Data-Driven Insights into FEMA Fund Allocation: Socioeconomic and Disaster Impact Analysis in the United States

Contributors: Michelle Cheng, Dylan Myaing, Helena Moore, Paula Delgado, Matt Lau, Anran Zhao

Abstract:

This study analyzes how the Federal Emergency Management Agency (FEMA) allocates disaster relief funds across U.S. states, focusing on the influence of socioeconomic factors, disaster frequency, and disaster type. With climate change increasing the frequency and severity of natural disasters, ensuring equitable and efficient fund distribution is essential for enhancing disaster resilience and recovery. Using reliable FEMA datasets combined with state-level data on unemployment rates, health outcomes, and disaster types and responses, this research employed regression and clustering techniques to identify patterns in funding allocation. Regression analysis revealed that frequent and severe disasters, such as hurricanes and floods, are strongly associated with higher funding levels, while less frequent but high-impact disasters, like tsunamis or freezing events, may be underfunded. Clustering methods supported that states with frequent, high-impact disasters received the most funding, while others with moderate disaster frequency and vulnerability showed inconsistent allocations. These findings suggest inefficiencies in FEMA's current practices and highlight opportunities for improvement. Thus, FEMA should align resource distribution with state vulnerabilities and disaster profiles to enhance preparedness and recovery. Future work could incorporate additional variables, such as infrastructure readiness, and develop predictive tools to guide funding decisions. This study provides actionable insights to support FEMA in optimizing disaster relief efforts, ensuring equitable recovery outcomes, and strengthening resilience in disaster-prone communities.

Introduction:

In recent years, the United States has faced an alarming rise in the frequency and intensity of natural disasters, a trend closely linked to the escalating effects of global warming. From devastating hurricanes to wildfires and floods, these disasters not only endanger lives but also place immense strain on federal relief programs like the Federal Emergency Management Agency (FEMA). As the climate crisis continues to contribute to the increasing frequency and intensity of natural disasters, the question of how FEMA allocates its funds has become more pressing. Effective disaster relief depends on a well-prioritized distribution of resources that accounts for the varying needs of states based on their socioeconomic conditions, financial resilience, health infrastructure, and disaster history.

This research paper examines the allocation of FEMA funds in relation to these critical factors, aiming to identify patterns that could be used in the future to create a data-driven framework for prioritizing funding. By analyzing datasets that provide insights into states' socioeconomic profiles, health insurance coverage, financial conditions, and disaster frequency and type, the study explores how FEMA currently distributes its funds and whether that distribution aligns with the most pressing needs. With global warming accelerating the frequency and severity of disasters, ensuring equitable and efficient fund allocation is extremely important. The central question driving this research is: How does FEMA prioritize its fund allocation to maximize disaster resilience, and how do current funding practices correlate with the factors that most significantly impact disaster vulnerability? By addressing this question, the paper aims to contribute to a more effective approach to disaster relief in the United States.

The analysis performed in the study provides a detailed exploration of FEMA’s funding patterns by utilizing statistical models and clustering techniques to examine the relationship between funding amounts, disaster types, and socioeconomic variables. Through regression analysis, the study identifies the disaster types most strongly associated with high funding allocations, such as hurricanes and floods, as well as disaster types that receive relatively less funding despite their potential for high impact, like tsunamis. In addition, the clustering analysis reveals distinct patterns among states, grouping them into clusters based on the frequency and severity of disasters and funding levels. These clusters highlight discrepancies in how FEMA’s funding is allocated, revealing inconsistencies in addressing vulnerability and disaster preparedness across different regions.

By examining results from both regression and clustering, the study reveals systemic inefficiencies in FEMA’s current allocation framework. For example, the findings indicate that some states with moderate disaster frequencies but high socioeconomic vulnerability may not be receiving funding proportional to their needs. Conversely, states experiencing frequent and severe disasters are often prioritized, but this may neglect smaller states with specific high-impact disasters. The results emphasize the need for a more equitable funding strategy, one that accounts for both disaster characteristics and the underlying vulnerabilities of affected populations.

