WORKING DRAFT

Analysis of

Extreme Weather and Disaster Events Impacts

on Key Public Health and Economic Attributes

2024 AI BOOTCAMP | MODULE 10 GROUP PROJECT

CONTRIBUTORS: Jamie Bond, Mike Szumski, Rajesh Velamala

UNC CHARLOTTE *Continuing Education*

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

CONTENTS

PROJECT OBJECTIVE

In this project, the University of North Carolina Charlotte / Chapel Hill AI Bootcamp project team conducted a data analysis to examine the impacts of extreme weather and disaster events on public health and economics.

EXECUTIVE SUMMARY

PROJECT OBJECTIVE

To examine the impacts of extreme weather and disaster events on public health and regional economics.

DATA SOURCES

The investigating team narrowed its efforts on the collection, transformation, analysis and visualization of daily, county-level data, where available, for at least five years including the period between 2018 - 2022, and focusing on the relationship between extreme events and the following:

FEMA Datasets:

- Disaster declarations
- Disaster summaries
- Public disaster assistance project data

CDC Datasets:

- US mortality counts
- National outbreak reporting

DATA TRANSFORMATION AND ANALYSIS

Using Python tools and libraries, the investigating team conducted the following key activities:

- **Data Identification and Exploration** – identified and reviewed available datasets, screening for data applicability and data quality
- **Data Transformation** – imported, cleaned and joined datasets by location and date attributes or disaster ID
- **Data Analysis** – explored the relationship of the various datasets using
    - pandas to investigation data correlation and dimensionality
    - numpy to perform numerical analyses
    - matplotlib and seaborn to visually explore datasets

KEY FINDINGS

- 2017 Hurricane Irma/Maria in Puerto Rico corresponded with the highest disaster assistance funding in our dataset
- 2019 winter storm in Oregon corresponded with the highest US death counts in our dataset
- 2020 COVID-19 in Texas corresponded with the highest hospitalization count in our dataset
- 2018 Hurricane Micheal in Florida corresponded with the highest illness count in our dataset
- **No compelling evidence to suggest a meaningful relationship between the peak death and illness counts the concurrent disaster declarations identified.**

OVERVIEW

Across the US, major weather and climate events appear to be occurring more frequently and with greater intensity.

Consequently, we expected to find observable impacts to the health, safety and economic data reported across the country and sought to identify various datasets to investigate this hypothesis.

Specifically, our aimed was to **examine the impacts of extreme weather and climate events on public health and regional economics** using data analysis.

APPROACH TO DATA EXPLORATION

DATA REQUIREMENTS

Datatype: time-based, location-based
Min scale: 5 years (to include 2018-2022)
Min resolution: county-level
Content: economic, health-related

DATA EXPLORATION

Sources: FEMA, CDC, NCEMS, SBA
Screening: data resolution, level of completeness, ease of accessibility, data available attributes

DATA TRANS-FORMATION

Import: dataset download, API calls
Cleanup: NA/null removal, attribute selection, key identification (for joining data)
Combine: merge by event ID or state/year

DATA ANALYSIS

Assess: trends over time (freq. and cost)
Compare: event count by state and year
Evaluate: outlier data for insights

DATA EXPLORATION *(Dataset 1)*

1

DATASET DESCRIPTION

FEMA Disaster Declaration Summaries

- 65,158 total records
- 25 data attributes, including
  - State
  - County
  - Disaster ID
  - Declaration name
  - Declaration date
  - Declaration type
- Collection via iterative API calls
  (7 API calls due to 10,000 per call limit)
- Data resolution
  - State and county level
  - 70 years by declaration date

