

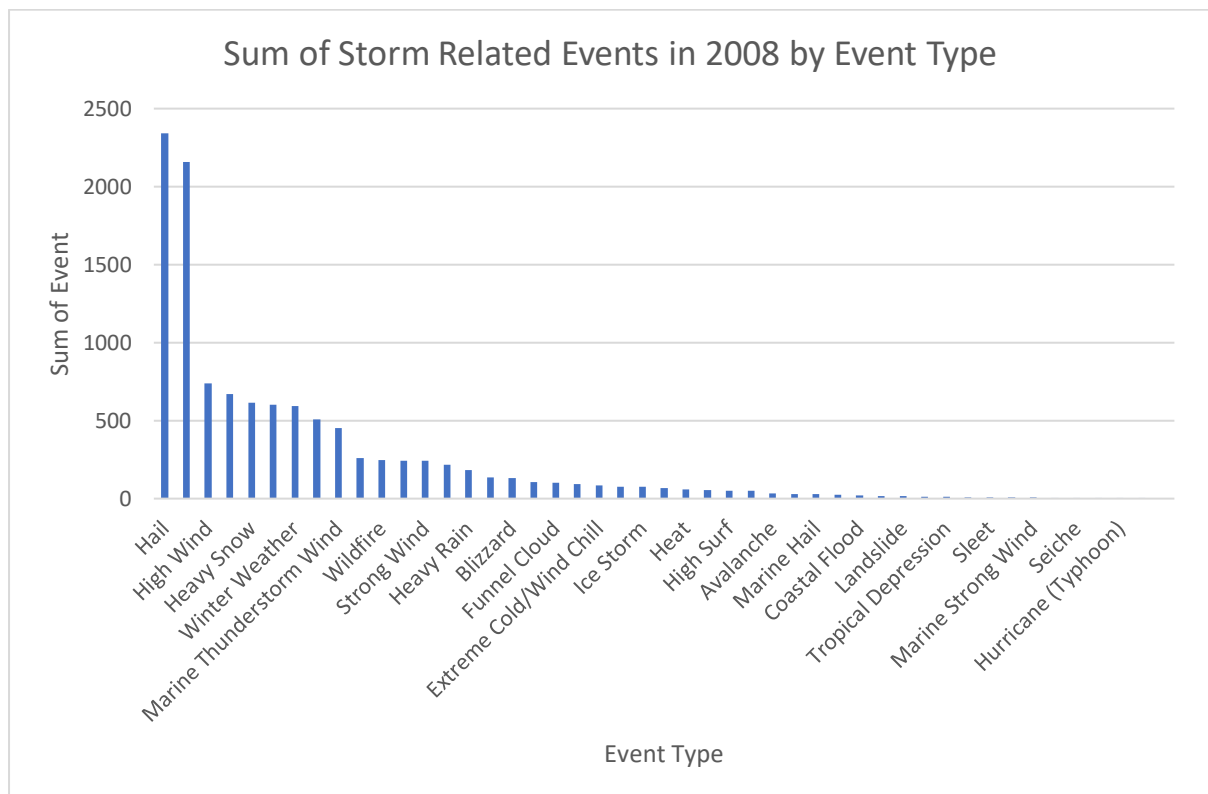
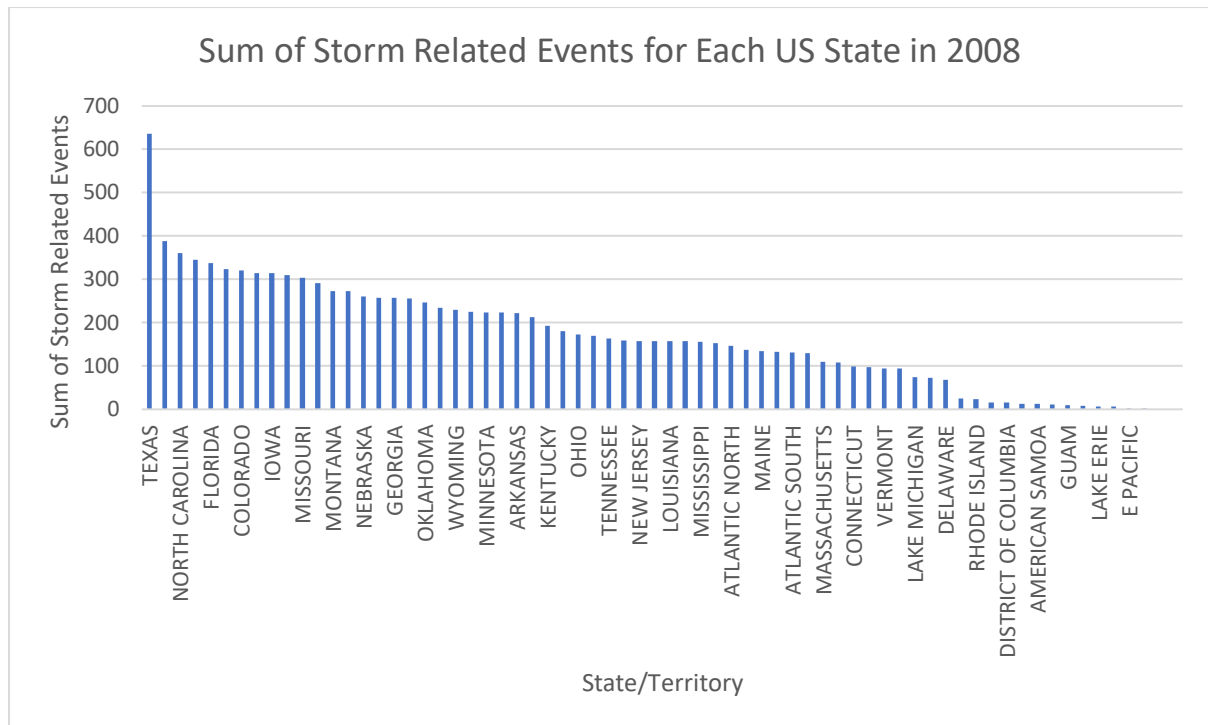
Database Design and Implementation
Professor Engel
Gerald Steven
Assignment 2, Part IV

