“There is no post-hurricane world”

Laura Brown
labr1398@colorado.edu

Alexander Gonzalez-Torres
algo9142@colorado.edu

Joo Young Han
joha4925@colorado.edu

PROBLEM STATEMENT

Natural disasters have profound effects on both the economic and social development of nations. Research on disaster impacts, such as those caused by Hurricane Maria in Puerto Rico, demonstrates how these disruptions can create long-term challenges in governance, human development, and economic recovery. The idea that "there is no post-hurricane world," as a 2022 exhibit at the Whitney Museum of American art put it, suggests that disasters like hurricanes, floods, and earthquakes fundamentally alter a society’s trajectory, leaving lasting scars on economic performance and social well-being. However, whether these disruptions produce temporary setbacks or long-term changes requires further exploration, particularly as it concerns social development metrics that are well-documented yet often omitted from disaster research. For example, access to clean cooking fuels, regulatory quality, access to financial institutions, etc.

This project aims to fill this research gap by applying data mining and statistical techniques to a combined dataset that includes: 1. information on natural disasters between 1960 and 2023 2. Economic and human development metrics for this time period. The analysis will address questions such as:

- Do different types of natural disasters produce distinct economic and social impacts?
- Are these impacts temporary, or do they result in long-term structural changes?
- How do disasters affect key economic indicators such as GDP growth, public debt, and trade?
- How do disaster affect social outcomes like education, healthcare access, and gender equality?
- What novel metrics, as defined by their omission from pertinent literature, are affected by disasters? To what extent?
- What attributes are shared among countries that recover more effectively from disasters?

Answering these questions will provide insight into how countries can mitigate the impact of natural disasters and areas where preparedness and disaster management can improve.

LITERATURE SURVEY

By the late 20th century, disaster studies emerged as an interdisciplinary field, integrating sociology, geography, economics, and public policy. Recent research has emphasized the global interconnectedness of economies and the increasing frequency and intensity of climate-related disasters. A review of the literature for Hurricane Maria, which made landfall in Puerto Rico on September 20, 2017, serves as a clear example and inspiration for this project.

In the age of intensifying climate change and natural disasters increasing in frequency and magnitude, we hope to build on the foundation set by prior studies to distill from the data whether climate change has intensified any previously known effects of natural disasters.

In “The Economy of Disasters? Puerto Rico Before and After Hurricane Maria” Carabello-Cueto 2021, it is discussed that while Puerto Rico had been trending into rising debt beforehand - Hurricane Maria had immediate effect on fiscal planning and economic factors thereafter. Statements within the research primarily relied on short-term before and after period trends to provide insights: such as debt payments doubled, unemployment insurance claims rose, poverty increased, increase in population migration, home reconstruction claims represented a majority of Puerto Rican households and the planned budget to education had been further reduced[1].

Keeping with trends in the literature, we shift our lens towards global-scale analysis [3]. Our project follows that path with the combination of the Emergency Events Database (EM-DAT) and World Development Index (WDI) international datasets. We plan to explore global patterns around short and long-term economic effects (grouped by disaster) and possibly confirm climate change disaster research as discussed in “The Economic Impacts of Natural Disasters: A review of Models and Empirical studies” [2]. Since working with global data creates analysis issues as there are some missing data values in the EM-DAT [1], we must first find a way to work around these impactful problems within our dataset.

PROPOSED WORK

The dataset for this project combines the complete World Development Index–a World Bank Group database–with entries between 1960 and 2023 in the The International Disaster Database–a project of the Center for Research on the Epidemiology of Disasters. Due to the size of the dataset, collection will require downloading it in subsets. These subsets must then be combined into a singular DataFrame using Pandas.

To eliminate redundancy and improve data integrity, database normalization techniques will be applied to the new DataFrame. In particular, new primary keys must be developed. Additionally, NumPy and SciPy statistical methods will be used to eliminate outlying/null values and reduce dimensionality where appropriate.

Furthermore, attributes that appear in both the World Development Index and International Disaster Database—such as dates—must be standardized in format.

The statistical methods of this project will focus on time-series analysis for all attributes to determine 1. their state before a natural disaster 2. their state for years after. We acknowledge that, in some cases, we must control for global financial events like oil price crises, the Great Recession, etc. Additionally, we must control for political instability unrelated to natural disasters. Development metrics with noticeable post-disaster trends, if any, will receive further statistical analysis to draw insights not produced during time-series analysis.

To best achieve the goals set out in this section, we believe that building a data pipeline in Python is the best way forward. This pipeline should consume the data from CSV files, produce a DataFrame, extract countries and indicators, and produce time-series visualizations.

We also hope to ‘work in the open’ and our Github repository will contain a number of Jupyter Notebooks discussing our processes, methods, and struggles for other researchers using this massive dataset.