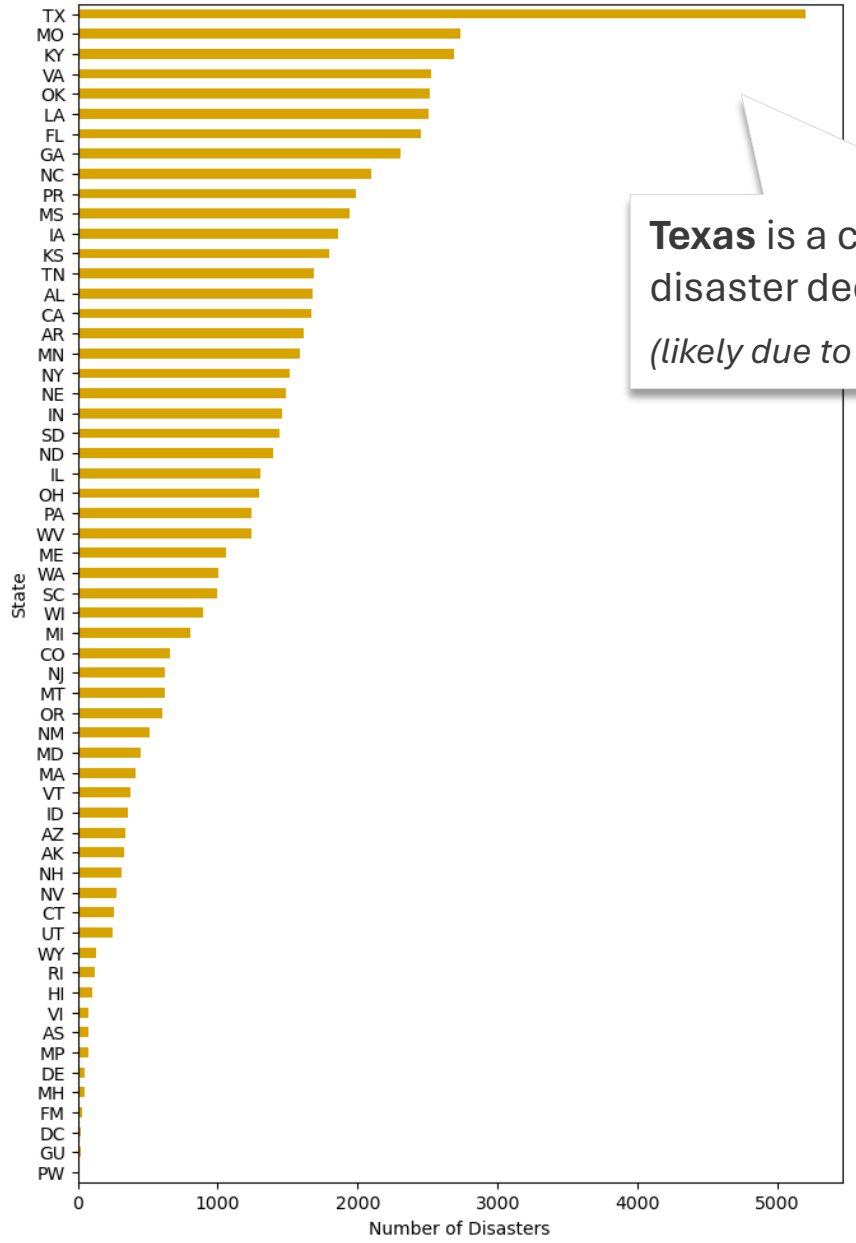
Key Data Profiling Activities

- Viewed column headers (.columns)
- Datatypes and non-null counts (.info)
- Listed unique states represented
- Confirmed first/last datetime values

Key Data Aggregation Activities

- Record count by state
- Record count by disaster ID no.
- Record count by state and disaster date

FIGURE 1 Number of US Disaster Declarations by State (1953 to 2023)



