

# Reproducible Research Peer Assignment 2

## Title

Summarizing the human and economic losses in the aftermath of severe weather events using NOAA Storms Events dataset

## Synopsis

Storms and other severe weather events can lead to both human (fatalities and injuries) and economic (crop and property damages) losses. Getting an estimate of the extent of damage caused by such events will help in ascertaining the key severe weather events. This can help in generating focus areas to alleviate the impact of such events in future [3 lines]

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. After the data cleaning process, we aggregate the number of fatalities, injuries, and property and crop damage as per event type. Using guidelines of National Safety Council, we assign an economic value to each fatality and injury, and combine this to the property and crop damages to arrive at a holistic economic loss for each event type. [4 lines]

We consider the top ten events which result in maximum extent of human and economic losses. We find that tornadoes have the maximum total number of casualties (both fatalities and injuries). However, tornadoes are ranked number three when we consider the total economic losses. [2 lines]

## Data Processing

The compressed dataset is downloaded into a temporary folder and uncompressed using the *bunzip2* function from **R.utils** package.

```
# load R.utils packages or install if not available
if(!require("R.utils")){
  install.packages("R.utils")
}
```

```
## Loading required package: R.utils
## Loading required package: R.oo
## Loading required package: R.methodsS3
## R.methodsS3 v1.6.1 (2014-01-04) successfully loaded. See ?R.methodsS3 for help.
## R.oo v1.18.0 (2014-02-22) successfully loaded. See ?R.oo for help.
##
## Attaching package: 'R.oo'
##
## The following objects are masked from 'package:methods':
##
##   getClasses, getMethods
##
## The following objects are masked from 'package:base':
##
##   attach, detach, gc, load, save
##
## R.utils v1.32.4 (2014-05-14) successfully loaded. See ?R.utils for help.
## Attaching package: 'R.utils'
##
## The following object is masked from 'package:utils':
##
##   timestamp
##
## The following objects are masked from 'package:base':
##
##   cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

```
library("R.utils")
```

```
# set the file url
fileurl <-
"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

# create a temporary directory
td = tempdir()

# create the placeholder file
tf = tempfile(tmpdir=td, fileext=".bz2")

# download into the placeholder file (curl method needed for Mac OS X)
download.file(fileurl, tf, method="curl")

# fpath is the full path to the extracted file
fname = "repdata_data_StormData.csv"
fpath = file.path(td, fname)

# unzip the file to the temporary directory
bunzip2(tf, destname=fpath, overwrite=TRUE)
```

Next, the dataset is loaded in the dataframe **df** and several data cleaning operations are performed as listed below. For example, there are several types of events which are represented by more than one name, such as *TSTM WIND* and *THUNDERSTORM WIND*.

```
# load the csv in data frame
df <- read.csv(fpath, as.is=TRUE)

# discard EVTYPE variables containing "Summary" term and store in df2
df2 <- df[!grepl("Summary", df$EVTYPE), ]

# convert all EVTYPE factors to upper case
df2$EVTYPE <- toupper(df2$EVTYPE)

# fix the issue of same EVTYPE having two names
df2[df2$EVTYPE == "TSTM WIND", ]$EVTYPE = "THUNDERSTORM WIND"
df2[df2$EVTYPE == "THUNDERSTORM WINDS", ]$EVTYPE = "THUNDERSTORM WIND"
df2[df2$EVTYPE == "RIVER FLOOD", ]$EVTYPE = "FLOOD"
df2[df2$EVTYPE == "HURRICANE/TYPHOON", ]$EVTYPE = "HURRICANE-TYPHOON"
df2[df2$EVTYPE == "HURRICANE", ]$EVTYPE = "HURRICANE-TYPHOON"

# convert EVTYPE from character to factor
df2$EVTYPE <- as.factor(df2$EVTYPE)
```

To calculate the economic costs, we need to assess the property and crop damage due to a severe weather event. In the data set, the amount and the units of the damage are present in different columns. The amounts in *PROPDMG* and *CROPDGMG* are scaled up using the units in *PROPDMGEXP* and *CROPDGMGEXP* respectively. We use the *recode* function from **car** package to perform the recoding of *PROPDMGEXP* and *CROPDGMGEXP*.

