

# An analysis of the fatalities and economic impacts caused from different events

## Synopsis:

This report is about a brief research about the impacts to people's safety and economy. Based on the dataset downloaded from **NOAA Storm Database**(URL: <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>), 902297 obs. of 37 variables ), we took a glimpse at the cases corresponding to each event types (**EVTYPE**). After we made two plots, we drew to the conclusion that most fatal & economical impact is caused by tornados. According to the first plot, most fatal impact of harmful events come from tornados, which causes approx. 100,000 people died in the last 60 years. According to the second plot, most economic impact come from floodess, about \$15 billions dmgl for 60 years.

## Prerequisites

Load the packages necessary for analysis

We will use **reshape**, **ggplot2**, **Hmisc**, **car** packages to accomplish our data analysis. And use **knitr** package to generate this report.

```
setwd("E:/R/repdata2")
library(knitr)
library(reshape)
library(ggplot2)
library(Hmisc)
library(car)
```

## Data Processing

### Load Data

I tried loading data from the original dataset as follows

```
setwd("E:/Rdir/repdata2")
storm_csv <- read.csv("repdata_data_StormData.csv.bz2", header =
TRUE, nrows=902297, comment.char="")
```

```
## warning: EOF within quoted string
```

```
dim(storm_csv) ##not 902297 obs.
```

```
## [1] 692288      37
```

But as you can see, I failed loading the data from the original csv. Struggled for 4 hours to solve the problem, I finally find that the error is because of the settings of my operating system, which seems a difficult task to deal with.

```
sessionInfo()
```

```
## R version 3.1.0 (2014-04-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.936
## [2] LC_CTYPE=Chinese (Simplified)_China.936
## [3] LC_MONETARY=Chinese (Simplified)_China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.936
```

I asked for help in the discussion forum, somebody in the forum offers me kindly help by sharing me an .RData file containing the raw dataset after we find no ways to deal with the problem that prevents me to start the mini-project. The original thread contains the .RData file is [[https://class.coursera.org/repdata-003/forum/thread?thread\\_id=31](https://class.coursera.org/repdata-003/forum/thread?thread_id=31)] The download link: [<https://dl.dropboxusercontent.com/u/31729865/mydata3.Rda>] I modified this .Rdata file in the console, changed the name of the dataset into “storm” (from “mydata” )

I certify that the .RData file only contains the raw data set.

So the process of loading the data for me is like that:

```
load("./storm.Rdata")
dim(storm)
```

```
## [1] 902297      37
```

## Data Process

STEP1:Turn the event type (EVTYPE) into upper case in order to lessen the possibility of error when categorizing the cases.

```
storm$EVTYPE <- toupper(storm$EVTYPE)
```

STEP2:Sum the fatalities and injuries (Safety Issues) by different event types, then transform it into a shape that's easy to process .

```

dmg1 <- aggregate(cbind(FATALITIES,INJURIES) ~ EVTYPE , storm,
sum)
dmg2 <- melt(head(dmg1[order(-dmg1$FATALITIES, -dmg1$INJURIES),
], 10))

```

```
## Using EVTYPE as id variables
```

STEP3: Multiply the damage of properties by the magnitude.

```

storm$PROPDGM <- storm$PROPDGM *
as.numeric(Recode(storm$PROPDGMEXP, "'0'=1;'1'=10;'2'=100;
'3'=1000;'4'=10000;'5'=100000;'6'=1000000;'7'=10000000;
'8'=100000000;'B'=1000000000;'h'=100;'H'=100;'K'=1000;
'm'=1000000;'M'=1000000; '-'=0; '?'=0;
+'=0",as.factor.result=FALSE))
storm$CROPDMG <- storm$CROPDMG *
as.numeric(Recode(storm$CROPDMGEXP, "'0'=1;'2'=100;
'B'=1000000000;'k'=1000;'K'=1000;'m'=1000000;'M'=1000000;
''=0;'?'=0",as.factor.result=FALSE))

