

# Fatalities, Injuries and Economic Impact of Weather Events in the US (1950-2011)

*Gaurav Tripathi*

*Saturday, December 20, 2014*

## Introduction

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.

## Synopsis

Following are the findings from this analysis: - Tornadoes caused the maximum number of fatalities between 1950-2011 - They also caused the maximum number of injuries, in the same period - Excessive Heat was the next in terms of number of fatalities - Thunderstorm (TSTM) Wind, Flood and Excessive Heat were the next 3, in order of number of injuries to people. The number of injuries for these 3 was close to each other - Flash Floods caused the maximum economic damage to property between 1950-2011 - Second in terms of economic damage to property was Thunderstorm Wind - Drought caused the maximum economic damage to crops in the same period - This was followed by Flood, which also caused economic damage to crops

## Data Processing

In order to process the data, Storm data was downloaded from the web URL: <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>. Given that it was a bz2 compressed file, the data was unzipped, and loaded into a data frame. Following is the code for the same:

```
# Download and unzip file
if (!file.exists("ReproResearch2")) {
  dir.create("ReproResearch2")
}

link <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
destin <- "./ReproResearch2/storm_data.csv.bz2"
download.file(link, destfile = destin, method="wget")

## Warning: running command 'wget
## "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
## -O "./ReproResearch2/storm_data.csv.bz2"' had status 127

## Warning in download.file(link, destfile = destin, method = "wget"):
## download had nonzero exit status
```

```
# Read CSV file
storm <- read.csv(bzfile("./ReproResearch2/storm_data.csv.bz2"))
```

On closer analysis of “storm” data set, multiple issues were seen: 1. Same event types were captured but with different cases 2. Same event types were captured but with different punctuations 3. Leading and trailing space 4. Typographical errors 5. Double spacing in between words

Issues #1 - #3 were resolved by the following code chunk, while #4 and #5 were not addressed:

```
# Identify and clean the event types information
## 1. Combine events mentioned in different cases (e.g. TORNADO and Tornado)
## 2. Combine events mentioned with different punctuation (e.g. 'waterspout tornado' and 'waterspout/to
## 3. Remove leading and trailing space
length(unique(storm$EVTYPE))      #Before cleaning of event type data
```

```
## [1] 985
```

```
event_type <- tolower(storm$EVTYPE)
event_type <- gsub("[[:blank:]][:punct:]+", " ", event_type)
event_type <- gsub("^\\s+|\\s+$", "", event_type)
storm$EVTYPE <- event_type

length(unique(storm$EVTYPE))      #After cleaning of event type data
```

```
## [1] 856
```

After cleaning the data, casualties (fatalities and injuries) were summed, grouped by event type.

```
# Sum casualties and injuries by event type into separate data sets
library(plyr)
casualties <- ddply(storm, .(EVTYPE), summarize, sum_fatality = sum(FATALITIES),
                    sum_injuries = sum(INJURIES))

fatal_events = casualties[order(casualties$sum_fatality, decreasing = T), ]
injury_events = casualties[order(casualties$sum_injuries, decreasing = T), ]
```

Next, Actual economic damage was calculated, for both property and crops, by combining values of “PROPDMG” with “PROPDMGEXP” and “CROPDMG” with “CROPDMGEXP”, as follows:

```
# Data carries units in different columns "PROPDMGEXP" and "CROPDMGEXP"
# Hence the power of exponent needs to be calculated separately
# h -> Hundred (10^2), k -> Thousand (10^3), m -> Million (10^6), b -> Billion (10^9)
pow_exponent <- function(e) {
  if (e %in% c('h', 'H')) return(2)
  else if (e %in% c('k', 'K')) return(3)
  else if (e %in% c('m', 'M')) return(6)
  else if (e %in% c('b', 'B')) return(9)
  else if (!is.na(as.numeric(e))) return(as.numeric(e))      #If Digit
  else if (e %in% c(' ', '-', '?', '+')) return(0)           #If Space or Special Character
  else stop("Invalid exponent value.")
}
```

```

# Calculate actual damage
prop_dmg_power <- sapply(storm$PROPDGMGEXP, FUN=pow_exponent)
storm$prop_dmg <- storm$PROPDMG * (10 ** prop_dmg_power)
crop_dmg_power <- sapply(storm$CROPDMGEXP, FUN=pow_exponent)
storm$crop_dmg <- storm$CROPDMG * (10 ** crop_dmg_power)

# Calculate economic loss and remove events which caused zero loss
econ_loss <- dply(storm, .(EVTYPE), summarize, prop_dmg = sum(prop_dmg), crop_dmg = sum(crop_dmg))
econ_loss <- econ_loss[(econ_loss$prop_dmg > 0 | econ_loss$crop_dmg > 0), ]

prop_dmg_events <- econ_loss[order(econ_loss$prop_dmg, decreasing = T), ]
crop_dmg_events <- econ_loss[order(econ_loss$crop_dmg, decreasing = T), ]