This research offers practical recommendations to improve FEMA’s disaster relief framework. These include leveraging predictive models to anticipate funding needs based on disaster trends and state vulnerability profiles, and incorporating additional socioeconomic variables into the allocation process. By addressing these gaps, FEMA can improve disaster relief and recovery outcomes, ensuring that all communities are better equipped to face the challenges of an increasingly unpredictable climate. This paper provides not only a critical evaluation of current funding practices, but also a proactive outlook on how FEMA's strategies can evolve to meet future demands.

Data:

The primary analysis in this study focuses on the different states in the United States. Datasets collected for this study include variables such as unemployment rates, FEMA fund allocation, and natural disaster details on the state level. The study specifically chose datasets from the FEMA website due to their comprehensive coverage, and reliable nature, as FEMA datasets are maintained and regularly updated by a trusted federal agency, ensuring accuracy, consistency, and a national standard for disaster-related data. Additionally, FEMA's data quality is validated through rigorous collection and reporting processes, making it an authoritative source for understanding disaster impacts and funding allocation patterns.

The following individual datasets were compiled into one unified dataset: FEMA emergency management performance grants, FEMA disaster declarations in the US by state, and GDP, household income/size, cancer (health outcomes), and unemployment in the US by state. The FEMA emergency management performance grant dataset provided funding amounts provided to individual projects across all 50 states. The FEMA disaster declarations dataset provided individual instances of disaster declarations across states since 1953. The various socioeconomic datasets provided data about average household size and income, and county and state level information on GDP, unemployment rates, poverty, health factors, and death rate.

This data initially was expected to provide a challenge in consistency and missing values. Since the unified dataset was pooled across many different sources, we predicted difficulties in standardizing it. Varying timeframes from each dataset, different naming conventions for common variables, and organization by state vs. county posed issues. Additionally, it was expected that missing data could be inconsistently spread across different states and years between the various sources. This required extensive cleaning and examination of variables.

From these datasets, we extracted and analyzed critical variables such as disaster type, frequency, severity, and FEMA relief funding amounts, as well as state-level indicators like unemployment rates and household size. The data cleaning process involved addressing inconsistent data types, renaming columns for consistent merging, and handling missing data through imputation or removal when necessary. For example, we validated data integrity by ensuring that each dataset included exactly 50 unique states, confirming comprehensive coverage. Duplicates were removed where identified, and FIPS codes were cross-referenced to ensure consistency and accuracy. Datasets were merged primarily on state names and FIPS codes, which served as unique identifiers across datasets. FIPS stands for Federal Information Processing Standards, and is used by the US government to provide standardized codes for each county and state in the country. All datasets shared FIPS data or state and county names, and were converted to standardized codes and combined using a state abbreviation and county name to FIPS table, and a state level fips code table. This negates the issue of problems with string matching for county names that could be labelled differently across datasets, which enabled the creation of a single unified dataset integrating socioeconomic and disaster-related variables.

Initially, we were hoping to examine the long term socioeconomic effects of disasters on a county and state level – for example, the effect on health outcomes of counties with lower education levels and higher poverty rates post-disaster. However, due to the nature of the data being more focused on events rather than causative long term outcomes, we pivoted our central question to investigate how we could specifically analyze FEMA disaster declaration and funding data combined with socioeconomic factors to suggest where FEMA could better allocate resources.

Methodology:

This study employs supervised and unsupervised learning techniques to model and predict the relationship between socioeconomic factors and FEMA fund allocation in disaster-prone areas. The analysis leverages a combination of k-means clustering and regression models to reveal further insights in the data regarding trends and correlation.