Due to the dataset’s size, we may focus our Proposed Work on a particular country or region to extract the most insights that we can from this dataset. Puerto Rico is top of mind, since it

serves as the inspiration for this project. We elaborate on the dataset in the following section.

DATA SET

1. World Development Index
2. The International Disaster Dataset

Both databases listed above produce CSV files. The World Development Index has 1488 attributes ("series"), including economic metrics such as GDP growth and GINI coefficient, and social development metrics such as HIV infection rates and access to clean cooking fuels. Changes in these attributes are described year over year ("time"), with data spanning between 1960 and 2023. Accordingly, we extract entries between 1960 and 2023 for The International Disaster Dataset. Of note, the Center for Research on the Epidemiology of Disasters indicates that entries before 1960 in their database are subject to reporting biases.

In total–without excluding null values in the World Development Index, which are common in older data–the complete dataset contains 26,480,562 data points spread between 1,533 attributes.

The data is quite granular as it concerns human development metrics, and we hope to make novel findings as we analyze these attributes following natural disasters.

EVALUATION METHODS

To ensure the rigor and reliability of this data mining project, four key evaluation methods will be employed: accuracy of insights, comparative analysis, reproducibility, and clarity of visualizations.

The accuracy of insights will be assessed by comparing our findings to existing literature on the economic and social impacts of natural disasters. This evaluation will determine whether the patterns we uncover—such as GDP trends, shifts in social metrics, or changes in governance—align with previously documented evidence. Discrepancies between our results and the literature will be addressed through comparative analysis to identify potential causes, such as differences in data sources, sample periods, or methodological approaches.

Reproducibility will be a cornerstone of the project, as it ensures that the findings can be validated by other projects. This involves transparent documentation of data sources, transformations, and methodologies. We will provide clear descriptions of the data mining and statistical techniques we applied to ensure that others can replicate our analysis in the dataset.

The use of Jupyter Notebooks plays a crucial role in facilitating the search, documentation, and validation of discoveries throughout this project, while also promoting reproducibility and transparency in both our technical work and thought process. Each step of the analysis—ranging from data acquisition and cleaning to transformation and visualization—is clearly documented, ensuring that others can easily follow, replicate, and build upon our findings.

Finally, the clarity of visualizations will be evaluated to guarantee that our results are digestible by a diverse audience. We aim to create visual tools–like graphs and dashboards– that effectively communicate our findings.

TOOLS

Programming Language: Python

Data Analytics: Pandas, NumPy, SciPy, Matplotlib, Seaborn

Collaboration: Github, Kanban, Jupyter

Database Management: SQL, SQLite, Google Drive

MILESTONES

COMPLETED

1. **October 11:** Collect complete dataset
2. **October 18:** Data preprocessing and storage in a normalized and standardized datasets
3. **October 28:** Create standardized work environments to ensure integrity in data and analysis

TO DO

4. **November 1:** Decide if we are narrowing down our analysis to a particular region or country
5. **November 8:** Complete data pipeline to normalize outlying values, eliminate empty values, initiate time-series analysis
6. **November 15:** Exploratory Data Analysis with a focus on corroborating completeness and accuracy of dataset, including time-series analysis of economic metrics
7. **November 22:** Time-series analysis of human development attributes, reduce dimensionality of dataset to focus on extracting further insights from metrics
8. **November 29:** Complete further analysis of human development metrics, corroborate resources against any existing literature
9. **December 6:** Employ Evaluation Methods outlined on this paper produce final report
10. **December 9**: Produce final report with findings and listing further research questions to be answered with dataset

REFERENCES

[1] Jose Caraballo-Cueto. The Economy of Disasters? Puerto Rico Before and After Hurricane Maria. Centro Journal. 33, 1. (Spring 2021), 66+. link.gale.com/apps/doc/A668270706/IFME?u=anon~d81b8026&sid=googleScholar&xid=e24c277b

[2] W. J. Wouter Botzen, Olivier Deschenes, and Mark Sanders. 2019. The Economic Impacts of Natural Disasters: A review of Models and Empirical studies. *Review of Environmental Economics and Policy* 13, 2 (May 2019), 167–188. https://doi.org/10.1093/reep/rez004

[3] Evi Susanti Tasri, Kasman Karimi, and Irwan Muslim. 2021. The effect of economic variables on natural disasters and the impact of disasters on economic variables. *Heliyon* 8, 1 (December 2021), e08678. https://doi.org/10.1016/j.heliyon.2021.e08678

[4] Jones, R.L., Guha-Sapir, D. & Tubeuf, S. Human and economic impacts of natural disasters: can we trust the global data?. *Sci Data* 9, 572 (2022). https://doi.org/10.1038/s41597-022-01667-x

RESULTS SO FAR

A critical aspect of our evaluation lies in data quality and pre-processing. Given the importance of reliable insights, significant effort was devoted to organizing, cleaning, and correlating datasets. This was done rather informally, as documented by the Jupyter Notebooks in our database, and we plan to publish the Python code for a data pipeline that standardizes these processes.