```

Given the large number of events, graph was plotted for the Top 20 events: - Causing Max. number of Fatalities - Causing Max. number of Injuries - Causing Max. damage to property - Causing Max. damage to crops

```

top20_fatal_events <- fatal_events[1:20,]
top20_injury_events <- injury_events[1:20,]
top20_property_dmg_events <- prop_dmg_events[1:20,]
top20_crop_dmg_events <- crop_dmg_events[1:20,]

# Plot Top 20 Fatalilty causing Events
library(ggplot2)
library(gridExtra)

g1 <- ggplot(data=top20_fatal_events, aes(x = reorder(EVTYPE, sum_fatality), y = sum_fatality)) +
  geom_bar(stat="identity") + ylab("Total number of fatalities") +
  xlab("Top 20 Fatalilty causing Events") + coord_flip()

# Plot Top 20 Injury causing Events
g2 <- ggplot(data=top20_injury_events, aes(x = reorder(EVTYPE, sum_injuries), y = sum_injuries)) +
  geom_bar(stat="identity") + ylab("Total number of injuries") +
  xlab("Top 20 Injury causing Events") + coord_flip()

grid.arrange(g1, g2, main="Top Fatal and Injury-causing events in the US (1950-2011)")

# Plot Top 20 Events by their Economic Impact on Property
g3 <- ggplot(data=top20_property_dmg_events, aes(x = reorder(EVTYPE, prop_dmg), y = log10(prop_dmg))) +
  geom_bar(stat="identity") + ylab("Total Economic Damage to Property (log Base 10 Scale)") +
  xlab("Top 20 Events") + coord_flip()

# Plot Top 20 Events by their Economic Impact on Crops
g4 <- ggplot(data=top20_crop_dmg_events, aes(x = reorder(EVTYPE, crop_dmg), y = crop_dmg)) +
  geom_bar(stat="identity") + ylab("Total Economic Damage to Crops") +
  xlab("Top 20 Events") + coord_flip()

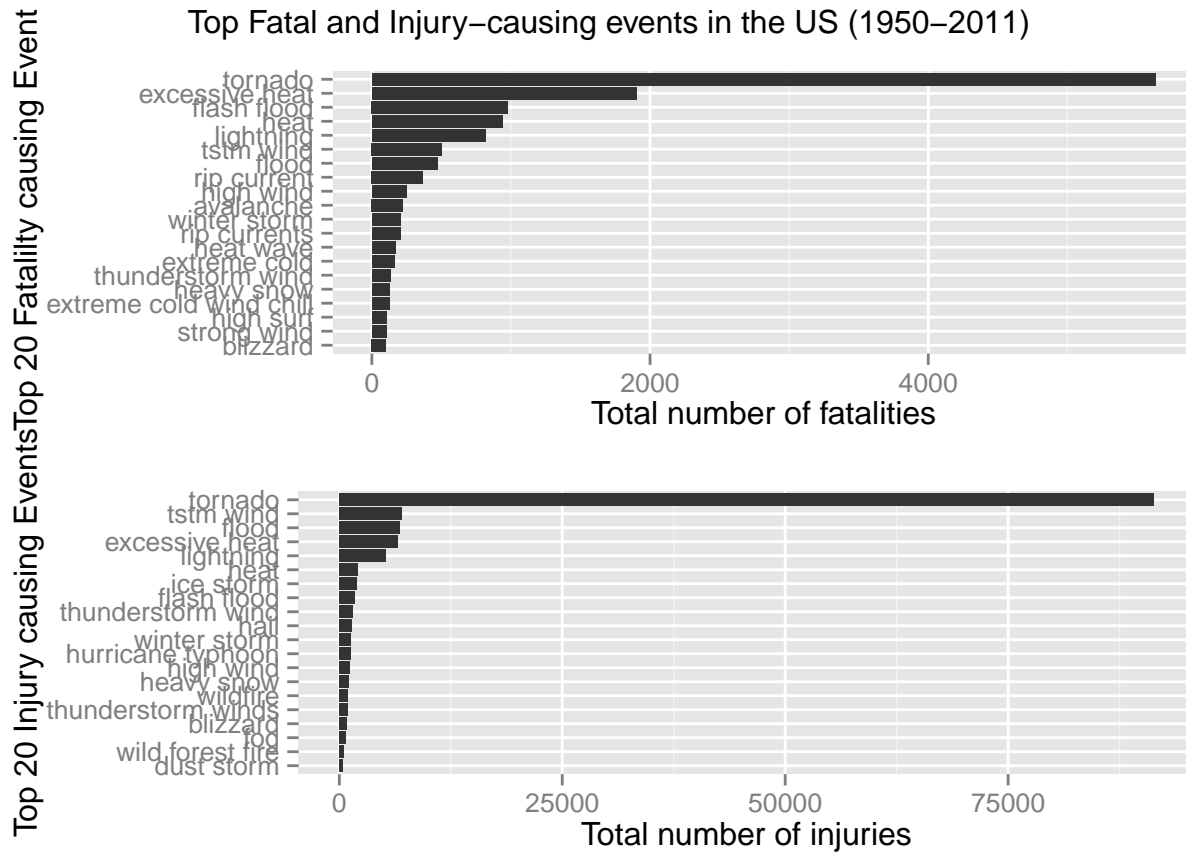
grid.arrange(g3, g4, main="Top events by their economic impact in the US (1950-2011)")

