

Prabhjot Singh

DS 5001

10 May 2021

Manifest File

Provenance: The data was pulled from the National Oceanic and Atmospheric Administration Storm Event Database. Currently it contains every storm between January 1950 to January 2021. For the purposes of this project the interval was limited to between 1993-2021.

Location: The database can be found at the following link:

<https://www.ncdc.noaa.gov/stormevents/>

After clicking on the link, navigate to 'Bulk Data Download' under the Data Access tab.

Description: Weather narratives were obtained from the National Oceanic and Atmospheric Administration storm database. The original file that was scraped featured every storm from the years 1950 to 2020 in the United States. However, after data cleaning and processing, the corpus was limited to the interval of 1993 to 2020.

The weather narratives themselves are descriptive accounts on what happened during a particular storm and formed by the National Weather Service. Beyond this, the meta data itself has many interesting features. It has highly specific markers such as the location, time, and day of the storm. Then depending on the storm type, other qualifying metrics such as the number of fatalities, magnitude of the tornado, cause of flood etc., were included.

Format: The format of the data is in csv.gz files. So, with the use of some other python code that was pulled from another project ([link](#)), I was able to compile the data.