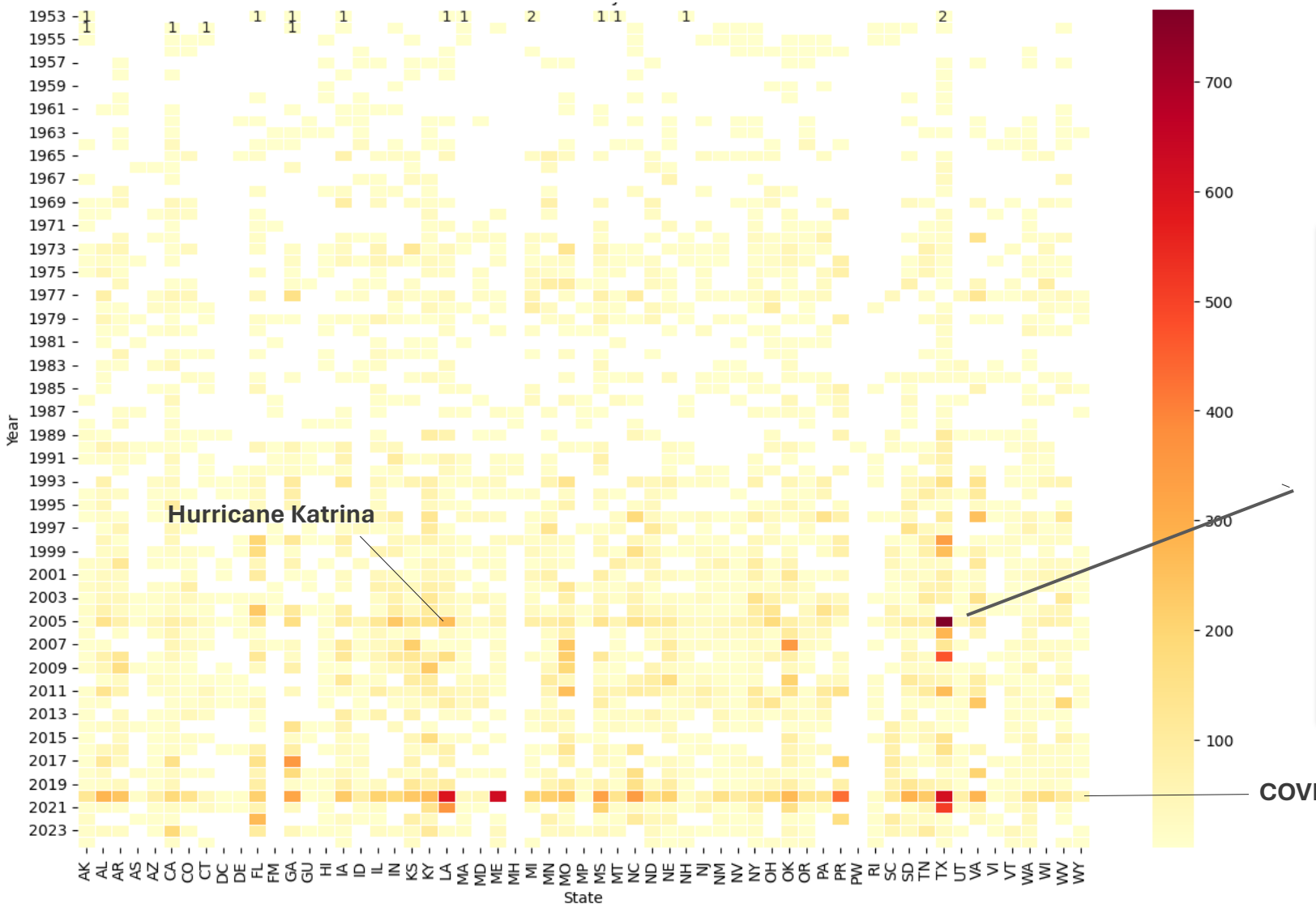
Texas is a clear outlier in number of disaster declarations

(likely due to land area and gulf coast proximity)

During the past 70 years, Texas experienced over 5000 disaster declarations – twice as many as the next

DATA EXPLORATION *(Dataset 1)*

FIGURE 2 **Number of Disaster Declarations by Year and State (1953 to 2023)**



Heatmap by state and number of disasters shows **increasing number** of disaster declarations over time.