```

## Results

### Impact of Events on Population Health

From the following graph, it can be seen that: - Top 3 Events leading to max. fatalities are as follows: Tornado, Excessive Heat, Flash Flood - Top 3 Events leading to max. injuries are as follows: Tornado, Thunderstorm (TSTM) wind, Flood

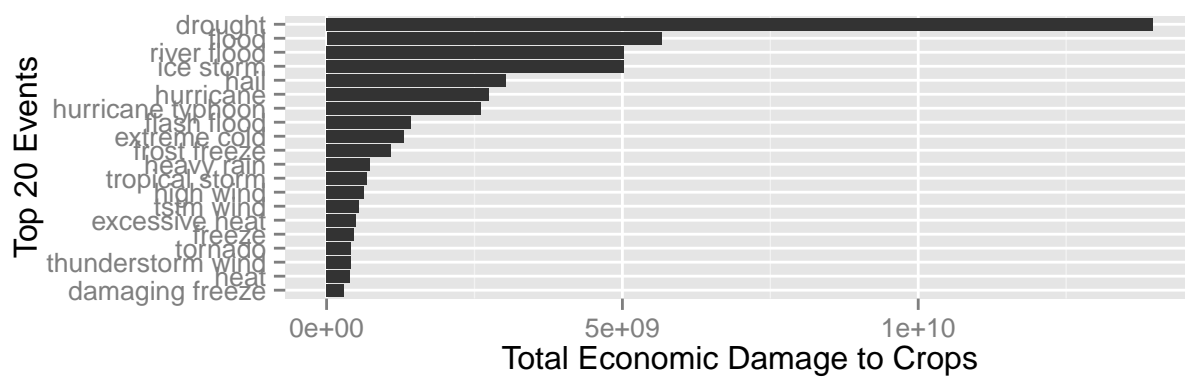
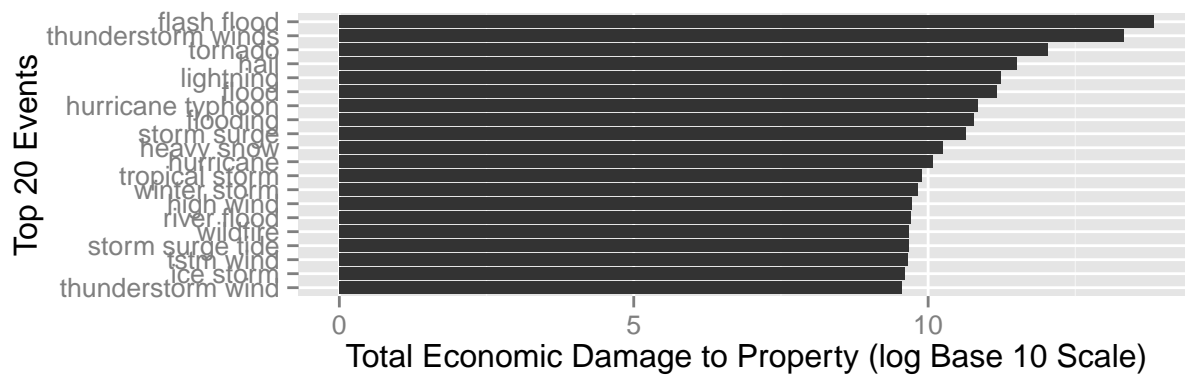


Hence, **Tornado** is by far the max. damaging event for population health.

### Economic Impact of Events on Property and Crops

From the following graph, it can be seen that: - Top 3 Events leading to max. Economic damage to property are: Flash flood, thunderstorm winds, tornado - Top 3 Events leading to max. Economic damage to crops are: Drought, Flood, Ice Storm

Top events by their economic impact in the US (1950–2011)



Hence, **Floods** of any kind cause the max. overall impact on Property and crops