**Weather analysis of United States from the year 1950 - 2011**

**Synopsis**: The analysis is done to predict the type of events that is the most harmful for human population as well the one which incurs the most property damage. This analysis is done using the R language in RStudio and uses some basic R packages to perform the analyses.

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site:

* [Storm Data](https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) [47Mb]

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

* National Weather Service [Storm Data Documentation](https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)
* National Climatic Data Center Storm Events [FAQ](https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf)

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

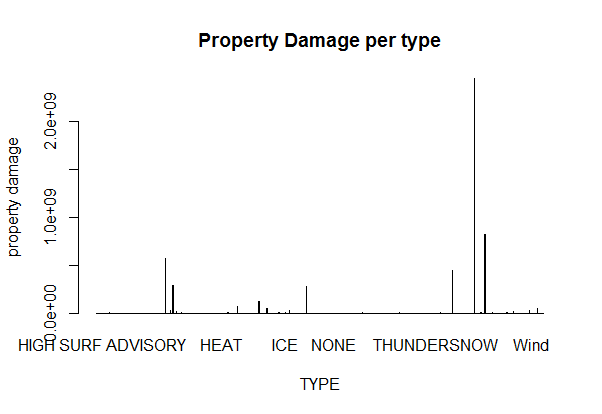
**Data Processing**: After the data was downloaded it was loaded using read.csv() function

*StormData <- read.csv("E:/My Stuff/R/repdata-data-StormData.csv/repdata-data-StormData.csv", stringsAsFactors=FALSE)*

After which using the summary() and str() function, the behavior of the dataset was studied.

**Results**: After applying some functions and processing the data it was found that

|  |
| --- |
| The maximum count of event happening in the US from the year 1950 to 2001 is the **HAIL** |
| The maximum number of people dead or injured due to this are because of **Tornado**. |
| The maximum number of property damage is because of **Tornado**. |

****