TX communities devastated by **Hurricane Rita**

What the data set looked like

DATA EXPLORATION *(Dataset 2)*

DATASET DESCRIPTION

FEMA Public Assistance Funded Project Details

- 790406 total records
- 22 data attributes, including
  - State
  - County
  - Disaster ID
  - Declaration name
  - Declaration year
  - Incident type
  - Damage category
  - Project amount
  - Project size
- Collection via iterative API calls (***80*** API calls due to 10,000/call limit)
- Data resolution
  - State-level (county data missing)
  - 70 years by declaration date
- **Final Dataframe: 792,938 x 14**

Key Data Profiling Activities
- Viewed column headers (.columns)
- Datatypes and non-null counts (.info)
- Viewed unique years and states

Key Data Cleaning Activities
- Dropped non-material columns
- Dropped County columns (high N/A count)
- Converted Project Amount to billion USD

Key Data Aggregation Activities
- Record count by state
- Record count by disaster ID no.
- Record count by state and disaster date

Each disaster may have many projects

Each state has many projects

TABLE 1. TOP 5 BY PROJECT AMOUNT

Year	Disaster Asst. Project	Project Amt	Disaster No.	Disaster Asst.
2020	TX Covid-19 Response	$9.8B	4485	$14.7B
2017	PR Hurricane Response	$6.1B	4339	$33.8B
2017	PR Hurricane Response	$2.5B	4339	$33.8B
2020	NY Covid-19 Response	$1.7B	4480	$14.1B
2017	PR Hurricane Response	$1.5B	4339	$33.8B

