Data Report: Impact of Climate-Related Events on Public Support for Climate Action in the United States

Main Question

Are people living in states with more casualties caused by climate-related events more likely to support countermeasures against climate change?

Datasources

1. Datasource: Storm Events Database 2020

This dataset, published by the National Centers for Environmental Information through the National oceanic and atmospheric administration, contains detailed statistics on personal injuries, fatalities, and financial damages caused by weather events in 2020, categorized by U.S. state. The dataset is well-structured and verified, with clear metadata on event types, damages, locations and source of claim and is stored as a csv file.

Furthermore each row contains information about the source of the reported data entry and more details about the respective weather event which are of lesser importance for this project. It represents all the data of the year 2020 published by the NCEI.

The data published by the National domain-specific data portal of the NOAA and is in a structured form with no missing or problematic values regarding the relevant data.

Example of data structure:

someColumn1	STATE	someColumn2	INJURIES_DIRECT	INJURIES_INDIRECT	DEATHS_DIRECT	DEATHS_INDIRECT	someColumn3
...	GEORGIA	...	3	0	0	0	...
...	KANSAS	...	0	0	0	8	...
...	KANSAS	...	0	3	0	0	...
...	KANSAS	...	5	0	0	0	...
...	COLORADO	...	0	0	0	0	...
...	COLORADO	...	0	0	5	0	...
...	KANSAS	...	6	0	0	0	...
...	COLORADO	...	102	0	0	0	...

- Metadata URL: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/Storm-Data-Bulk-csv-Format.pdf
- Data URL: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/StormEvents_details-ftp_v1.0_d2020_c20240620.csv.gz
- Data Type: CSV
- License: NCEI is part of NOAA, an U.S.A Goverment agency. The NOAA website (under license) states that all their data is freely accessible and usable as it's a goverment agency and all goverment data made public by employees in offical duty is freely accessible and redistributable (which this dataset is because "Storm Data is provided by the National Weather Service (NWS) [...]"https://www.ncdc.noaa.gov/stormevents/details.jsp. The general policy of NOAA states, that per default the data is openly accessible and usable for everyone unless explicitly stated otherwise (mostly CC0-License) (https://www.ncei.noaa.gov/archive). While the dataset does not have a license attachted to it, it is published on data.gov the us open data portal. Since it's published through NCEI, consequently their data is under "U.S. Public Domain" or "CC0", hence "open" and freely accessible https://www.ncei.noaa.gov/sites/g/files/anmtlf171/files/2023-12/NCEI%20PD-10-2-02%20-%20Open%20Data%20Policy%20Signed.pdf. NOAA asks to include a notice of what agency the data originates from which will be complied with in this project through a textfile.

2. Datasource: Yale Climate Change Opinion Map 2020

This dataset represents the summarized statistics about the public opinion on climate change related question regarding importance, countermeasures, awareness on the topic, etc. It averages the importance from 0% to 100% for each question per U.S. state and on county level in the year of 2020. The data was published and evaluated through polls by the private Yale University through their climate change action program. Each row represents the avg. values for each U.S. state and county with over 30 columns each representing a question. The data quality is well structured with no missing values or errors regarding the relevant data collected through a representative study and hence was chosen.

Example of data structure:

GeoType	GeoName	CO2limits	drilloffshore	...
State	Wyoming	70,49	85,49	...
County	Houston County, Alabama	45,6	32,45	...
State	Montana	33,45	68,98	...

- Metadata URL: https://raw.githubusercontent.com/yaleschooloftheenvironment/Yale-Climate-Change-Opinion-Maps/refs/heads/main/YCOM5.0_2020_webdata_Metadata.csv
- Data URL: https://raw.githubusercontent.com/yaleschooloftheenvironment/Yale-Climate-Change-Opinion-Maps/refs/heads/main/YCOM5.0_2020_webdata_2020-08-19.csv
- Data Type: CSV
- License: MIT License on GitHub issued by Yale officials. Hence it's freely usable for this project https://github.com/yaleschooloftheenvironment/Yale-Climate-Change-Opinion-Maps. The License is to be included in the project and noted in all corresponding presentations.

