# Analysis of severe weather events in the united states between the years 1950 and 2011

MATIAS, V. A.

28/03/2021

## 1. Synopsis

This is the final project of the Reproducible Research Course offered by Johns Hopkins University on the Coursera platform.

This paper will present an analysis of climate events based on the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database.

The programming language R will be used as the basis for all the analysis to be presented, this document being developed with R markdown (Rmd).

### 1.1 Data

The available data are in a compressed CSV file for type .bz2, resulting in a 42Mb file.

Data: Storm Data

Documentation: National Weather Service Storm Data Documentation

FAQ: National Climatic Data Center Storm Events FAQ

## 2. Data Processing

This section covers the techniques used in the collection and processing of data

### 2.1 Required libraries

Before starting the analysis, some packages will be needed to facilitate development:

- dplyr: To facilitate data manipulation
- ggplot2: To create the graphics

```
library(dplyr)
library(ggplot2)
```

## 2.2 Geting the data

The project requires that the compressed CSV files be in the directory where the analysis will be performed. This step can be performed by the user himself, downloading the file and placing it in the analysis directory, or by the code below.

```
data_url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(data_url, destfile = "storm_data.csv.bz2")
```

## 2.3 Loading the data

Now the downloaded file with the name *storm_data.csv.bz2* will be stored in memory. Note that this is a large file when uncompressed, requiring almost 500Mb to allocate it in RAM.

```
weather <- read.csv("storm_data.csv.bz2")
object.size(weather)
```

```
## 494295216 bytes
```