FIGURE 3

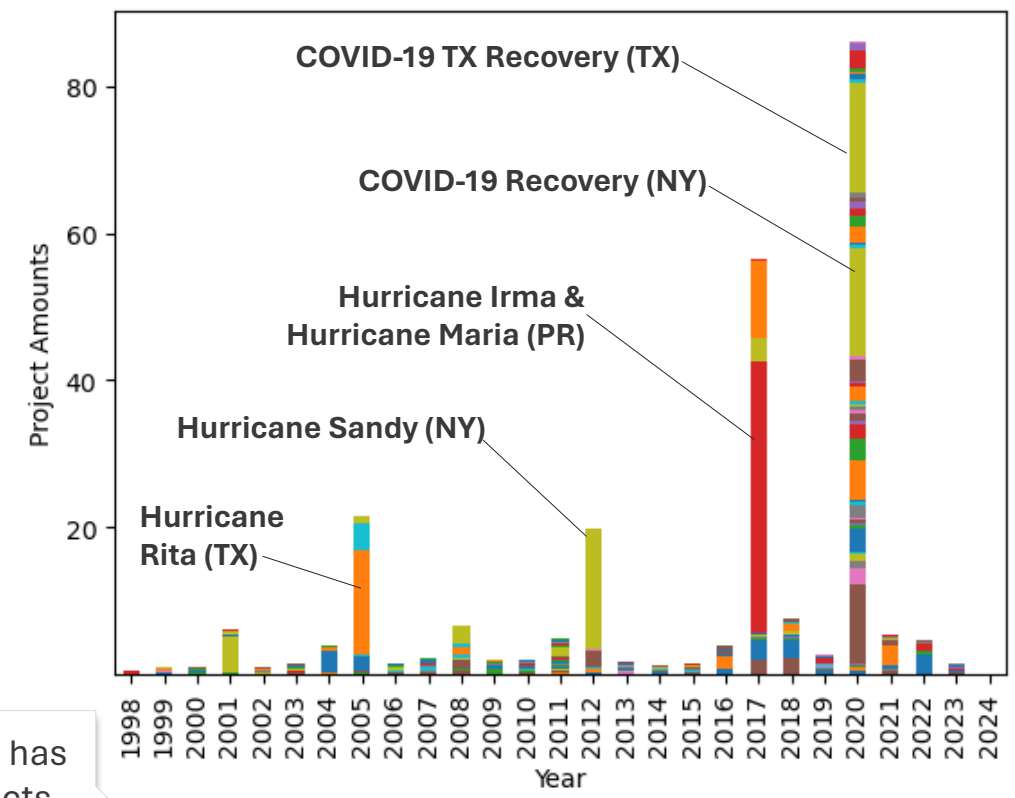
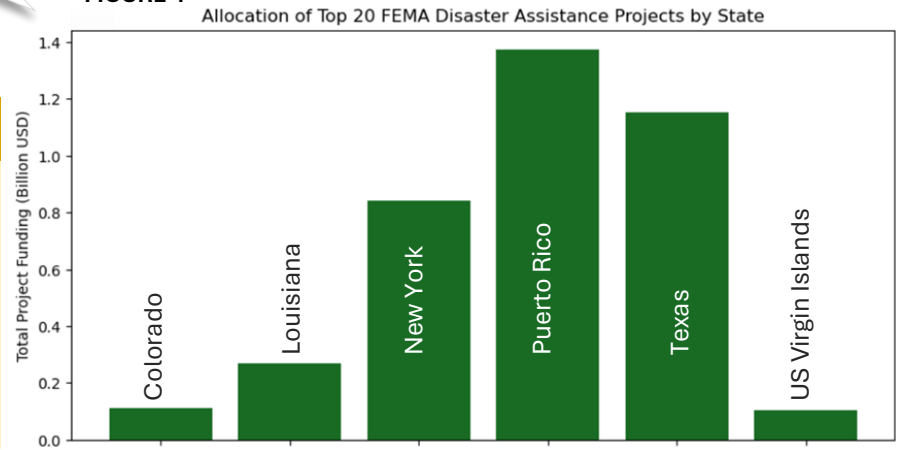
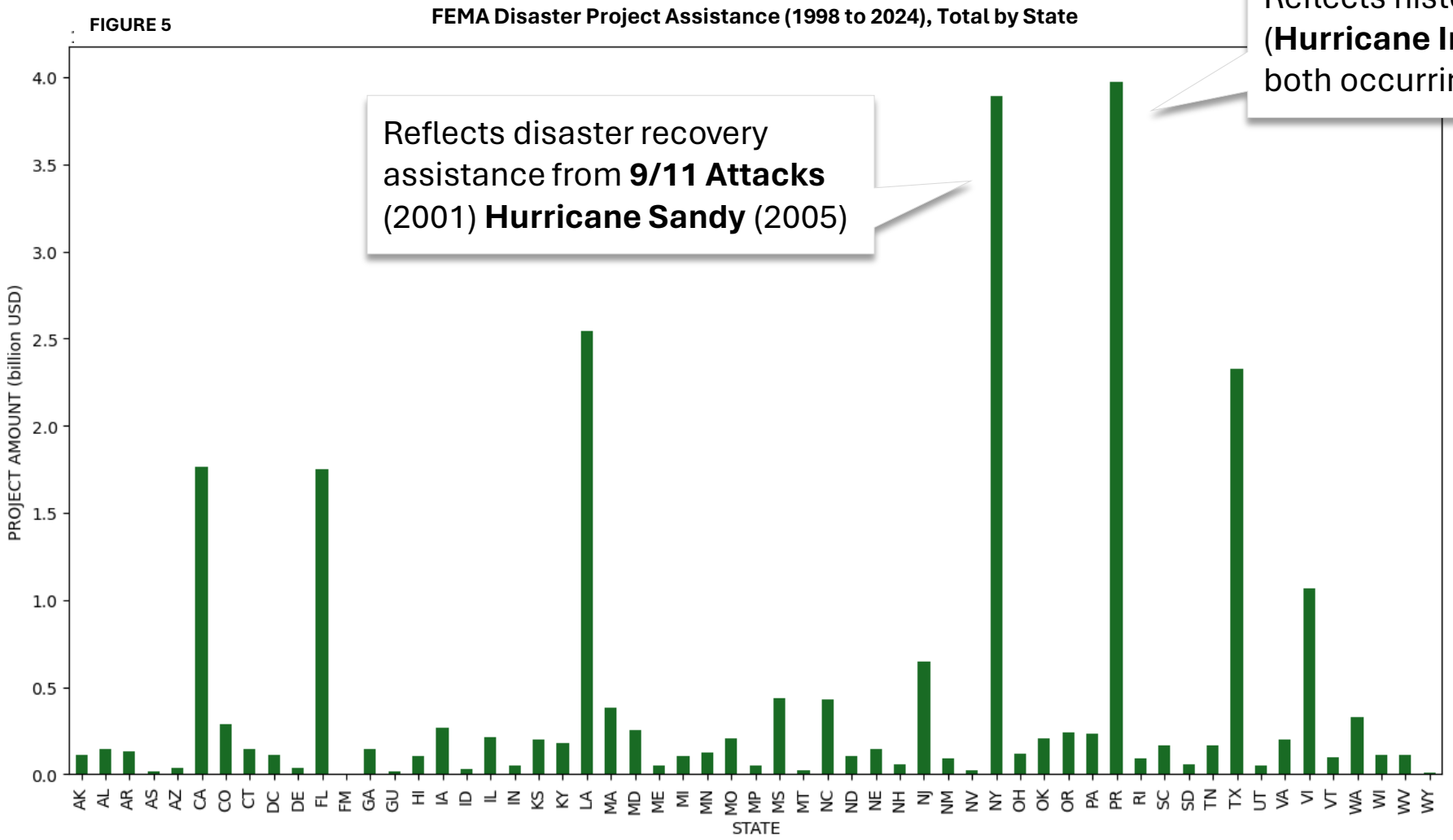