```
# load car packages or install if not available
if(!require("car")){
  install.packages("car")
}
```

```
## Loading required package: car
##
## Attaching package: 'car'
##
## The following object is masked from 'package:rgl':
##
##     identify3d
##
## The following object is masked from 'package:psych':
##
##     logit
```

```
library("car")
```

```
# use PROPDMGEXP to get property damage costs
df2$propdmgUnits <- recode(df2$PROPDMGEXP, " ''=0; '- '=0; '? '=0; '+ '=0;
'0'=0; '1'=10; '2'=100;

'3'=1000; '4'=10000; '5'=100000; '6'=1000000; '7'=10000000;
'8'=100000000; 'B'=1000000000; 'h'=100; 'H'=100;
'k'=1000;
'K'=1000; 'm'=1000000; 'M'=1000000",
as.factor.result = FALSE)
df2$PROPDMG <- df2$PROPDMG * df2$propdmgUnits

# use CROPDMGEXP to get crop damage costs
df2$cropdmgUnits <- recode(df2$CROPDMGEXP, " ''=0; '- '=0; '? '=0; '+ '=0;
'0'=0; '1'=10; '2'=100;

'3'=1000; '4'=10000; '5'=100000; '6'=1000000; '7'=10000000;
'8'=100000000; 'B'=1000000000; 'h'=100; 'H'=100;
'k'=1000;
'K'=1000; 'm'=1000000; 'M'=1000000",
as.factor.result = FALSE)

df2$CROPDMG <- df2$CROPDMG * df2$cropdmgUnits
```

## Data Analysis

The study aims to assess the impact of adverse weather events on population health and the economic lossess associated. Using *ddply* function from **plyr** package, the total fatalities, injuries and economic losses (property and crop damages) are summarized as per event type.

The National Safety Council in Estimating Cost of Unintentional Injuries gives the estimated economic cost of a fatality as \$1,410,000 and that of injury as \$78,900 [[http://www.nsc.org/news\\_resources/injury\\_and\\_death\\_statistics/Pages/EstimatingtheCostsofUnintentionalInjuries.aspx](http://www.nsc.org/news_resources/injury_and_death_statistics/Pages/EstimatingtheCostsofUnintentionalInjuries.aspx)]. We use these figures to calculate the holistic economical cost of an adverse weather event. To assist in analysis, the economical costs are scaled by 1 billion.

```
# load plyr packages or install if not available
if(!require("plyr")){
  install.packages("plyr")
}
```

```
## Loading required package: plyr
```

```
library("plyr")
```

```
# use ddply from plyr to summarize data for fatalities, injury and damage
grouped per event type
aggrByEvent = ddply(df2, .(EVTYPE), summarize,
                    totalFatalities = sum(FATALITIES),
                    totalInjuries = sum(INJURIES),
                    totalPropdmg = sum(PROPDMG),
                    totalCropdmg = sum(CROPDMG)
                    )