Data Pipeline

The core tool for the data pipeline is jayvee. The project consist out of two pipelines. The entrypoint is the pipeline.sh file which calls a downloadAndUnzipData.sh. This is nessecary to download and unzip one of the two data sets with curl and gunzip.

Then the main jayvee project file is executed. It consists out of two sequential pipelines where one picks up the downloaded file with a LocalFileExtractor and one downloads it's file with an HttpExtractor. After some configuration of the csv interpretation only the needed columns are selected by explicitly listing the predefined names in a TableInterpreter and ignoring all unnessecary columns and their values. The rows are saved with the most fitting built-in value types, except for the GeoType of the "Climate Change opinion map" data set where a custom data type is used to only allow the type "State" and thus filtering for only rows that belong to states and excluding counties. Finally both data sets get each an additional column with a transformer. The Storm Events data set gets a "TOTAL_CASUALITIES" data column with the sum of all 4 casulties categories to summrize the total amount of casualties. The opinion map data set gets a "AverageOpinionTrend" column to save the average relative climate change countermeasures support trend averaged for all the relevant climate change items. Both columns and each values will be the key variables for the analysis. No data was nessecary to be cleaned.

- Problems: As mentioned above, i couldn't download one of the data sets via the HttpExtractor due to it being downloaded as a .gz compressed file which the HttpExtractor can't handle. Hence my solution was to download it with curl and unzip it to provide it localy. Nonetheless the setup provides dynamic downloading of the file each time the pipeline is run from the script.

- Meta-quality measures: The retries of the HttpExtractor are set to 3 to avoid race condition and temporary failures. The curling of the compressed data set has no retries which might lead to issues, but the succesful downloading of the first data set is required for the pipeline start which avoids a half-executed pipeline. All the files that are created localy are replacing a (possible) prior one so multiple sequential executions of the pipeline are possible.

The following tables show how the data sets would look like after the pipeline in the sqlite files:

GeoType	GeoName 2	CO2limits	drilloffshore	...	AverageOpinionTrend
State	Wyoming	70,49	85,49	...	77,99
State	Montana	33,45	68,98	...	...

STATE	INJURIES_DIRECT	INJURIES_INDIRECT	DEATHS_DIRECT	DEATHS_INDIRECT	TOTAL_CASUALITIES
GEORGIA	3	0	0	0	3
KANSAS	100	212	0	8	312
KANSAS	0	3	0	0	3
KANSAS	5	0	0	0	5
COLORADO	0	0	0	0	0
COLORADO	0	0	5	0	5
KANSAS	6	0	0	0	6
COLORADO	102	0	0	0	102

Result and Limitations

The data pipeline results in two structured data sets as sqlite data base files. They contain the key variables which are nessecary for the analysis aswell as columns that are nessecary to map the data rows to eachother and analyse the combined data sets later on. The data sets have no empty or unintended values. SQLite was chosen for the upcoming analysis. Due to this data type it will be possible to first, group the values of the storm event data set on a state level. This is nessecary due to the presence of multiple rows with different casulties values per state. The rows in the original data set do not summarize the values per state. Second, to join the AverageOpinionTrend and TOTAL_CASUALITIES values per row where the value of STATE, a text value of the state name which is identical in each data set, will be used as the variable to join the tables on. This is very easily possible with sql and hence sqlite is chosen. This can result again in an sqlite db or similar files, which can be used in other software like R which works well with sqlite aswell.

- Potential issues: Some temporary downtimes of the data source could lead to problems. Especially the public opinion map which is on a GitHub repository of a yale university account. The data set could technically easily be deleted by yale staff or some temporary downtime of GitHub which is a private company. On the other hand the NOAA as a governmental data set provider could also have some temporary downtimes. Regarding future proceedings for analyzing the data it's not fully clear whether joing the tables will go entirely smoothly due to maybe some small differences in the names of states which will be used to join the tables.