FIGURE 4



DATA EXPLORATION *(Dataset 2)*

FIGURE 5



Reflects disaster recovery assistance from **9/11 Attacks** (2001) **Hurricane Sandy** (2005)

Reflects historic catagory 4+ hurricanes (**Hurricane Irma** and **Hurricane Maria**), both occurring September 2017

Over the past 30 years, may of the states receiving outsized funding have been coastal states, begging the question:

- Project amount driven greater hurricane damage?
- Project amount driven by higher populations among costal states?

DATA EXPLORATION *(Dataset 3)*

DATASET DESCRIPTION

National Center for Disease Control National Outbreak Reporting

- 59736 total records
- 21 data attributes, including
  - Year
  - Month
  - State
  - Hospitalization count
  - Death count
  - illnesses
- Ingested .xlxs file
- Data resolution
  - State-level records
  - Monthly records (1971 to 2023)

Key Data Profiling Activities

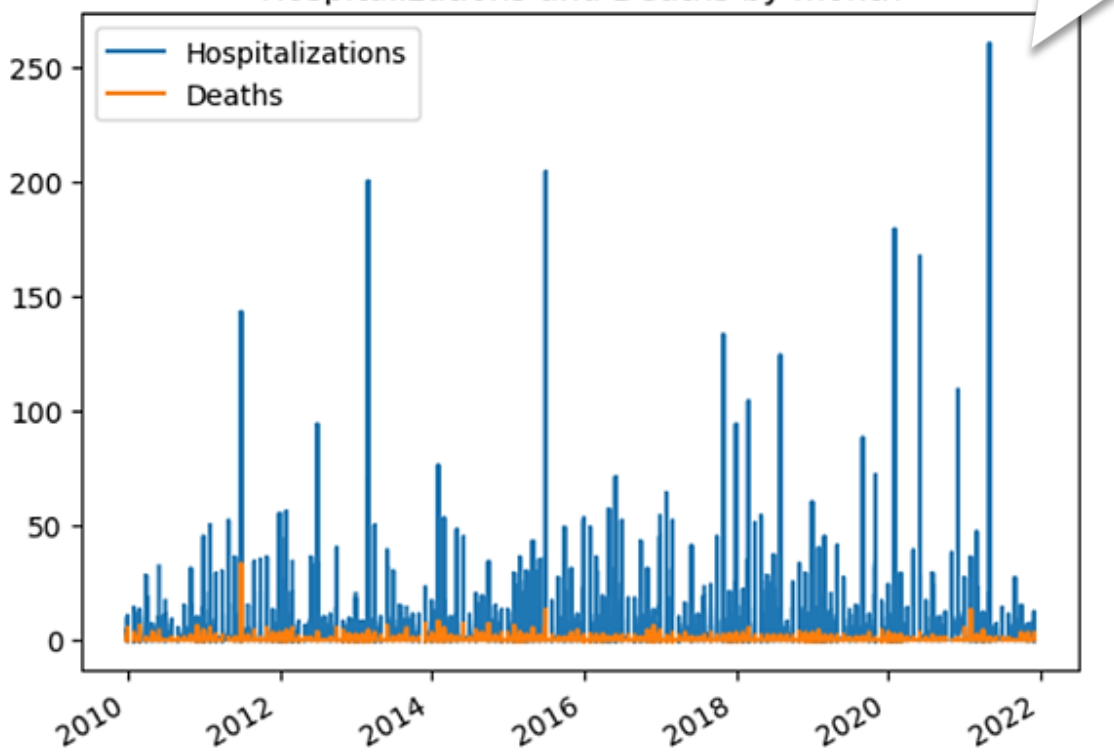
- Viewed column headers (.columns)
- Datatypes and non-null counts (.info)
- Viewed datashape