K-Means Clustering: This technique was used to group counties based on disaster type, frequency, funding levels, and public assistance applicants. By identifying counties with similar characteristics, this clustering method provided actionable insights into shared vulnerabilities and funding needs, which informed potential policy priorities for FEMA fund allocation. Hierarchical clustering was also explored to validate the robustness of the groupings. These clusters allowed us to identify patterns in FEMA fund allocation and group states or counties with similar vulnerabilities and resource needs. In addition, as seen in Figure 2, hierarchical clustering was explored to provide complementary insights and validate the robustness of the identified groupings. These clusters were used to identify patterns and can be used to inform potential policy priorities for FEMA fund allocation.

Regression Analysis: This technique was used to model the relationship between FEMA fund allocations and various factors such as disaster frequency, type, and socioeconomic indicators. The process began with data preparation, where datasets were cleaned and merged –- this involved standardizing column names and handling missing data through imputation or removal. The merged dataset included key variables like the number of disasters, FEMA funding amounts, unemployment rates, and household sizes. For the regression model, we used linear regression techniques to predict FEMA fund allocations. The dependent variable was the FEMA fund allocation amount, while independent variables included disaster frequency, type, and socioeconomic factors. We split the data into training and testing sets using an 80-20 split to train the model and evaluate its performance. The model was trained using the training set and evaluated using metrics such as Mean Squared Error (MSE) and R-squared values. This method allowed us to measure the impact of each factor on funding levels.

The success of these approaches can be measured by the model's ability to predict FEMA funding allocations and long term outcomes based on socioeconomic and disaster-related factors. Metrics such as R^2 and RMSE are used to assess prediction accuracy and variance explained. For potential classification tasks, such as predicting "high" versus "low" vulnerability, metrics like accuracy and sensitivity. In addition, several challenges were anticipated during this study. Data sparsity and missing values were addressed through imputation or by excluding counties with substantial gaps in data. Geographical aggregation issues from mismatched data granularity, such as state-level versus county-level data, were also mitigated by aggregating or supplementing datasets as needed. Results were presented through a combination of performance metrics and visualizations. Regression coefficients, cluster visualizations, scatter plots, as well as other graph types were used to effectively display our results.

Results:

Our aim was to analyze how FEMA allocates funding aid to states, based on the type of natural disaster that occurs. Though there are many other factors that may come into play regarding funding allocation, natural disaster type clearly plays a significant role and it's important to dissect these allocations. Figure 1 represents the correlation between disaster type and allocated funding. For example, in this linear regression model, coefficients represent the relationship between natural disasters (under "Feature") and the funding amount (represented by the "Coefficient") as follows: If the coefficient for Earthquake is is 6.78 x 106, it means that for each additional earthquake in a state during a given year (all else being equal), the model predicts an additional $6,580,000 in disaster relief funding. A negative coefficient is associated with a decrease in predicted disaster relief funding (this could occur due to prioritization or other complex funding dynamics). It's important to note that these coefficients do not imply causation, only statistical association. It's also important to note that if disaster types are highly correlated (e.g., hurricanes often cause floods), the coefficients might not reliably isolate the effect of each variable.

To further analyze how state funding is allocated and understand how natural disasters play a role in funding, we conducted a k-cluster analysis. Figure 5 demonstrates 3 clusters pulled from the 50 states based on disaster type, frequencies, and allocated funding. As seen in Cluster 0, there are a high average number of disasters, and this represents areas with frequently occurring disasters. Cluster 1 has fewer disasters than Cluster 0, and moderate funding levels which probably represents regions with moderate disaster frequency. Cluster 3 has high values for specific disasters (such as Freezing), which suggests a focus on areas prone to specific but impactful disasters.