Due to the sheer size of our database, we found it necessary to create a project.db file that can be queried with for quick access to information. Opening the CSV files and conducting search-bar queries leads to crashes quite often.

Furthermore, we began conducting preliminary Exploratory Data Analysis to test the accuracy and robustness of our dataset. We were

particularly interested in Puerto Rico's GDP Growth after Hurricane Maria.

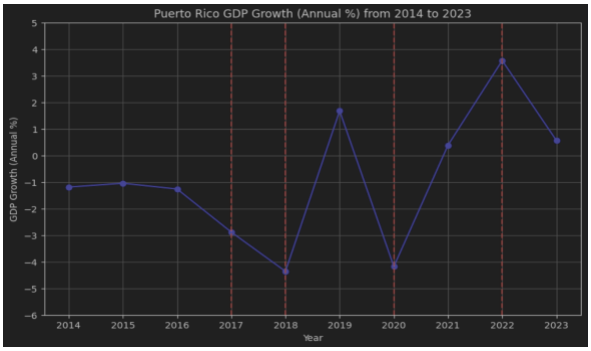

Figure [1]. Puerto Rico GDP Growth (Annual %) from 2016 to 2022 and each red vertical line indicating when a disaster had occurred. Hurricane Maria occurring in 2017.

The figure above describes annual GDP growth trends for Puerto Rico as an annualized percentage. Our dataset closely matches annualized GDP Growth found online and in the literature. This suggests that our dataset is robust, complete, and accurate. Further EDA is to be conducted.

Interestingly, the sharp downturn in GDP Growth after Hurricane Maria (2017) begins the prior year and continues into 2018, followed by sub-1% growth in 2019. GDP Growth in Puerto Rico seems decoupled, year-over-year, from the effects of Hurricane Maria.

Yet, Hurricane Maria profoundly affected the Commonwealth of Puerto Rico, as the Whitney Museum exhibit that inspires this project's title suggests. Anecdotally, one of our authors hails from Puerto Rico, and sees the effects of Hurricane Maria today, with houses that were never rebuilt and businesses that never reopened. These effects, we hope, are visualized in human development metrics that compose our dataset and that are often ignored by the literature. As written above, we will be focusing on these in further milestones.

Since the preliminary analysis detailed above challenges our operating assumptions, Hurricane Maria drastically changing the economic fate of Puerto Rico, for example, we believe that focusing this project on this country may be most productive for developing answers to the questions outlined in our Problem Statement. Additionally, it may solve the database size issues that made it necessary to bring SQL into our tech stack.

As a proof of concept of our Evaluation Methods, we conducted some temporal analysis for Japan. Figure[2] presents visualizations of data trends such as Japan's power consumption before and after the 2011 Tōhoku earthquake (Figures 1, 2, 3, and 4). This temporal analysis allows for precise insights into how disasters influence both immediate and long-term societal factors. Our approach to visualization—using side-by-side plots of metrics like total damage, deaths, and power consumption—was guided by the goal of ensuring clarity. Effective communication of these time-sensitive trends provides a nuanced understanding for policy-makers, researchers, and the general public.

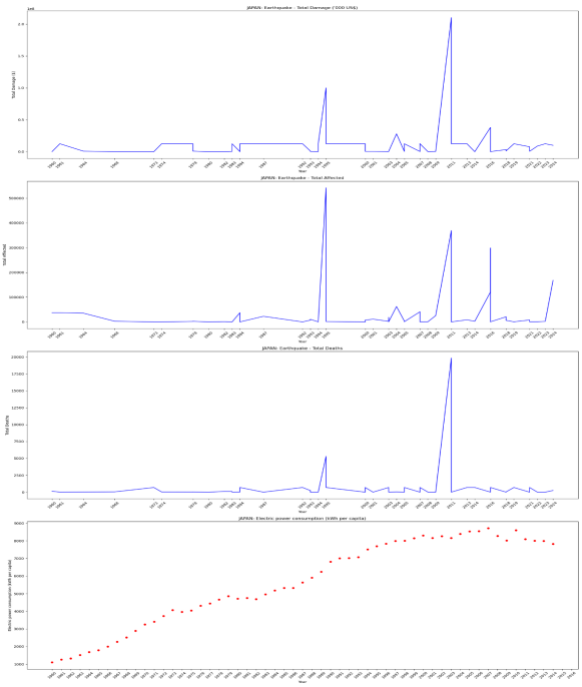

Figure [2]. The first three graphs show significant spikes in total damage, affected population, and deaths in Japan, notably peaking in 2011, reflecting the devastating impact of the Tōhoku earthquake and tsunami. The last demonstrates a steady rise in electric power consumption per capita over time, with a noticeable plateau and decline around 2011, highlighting the disruption caused by the Fukushima nuclear disaster.