For Assignment 2, I chose a dataset from the National Oceanic and Atmospheric Administration (NOAA). The dataset, which can be retrieved from the following URL: <https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/>, contains a summary of all storm related events that happened in 2008. The original unfiltered dataset had 51 columns and over 71,000 records, and after some cleaning, the refined dataset now has 13 columns and over 11,000 records. I chose this dataset because of my interest in environmentalism and climate change. With extreme and unpredictable weather happening more frequently, analyzing data of storm events in the US can be helpful in finding patterns over the years. The data source - NOAA's public database - contains storm data from 1950 to 2019, but I chose the year 2008 in particular because of its large size, which allows me to experiment and manipulate a larger volume of data. The dataset also has a mix of both quantitative data (number of injuries, number of deaths) and categorical data (state, event type), which allowed me to experiment and manipulate different types of data.

I believe that this data source is reliable because NOAA is one of the leading American scientific agencies that collect data on the ocean, atmosphere, and weather. NOAA updates their website annually, as they already have dataset for the year 2019. Each dataset is also detailed and consistent in formatting, as each record contains 51 variables (or columns) in the dataset. This dataset in particular can be useful for analyzing patterns in weather changes due to climate change. Because NOAA has data all the way back from 1950, the number of storm related events can be calculated for every year to see if extreme weather happens more frequently because of climate change. The severity of each storm event can also be documented by analyzing the number of injuries, deaths, and by reading the EVENT_NARRATIVE and EPISODE_NARRATIVE columns, which contain descriptions of each record. Although

climate change itself can be difficult to observe, NOAA's storm data could be compared with other factors that are related to climate change, such as global carbon emissions over the years, to find the correlation between extreme weather and climate change.

I made two tests on Excel to check for outliers, and used three charts to visualize the tests. The first test is to make a bar chart that displays the total number of storm related events in 2008 by state. I first created a pivot table to show the result in descending order, and then create a bar chart from the table. The result shows that Texas by far has the higher number of storm related events in 2008, over 636 events in one year. The state with the second highest number is Illinois with 388 events, and the state or territory with the lowest number is Gulf of Alaska with 2 events in 2008. My second test was to find the event type that occurred the most frequently in 2008. I used a similar technique where I used a pivot table in descending order, and then creating a bar chart. The result shows that hail was the most common occurring event in 2008, with 2343 reported cases. The second most frequent event was thunderstorm wind with 2158 cases, and the least was volcanic ashfall, with 2 reported cases. To give perspective to the proportion of these event types, I also created a pie chart from the same table to see what percentage of all the storm events in 2008 were hail or thunderstorm wind. The pie chart reveals that 21% of all storm events in 2008 was hail, and 19% was thunderstorm wind. The three charts are appended below for reference.



Storm Related Events in 2008 by Event Type