# calculate total economic costs by combining cost of fatality/injury and
crop/property
# damage due to an adverse weather event
costFatality <- 1410000
costInjury <- 78900
aggrByEvent$totalDamageCost <- (aggrByEvent$totalCropdmg +
aggrByEvent$totalPropdmg)/1e9
aggrByEvent$totalHumanCost <- (costFatality * aggrByEvent$totalFatalities +
costInjury * aggrByEvent$totalInjuries)/1e9
aggrByEvent$totalEconomicCost <- aggrByEvent$totalDamageCost +
aggrByEvent$totalHumanCost
```

## Results

The two research questions are as follows:

1. Across the United States, which types of events are most harmful with respect to population health?

## 2. Across the United States, which types of events have the greatest economic consequences?

The first two figures depict the total injuries and fatalities respectively which happened during the adverse weather events. The list of event types is sorted in descending order enabling us to find out the events which cause the most disruption in population health. Therefore, these two figures assist us in answering research question 1.

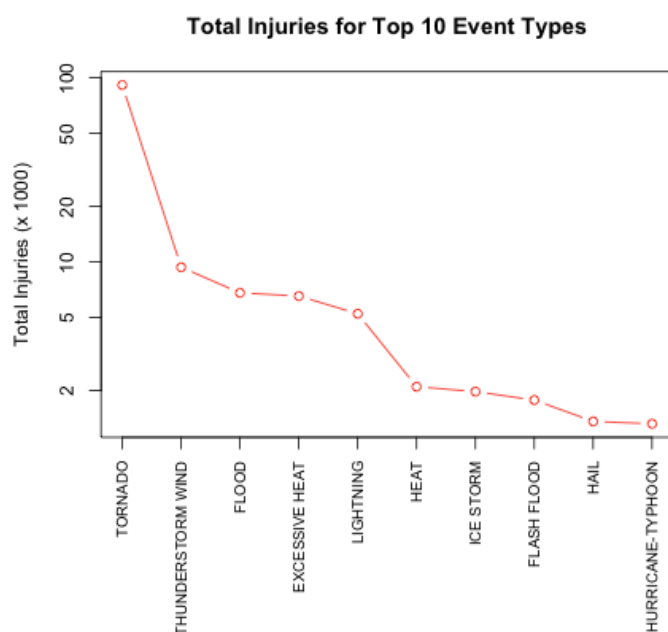
The last figure depicts the total economic costs of the adverse weather events. Again, the list of event types is sorted in descending order enabling us to find out the events which cause the most economical loss. Thus, this figure assists us in answering research question 2.

We limit the list of events to 10 for clarity of visualization and interpretation. Note that for each of the output parameters (injuries or fatalities or economic losses), the top 10 events might be different.

```
# limit to top 10 event types
n <- 10

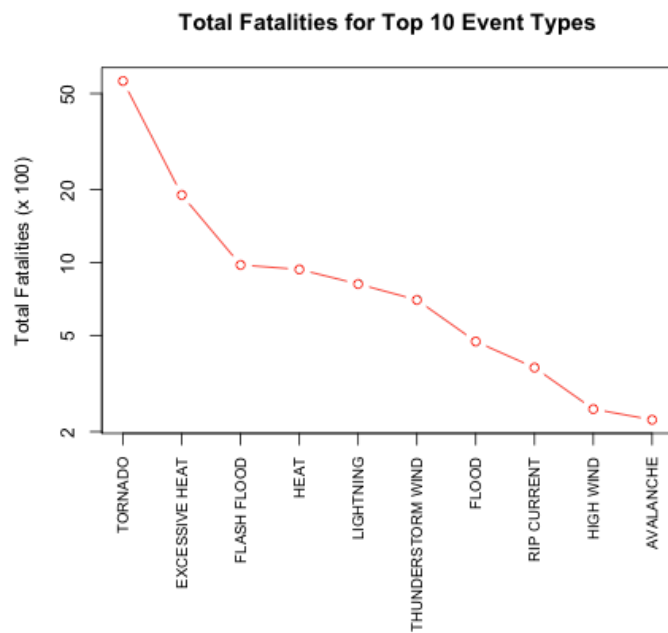
par(mfrow=c(1,1), mar=c(10,8,4,2))

