward wealthier regions. A potential policy recommendation is to simplify the Individual Assistance application process for Tier 1 (Experienced/Low-Income) counties, reducing the administrative burden that may be preventing poorer regions from accessing funds they qualify for.

3. Limitations and Future Work

- **Temporal Bias:** This study merged historical disaster data (1953–2021) with a static snapshot of Census data from 2021. This assumes that the relative wealth ranking of counties has remained stable over 70 years, which introduces temporal bias. A more robust study would utilize decennial Census data to match economic conditions to the specific year of the disaster.
- **Damage Thresholds:** The dataset contains binary flags for aid programs but lacks specific dollar amounts for damages. Future analysis should incorporate Public Assistance (PA) dollar obligations to measure the magnitude of aid, not just the probability of declaration.