This data provides crucial insights into patterns of disaster frequency, severity, and associated funding, allowing for a more strategic approach to resource allocation (reference Figure 6 for a visualization of the disaster clustering). By clustering disasters into groups based on their characteristics, decision-makers can identify regions or disaster types that require the most attention. For instance, Cluster 0, with the highest total number of disasters and significant funding allocation, likely represents areas with frequent and severe events, such as floods or fires, necessitating sustained investment. Conversely, Cluster 1, with fewer disasters and lower funding, may highlight regions where resources can be allocated more efficiently. The funding discrepancies across clusters suggest opportunities to reassess how resources are distributed, ensuring that high-impact but less frequent disasters (e.g., tsunamis in Cluster 2) are not underfunded. Overall, this analysis enables targeted disaster preparedness and relief strategies, ensuring funding aligns with the unique needs of each cluster to maximize resilience and minimize long-term recovery costs.

Conclusion:

Our findings indicate that disaster frequency and severity are significant predictors of FEMA fund allocation, with frequent disasters consistently associated with higher funding levels. However, the relationship between funding and socioeconomic factors, such as poverty rates, educational attainment, and health outcomes was less consistent. This suggests that FEMA's current allocation framework may not adequately account for the underlying vulnerabilities of disaster-prone areas. For instance, states experiencing moderate disaster frequency but high socioeconomic vulnerability often received less funding than their needs might warrant. This highlights a systemic inequity in resource distribution that may exacerbate recovery challenges for already disadvantaged communities.

The clustering analysis further revealed distinct patterns in funding behavior, grouping states based on disaster frequency, type, and funding levels. High-frequency disaster states consistently received significant funding, while areas with specific but less frequent disasters appeared to be unaccounted for. This analysis provides actionable insights for FEMA to refine its funding strategy, ensuring that less visible but high-impact disasters are not overlooked in favor of more immediate crises.

FEMA's current funding behavior is effective in regards to addressing immediate needs during high frequency disasters. These events normally affect large populations, cause widespread damage, and require substantial resources for recovery. From this perspective, prioritizing frequent disasters aligns with the agency's mandate to respond quickly to widespread emergencies. However, this approach also includes shortcomings as disasters that are less frequent but equally catastrophic often receive disproportionately less funding. Prioritizing high-impact disasters is crucial, as their long-term effects, such as infrastructure damage and population displacement, can severely delay recovery and deepen vulnerabilities. This inconsistency suggests that FEMA adopts reactive behavior, rather than proactive behavior, favoring visible and widespread crises over high-impact ones. While this prioritization is logical in respects to the scale of disasters, it ultimately fails to account for long-term vulnerabilities and the systemic inequities across states. FEMA's funding behavior is rational in some regards, but it does not fully align with the data, which indicates a need for a balanced approach that factors in both frequency and socioeconomic vulnerabilities of impacted areas.

While the study offers valuable insights, it also faced notable limitations stemming from data quality and scope. The integration of diverse datasets introduced challenges related to inconsistent timeframes, mismatched granularity, and missing values. These issues required extensive data cleaning and imputation, which may have affected the precision of the findings. Moreover, the reliance on event-focused datasets constrained our ability to explore the long-term socioeconomic impacts of disasters, such as changes in health outcomes or economic recovery rates. Addressing these gaps would require more granular, longitudinal datasets that track post-disaster recovery metrics over time.

Our initial research question, which sought to explore the health impacts of natural disasters, was ultimately revised due to data constraints. This pivot highlights the difficulty of linking disaster data with long-term outcomes, but it also underscores an opportunity for future research. Incorporating more comprehensive variables, such as infrastructure readiness and climate adaptation efforts, could provide a deeper understanding of how disasters affect communities and how resources can be optimally allocated.

This study provides a foundational framework for improving FEMA’s disaster relief strategies, but it also highlights avenues for future research and development. One potential extension would involve analyzing the cost-benefit trade-offs of investing in disaster preparedness versus recovery. By quantifying the economic and social impacts of proactive resilience-building measures, FEMA could allocate resources more effectively to minimize long-term recovery costs.

