

## Storm Events Dataset Write-Up

Kerim Celik, Mandy Cisler, Elizabeth Morgan, Kevin Nazario, David Robles

01/18/2023

From our initial three research questions, our investigation evolved to answering the final three questions found in our presentation. The first question on trends stayed largely the same, but expanded to look at the most recent 15 years. Our initial second question about regions changed into our third about El Nino / La Nina, and dropped consideration of demographics due to time constraints. Our initial third question changed into our second question, and looked at tornados in particular due to their presence across all years in the dataset. The following were our findings:

Research Question 1: What has been the impact of storms over recent years? Is there a correlation between storm fatalities and total damages? What are the differences in storm impacts between the most recent 15 years and the initial 15 years?

- The deadliest storms between 2008 - 2022 were caused by the following kinds of storm events: tornado, excessive heat, wildfire, hurricane, and heat.
- Frequency of storms does not align with which storms were the deadliest (though there was some overlap)
- Across the 2008-2022 dataset, there were a few years that had higher death/injury totals. 2011 in particular stands out and may be explained by a few very severe storms in that year (3 of the deadliest storms were from 2011).
- Our analysis found a moderate relationship between deaths and damages which predicts that as deaths increase, total cost of damages will increase. This confirms our initial expectation.
- Although there is a moderate relationship between the deaths and damages, more analysis needs to be conducted to understand the nuances of this relationship and its reflection inside the dataset.
- After comparing the total number of deaths for a 15-year span for the earliest set of data available (1950-1964) and the most recent 16 years of data (2009-2023), we can conclude that storms recorded in the most recent data set (2009-2023) were more deadly.
- The average amount of total deaths per year between 1950 and 1964 was 107, and the average amount of total deaths per year between 2009 and 2023 was 475.27. The 2009-2023 data set had no outliers, while the 1950-1964 data set had two outliers: 231 deaths in 1952 and 517 deaths in 1953.

Research Question 2: Given that tornadoes are one of the most frequent event types and among the most destructive and deadly in our dataset, do tornadoes with greater magnitudes result in increased property damage?

- The R-squared value for the linear regression model without outliers is approximately 0.0763.

- Urban areas constitute a significant portion of the population but occupy a small land space. Tornadoes that occur in urban areas may cause more property damage due to the concentration of infrastructure and higher property values. On the other hand, large tornadoes in rural areas may not lead to substantial property damage as there might be fewer structures and less valuable property.

Research Question 3: Can we answer any larger related questions? For example, can we observe the effects of larger weather patterns in this dataset (i.e. El Nino / La Nina)?

- Finding statistical relationships with phenomena not directly related to what is contained in our dataset is very difficult. Due to the uncertain nature of those relationships, any seemingly significant data is called into question.
- Having noted that, many notable recent years show a significant difference from the older data. In particular, when looking at storm count per region affected by El Nino / La Nina, years classified as strong are significantly different from the mean of counts for 2000-2011 using a chi-square analysis.
- Attempting to correlate the strength of El Nino / La Nina to the number of storm events per region was largely unsuccessful. The greatest Pearson R-value obtained was 0.52 and the greatest R-squared value was 0.27; many of the other values were much worse.
  - If more time was available, we might consider splitting types of storm events to those caused by dry/wet or hot/cold climate. We could then count only those events in the regions with that climate from El Nino / La Nina for the year to try and obtain better correlation results.
  - Overall, it seems like the way El Nino / La Nina phenomena are rated is not very conducive to the types of analysis we tried. A full reconsideration of approach here would be best.