Key Data Cleaning Activities

- Dropped non-material columns

Loss of overall data resolution when comparing county-level disaster with state-level public health data

FIGURE 7 Hospitalizations and Deaths by month



Publicly accessible national health data was available only at low geospatial resolution due to privacy protections.

DATA EXPLORATION *(Dataset 4)*

DATASET DESCRIPTION

National Center for Disease Control Mortality Multiple Cause-of-Death

- 15,859,795 records
- 30 data attributes, including
  - Total death count
  - Death from cataclysmic event
  - Deaths due to other causes
  - Year
  - Month
- ingested fixed format text files
- Data resolution
  - Monthly death counts
  - Locational data not provided

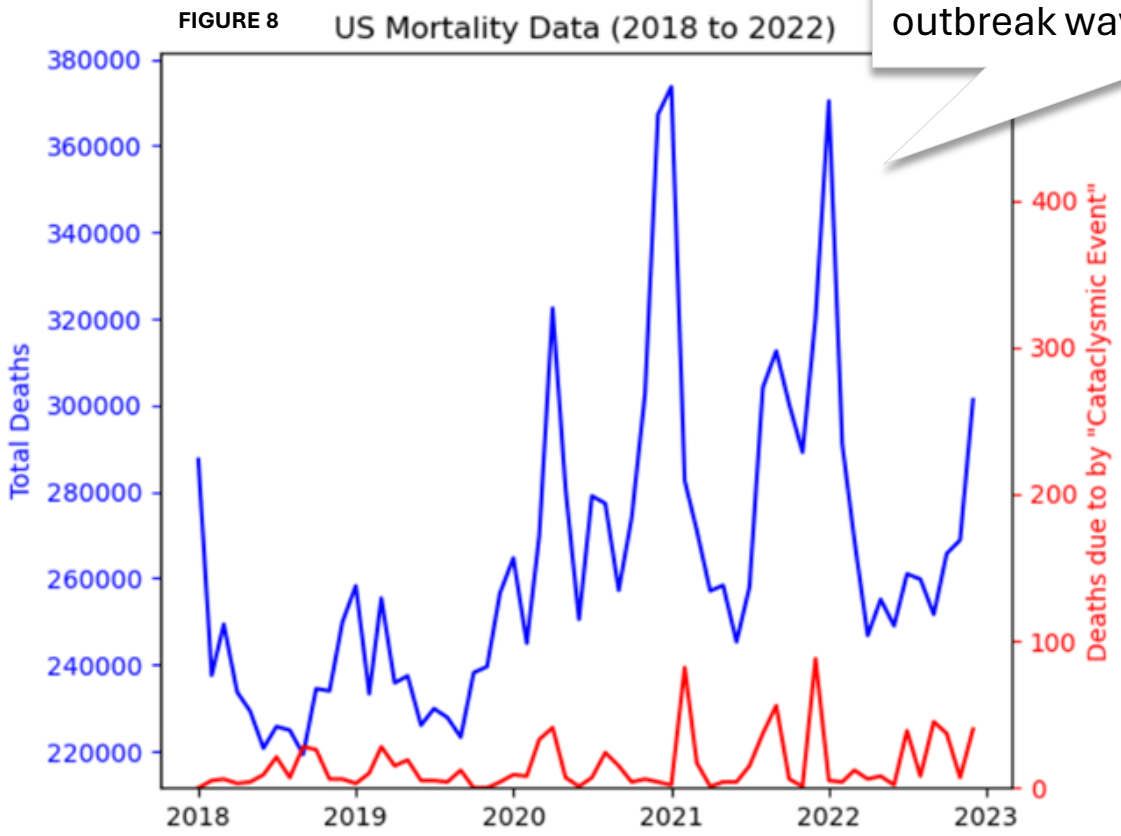
Key Data Profiling Activities

- Viewed column headers (.columns)
- Datatypes and non-null counts (.info)
- Viewed sum for each cause of death