Real-time disaster data combined with socioeconomic indicators could further enhance FEMA’s ability to dynamically adjust funding in response to emerging vulnerabilities. Finally, longitudinal studies tracking post-disaster recovery metrics, such as employment rates, educational attainment, and health outcomes, would provide a more comprehensive understanding of the long-term impacts of disasters and the effectiveness of relief efforts.

While FEMA’s current funding practices address immediate disaster response effectively, they fall short of equitably addressing the long-term vulnerabilities of disaster-affected communities. The study emphasizes the need for a balanced approach that accounts for both disaster frequency and the socioeconomic characteristics of impacted areas. By adopting data-driven tools and predictive models, FEMA can improve its disaster relief framework to better align with the diverse needs of states and ensure a more resilient and equitable recovery process. These recommendations serve as a roadmap for FEMA to enhance its resource allocation strategies, addressing both current inefficiencies and the evolving challenges of an increasingly unpredictable climate.

Appendix:

```
Mean Squared Error: 75236351229713.56
R-squared: -1.6420160847190273
              Feature   Coefficient
13              Other  7.458012e+06
5          Earthquake  6.578130e+06
19  Toxic Substances   3.527077e+06
3     Dam/Levee Break  2.175235e+06
6                Fire  2.714233e+05
18            Tornado  6.178786e+04
8               Flood  4.682506e+04
14   Severe Ice Storm  3.367254e+04
11          Hurricane  2.696142e+04
0          Biological  8.313693e+03
15       Severe Storm  1.473953e+03
1            Chemical  2.997695e-09
10        Human Cause  9.313226e-10
9            Freezing  1.164153e-10
22            Typhoon  0.000000e+00
17          Terrorist  0.000000e+00
23  Volcanic Eruption  0.000000e+00
20     Tropical Storm  0.000000e+00
21            Tsunami  0.000000e+00
7      Fishing Losses -4.656613e-10
4             Drought -2.153683e-09
16          Snowstorm -4.004818e+04
24       Winter Storm -1.177787e+05
2       Coastal Storm -1.715085e+05
12      Mud/Landslide -3.614730e+05
```

Figure 1. A linear regression demonstrating the correlation between natural disaster occurrence within a state and the expected change in funding allocated to that state as a result.

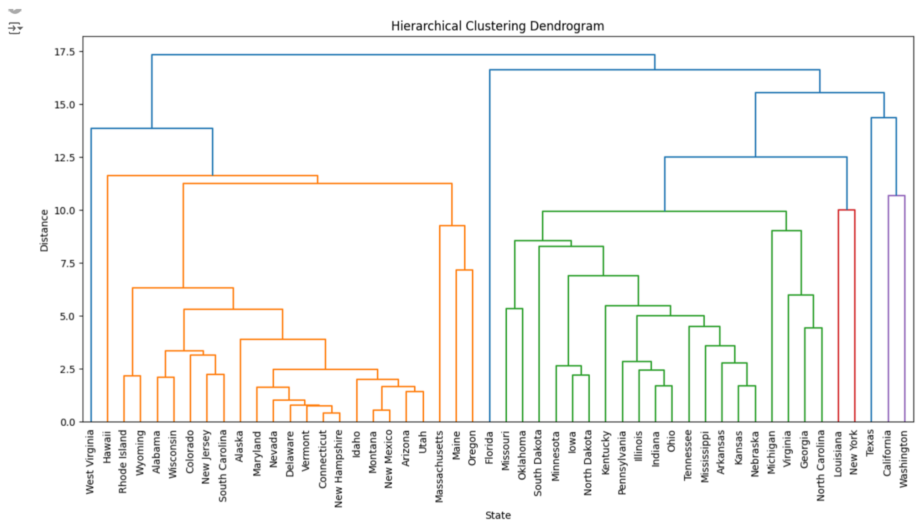

Figure 2. Hierarchical clustering of states based on natural disaster type, occurrence, and FEMA Funding allocation.