# total injuries
plotdata <- aggrByEvent[order(aggrByEvent$totalInjuries, decreasing = T), ]
[1:n, ]
plot(plotdata$totalInjuries/1000, col = "red", type = "b", xaxt = "n", log =
'y', xlab = "",
main = "Total Injuries for Top 10 Event Types", ylab = "Total Injuries (x
1000)")
axis(1, labels = plotdata$EVTYPE, at = 1:length(plotdata$EVTYPE), las = 2,
cex.axis = 0.8)
```



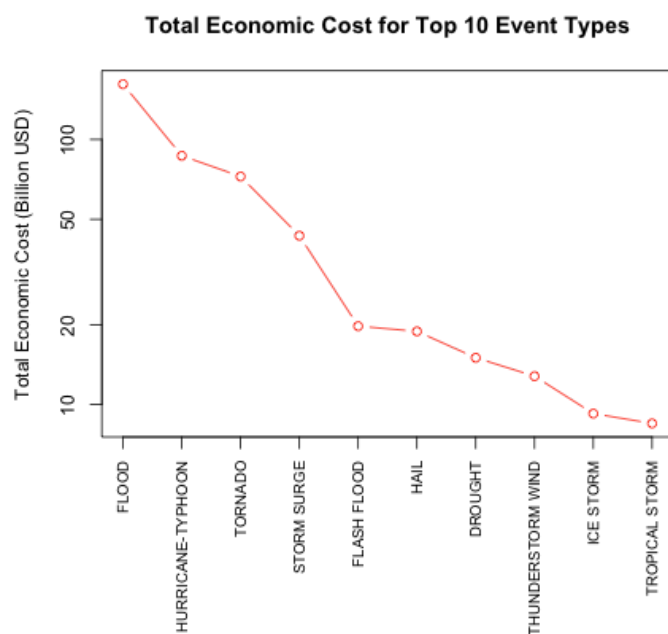
```
par(mfrow=c(1,1), mar=c(10,8,4,2))

# total fatalities
plotdata <- aggrByEvent[order(aggrByEvent$totalFatalities, decreasing = T), ]
[1:n, ]
plot(plotdata$totalFatalities/100, col = "red", type = "b", xaxt = "n", xlab =
"", log = 'y',
main = "Total Fatalities for Top 10 Event Types", ylab = "Total
Fatalities (x 100)")
axis(1, labels = plotdata$EVTYPE, at = 1:length(plotdata$EVTYPE), las = 2,
cex.axis = 0.8)
```



```
par(mfrow=c(1,1), mar=c(10,8,4,2))

# total economic losses
plotdata <- aggrByEvent[order(aggrByEvent$totalEconomicCost, decreasing = T),
][1:n, ]
plot(plotdata$totalEconomicCost, col = "red", type = "b", log = 'y',
      xaxt = "n", xlab = "", main = "Total Economic Cost for Top 10 Event
Types",
      ylab = "Total Economic Cost (Billion USD)")
axis(1, labels = plotdata$EVTYPE, at = 1:length(plotdata$EVTYPE), las = 2,
      cex.axis = 0.8)
```



```
# get the first rows of data to see the numbers
head(plotdata)
```

##	EVTYPE	totalFatalities	totalInjuries	totalPropdmg
## 153	FLOOD	472	6791	1.498e+11
## 370	HURRICANE-TYPHOON	125	1321	8.117e+10
## 691	TORNADO	5633	91346	5.695e+10
## 596	STORM SURGE	13	38	4.332e+10
## 137	FLASH FLOOD	978	1777	1.682e+10
## 211	HAIL	15	1361	1.574e+10
##	totalCropdmg	totalDamageCost	totalHumanCost	totalEconomicCost
## 153	1.069e+10	160.47	1.20133	161.67
## 370	5.350e+09	86.52	0.28048	86.80
## 691	4.150e+08	57.36	15.14973	72.51
## 596	5.000e+03	43.32	0.02133	43.34
## 137	1.421e+09	18.24	1.51919	19.76
## 211	3.026e+09	18.76	0.12853	18.89

We find that tornadoes have the maximum total number of casualties (both fatalities and injuries). However, tornadoes are ranked number three when we consider the total economic losses. The first two places are taken by flood and hurricane-typhoon. A possible reason can be somewhat longer duration of flood as compared to tornado. Due to the longer duration of the event, the extent of damage is much more leading to higher economic impact.

It is also interesting to note that there are more injuries during flood than flash flood (flood without prior warning) whereas fatalities are more for flash flood as compared to flood (expected since there is less time to react in the event of flash flood). Excessive heat also takes a toll on human population. It's ranked number four for injuries and number two for fatalities.