Key Data Aggregation Activities

- Parsed and aggregated 5 fixed format text files into 5 csv summary files.
- Concatenated 5 years into single datafame

FIGURE 8



Double peak in total death count corresponding with **COVID-19** outbreak waves

Total US Deaths over 5-year period: 15,859,795
Death by cataclysmic event: 911
Death by heart failure: 79*
Death by stroke: 302*

Because the COVID-19 disaster was nation-wide, its was the only noticeable public health impact of county-level disasters.

ANALYSIS APPROACH

COMBINED DATASET

Following data exploration and transformation, the investigating team analyzed a final dataset consisting for **430,142 records** with 9 attributes:

- Disaster number
- Disaster name
- Project Amount
- Year
- Month
- State
- Hospitalization count
- Illnesses
- Death count

Datasets Merged into Final Data Set

```
----------------------------------------
Size #  (793031, 14)
Date Range # 1998 2024
Data States # 57

_____________________________________________
Mortality Summary
----------------------------------------
Size #  (60, 6)
Date Range # 2018 2022

_____________________________________________
Outbreak Summary
----------------------------------------
Size #  (59736, 21)
Date Range # 1971 2021
Data States # 58

_____________________________________________
```

Analyzed the relationships between several attributes to determine if relationships between disaster declarations and public health could be detected.

Using the dollar amounts awarded for disaster recovery projects as a measure of disaster cost, the team investigated relationships between disaster declarations and the following:

- Public Assistance Project Amounts (economic)
- Deaths (public health)
- Hospitalizations (public health)
- Reported Illnesses(public health)

RESULTS AND CONCLUSION

KEY OBSERVATIONS

- Texas to be a clear outlier for number of disaster declarations
- 2017 disaster declaration in Puerto Rico (Hurricanes Irma and Maria) and 2020 disaster declarations in TX and NY (COVID-19) three of the highest publicly funded disasters in US history
- Proliferation of privacy protection rules resulted in low resolution public health data being made available for public use (i.e., data not reported by location), dampening its utility for this analysis.

KEY FINDINGS

- 2017 Hurricane Irma/Maria disaster declaration in Puerto Rico corresponded with the highest disaster assistance funding in our dataset
- 2019 winter storm disaster declaration in Oregon corresponded with the highest US death counts in our dataset
- 2020 COVID-19 disaster declaration in Texas corresponded with the highest hospitalization count in our dataset
- 2018 Hurricane Micheal disaster declaration in Florida corresponded with the highest illness count in our dataset

While it may be likely that the 2020 COVID-19 disaster declaration in Texas and peak hospitalization counts are related, the investigating team saw **no compelling evidence to suggest a meaningful relationship between the peak death and illness counts the concurrent disaster declarations identified**

ADDITIONAL INVESTIGATION AREAS

Given access to higher fidelity public health data and additional time to explore more public data, the investigating team may have opted to explore the following areas:

- Evaluation of state-level disaster data normalized by land area and per capita
- Relationship between extreme heat events (historically excluded from disaster declarations) on mortality and agricultural and productivity losses
- Relationship between major flood events and prevalence of mosquito born illnesses
- Possibility of a delayed relationship between public health data and disaster declarations

Reference Content

GitHub

Project Repository: github.com/mikeszumski/Project_One_Repo

Contributors

- Jamie Bond | @JBondAI (dataset identification, storytelling)
- Michael Szumski | @mikeszumski (data wrangling, public health analysis)
- Rajesh Velamala | @REJESHVELAMALA, (data wrangling, disaster analysis)

Data Providers

- Federal Emergency Management Agency (FEMA)
- National Center for Disease Control