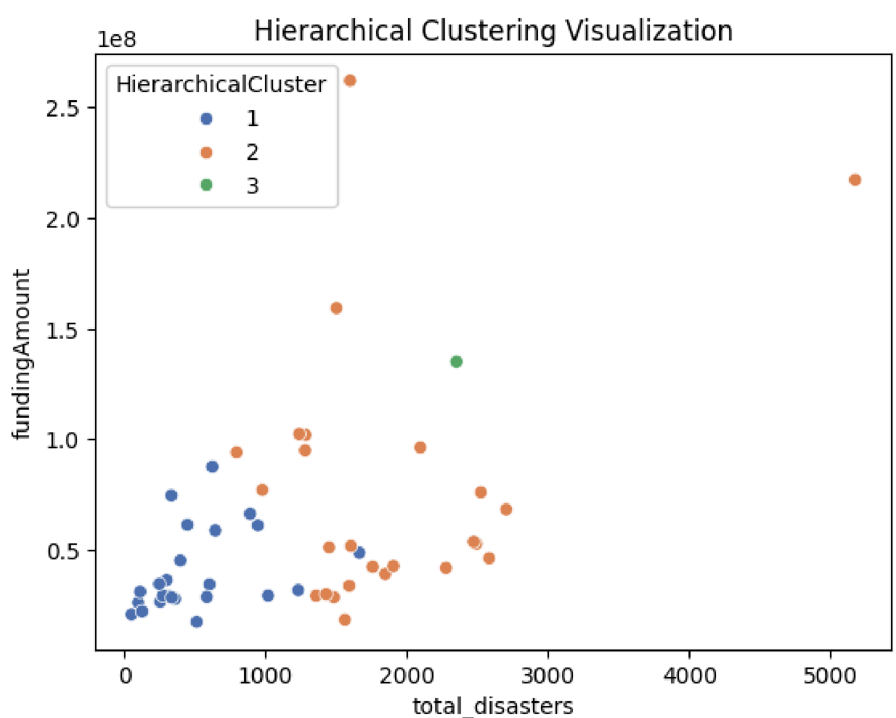

Figure 3. Correlations within the cluster between funding amount and total number of disasters.

```
            Biological  Chemical  Coastal Storm  Dam/Levee Break  Drought  \
state
Alabama            139         0              0                0       67
Alaska             121         0              2                0        0
Arizona             73         0              0                0        8
Arkansas           152         0              0                0       32
California         116         0             43                5       47

            Earthquake  Fire  Fishing Losses  Flood  Freezing  ...  Terrorist  \
state                                                          ...
Alabama              0    11               0    104         0  ...          0
Alaska              13    30               0     50        14  ...          0
Arizona              0    88               0     82         1  ...          0
Arkansas             0     0               0    271         0  ...          0
California          32   460               4    448        57  ...          0

            Tornado  Toxic Substances  Tropical Storm  Tsunami  Typhoon  \
state
Alabama          62                 0               0        0        0
Alaska            0                 0               0        0        0
Arizona           0                 0               0        0        0
Arkansas        113                 0               0        0        0
California        0                 0               0        3        0

            Volcanic Eruption  Winter Storm  total_disasters  fundingAmount
state
Alabama                     0             0             1665   4.887564e+07
Alaska                      0             0              318   2.934264e+07
Arizona                     0             0              335   7.473417e+07
Arkansas                    0             0             1593   3.395693e+07
California                  0             0             1599   2.618396e+08

[5 rows x 27 columns]
```

