Flood Disasters in 2020 and 2021: Exploratory Data Analysis Report

Maysen Pagan

2023-11-06

1 Motivations and Initial Questions

2 The Data

The data for this natural disaster analysis begins with two datasets from the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) and two datasets from the Federal Emergency Management Agency (FEMA).

2.1 NOAA

Data from NOAA provides recordings of all storm events from 2020 and 2021. Data sets from each year provide significant information including the beginning and ending time and date of the disaster, the type of disaster, the location of the disaster, as well as injuries, deaths, and property damage as a result of the disaster. In each year, around 61,000 storm events occurred resulting in around 61,000 rows with 17 variables for each dataset. The raw data can be found here.

2.2 FEMA

One of the datasets from FEMA is the Disaster Declarations Summary which provides information on all federally declared disasters beginning from the year 1953. The data lists the disaster declaration type of either a major disaster, fire management, or emergency declaration. Additionally, a more specific description of the incident is provided as variable `incidentType` as well as what recovery program was declared for the disaster. The other FEMA Disaster Summaries dataset proides information on financial assistance provided to the communities affected by select events. It includes the number of approved Individual Assistance requests,

and individual, public assistance, and hazard mitigation grant amounts. The raw data can be found here.

3 Data Organization and Cleaning

3.1 Dates and Disaster Event

The first step in organizing the data to prepare for exploratory data analysis is filtering all the data sets to only include the disaster event of floods in the years 2020 and 2021. Both NOAA datasets are already filtered by year, so they were then filtered to contain rows where the variable `EVENT_TYPE` included the string `Flood`. Following this filter, the NOAA datasets in 2020 and 2021 include the injury, deaths, and property damage for flash floods, floods, coastal floods, and lakeshore floods. Additionally, the data from the FEMA Disaster Declarations Summary was filtered so that the variable `incidentType` only includes floods, severe storms, coastal storms, tropical storms, tsunamis, and typhoons that began in 2020 and 2021. Although all of these disasters are not explicitly floods, they were still included in analysis as they are disasters may have resulted in flooding in communities.

3.2 Joining

The assistance data from FEMA includes aid provided to communities from select disasters. This data set does not provide the location or type of disaster but rather only includes the unique number assigned to each disaster by FEMA. The Disaster Declarations Summary from FEMA also includes the unique disaster number along with location and disaster type. These two data sets were combined using a `left_join` so that every row in the Disaster Declarations Summary data now includes the aid provided from the assistance data.

3.3 Damage to Property and Crops

The final step in cleaning the data for exploratory data analysis was converting `DAMAGE_PROPERTY` and `DAMAGE_CROPS` from the NOAA datasets to dollar values. The values for these variables were characters with a decimal and a "K" for thousands of dollars or "M" for millions of dollars. To convert to numerical dollar values, the characters "K" and "M" were extracted from each variable column and each column was then converted to a numeric value then multiplied by 1000 or 1000000 depending on whether the extracted character was a "K" or "M" respectively.