Figure 4. Example of State mapping of natural disasters, total disaster occurrence, and funding allocated to each state

```
         Biological  Chemical  Coastal Storm  Dam/Levee Break    Drought  \
Cluster
0        569.000000  0.000000      25.500000         0.000000  74.500000
1        135.829787  0.191489      11.914894         0.276596  22.893617
2        150.000000  0.000000      25.000000         0.000000   0.000000

         Earthquake        Fire  Fishing Losses       Flood    Freezing  ...  \
Cluster                                                                  ...
0          0.000000  613.500000        8.000000  299.500000   13.500000  ...
1          2.319149   44.744681        0.553191  205.085106    2.702128  ...
2          0.000000  277.000000        0.000000   73.000000  147.000000  ...

           Tornado  Toxic Substances  Tropical Storm   Tsunami   Typhoon  \
Cluster
0        25.000000          0.000000             0.0  0.000000  0.000000
1        31.510638          0.191489             0.0  0.191489  0.042553
2        38.000000          0.000000            54.0  0.000000  0.000000

         Volcanic Eruption  Winter Storm  total_disasters  fundingAmount  \
Cluster
0                 0.000000      0.000000      3096.000000   1.233323e+08
1                 1.085106      0.914894      1132.234043   5.596522e+07
2                 0.000000      0.000000      2352.000000   1.350564e+08

         HierarchicalCluster
Cluster
0                   1.500000
1                   1.510638
2                   3.000000

[3 rows x 28 columns]
```

Figure 5. Mapping of unique natural disasters, total disaster occurrence, and allocated funding per 3 clusters of the 50 states (reference Figure 2)

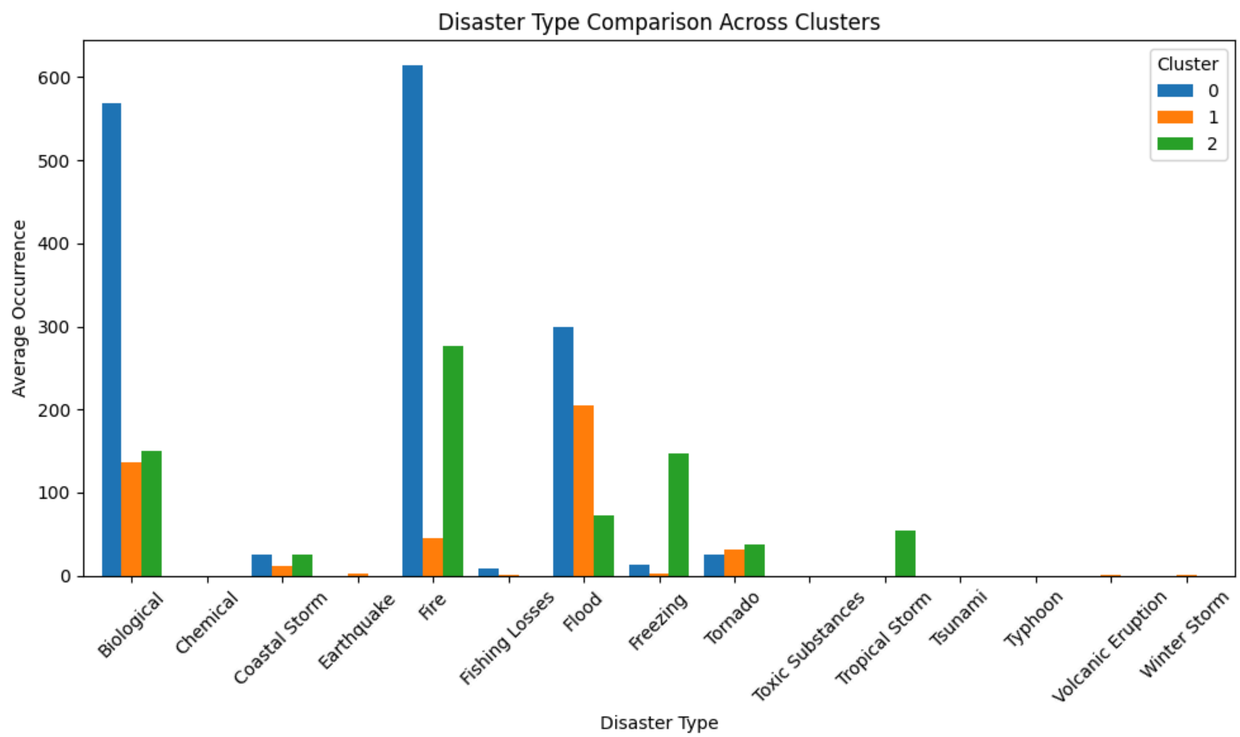

Figure 6. Bar chart visualization of Figure 5

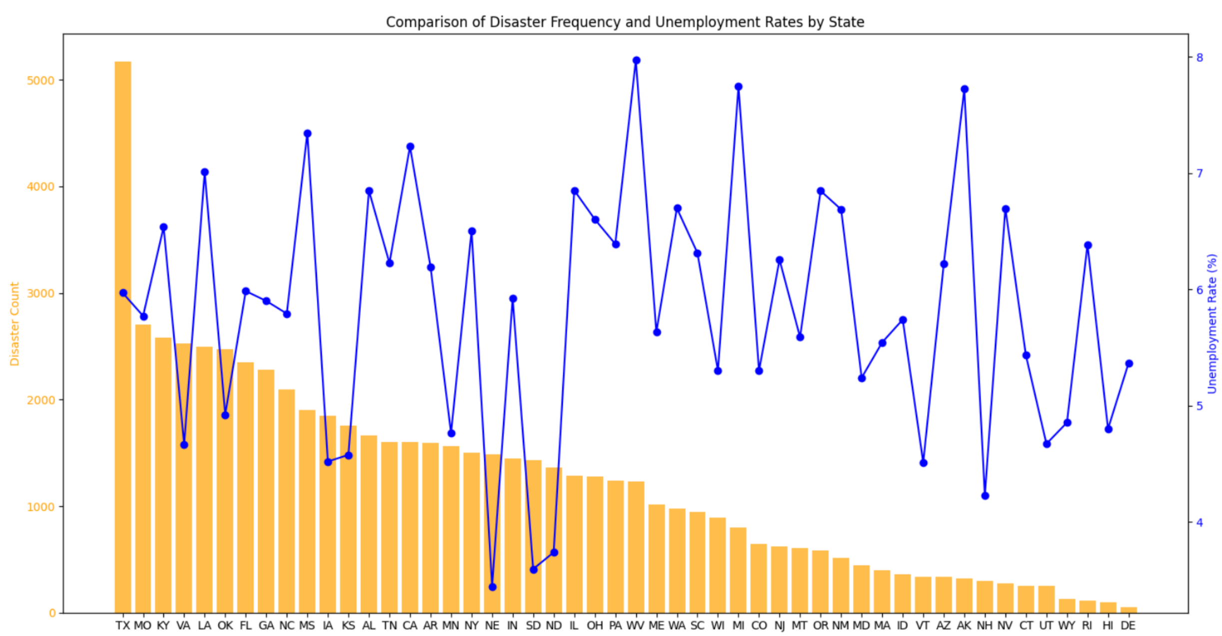

Figure 7. Comparison of disaster frequency and unemployment per state

References

Federal Emergency Management Agency. *Disaster Declarations Summaries*. Retrieved from *DisasterDeclarationsSummaries.parquet*.

Federal Emergency Management Agency. *U.S. Disaster Declarations*. Retrieved from *us_disaster_declarations.csv*.

Federal Emergency Management Agency. *U.S. Disasters (M5 Dataset)*. Retrieved from *us_disasters_m5.csv*.

U.S. Bureau of Labor Statistics. *Unemployment Rate by State*. Retrieved from *unemployment_per_state.csv*.

U.S. Census Bureau. *State and County FIPS Code Master List*. Retrieved from *state_and_county_fips_master.csv*.

U.S. Census Bureau. *State FIPS Master List*. Retrieved from *state_fips_master.csv*.