```

STEP4: Sum the economic damages by different event types, then transform it into a shape that's easy to process.

```

eco1 <- aggregate(cbind(PROPDGM,CROPDMG) ~ EVTYPE , storm, sum)
eco2 <- melt(head(eco1[order(-eco1$PROPDGM, -eco1$CROPDMG), ],
10))

```

```
## Using EVTYPE as id variables
```

## Results

### Fatal Impacts

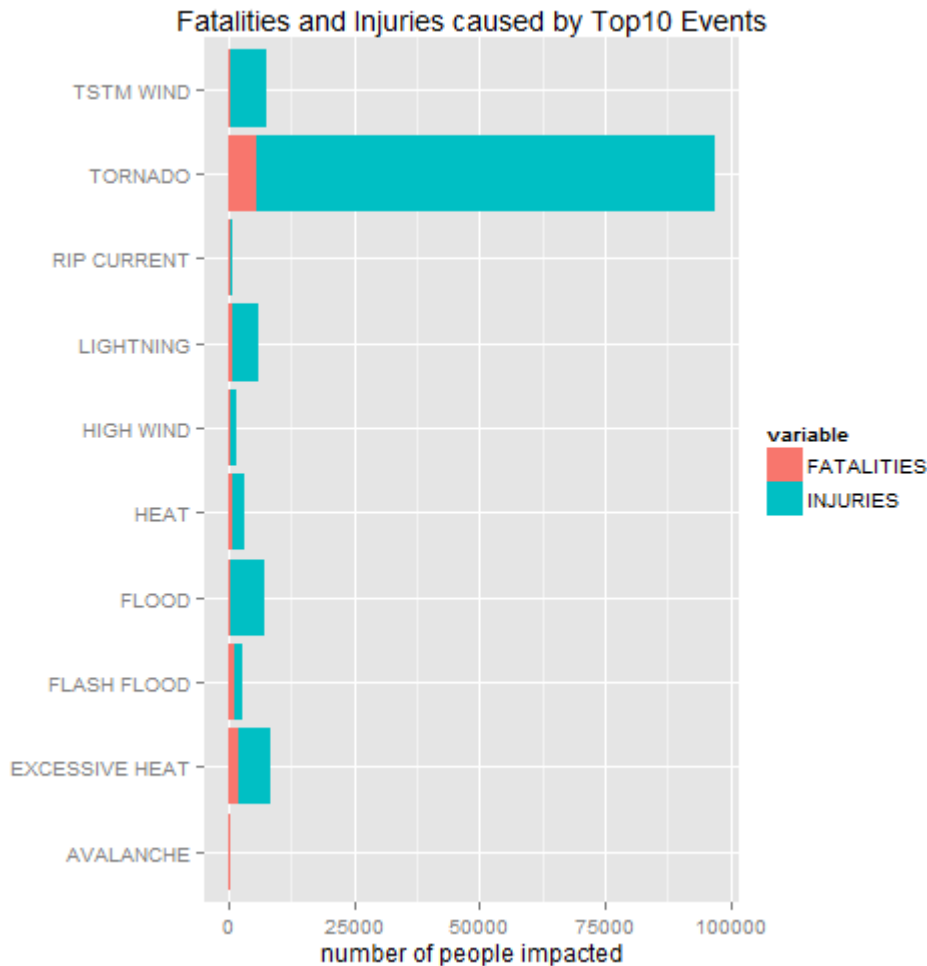
Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

We made a plot indicates the casualties and injuries by event types, using ggplot2 package. We draw to a conclusion that **TORNADO CONTRIBUTES MOST** in the fatalities and injuries.

```

ggplot(dmg2, aes(x=EVTYPE, y=value, fill=variable),color =
variable) + geom_bar(stat = "identity") +
  ggtitle("Fatalities and Injuries caused by Top10
Events") + labs(x = "", y="number of people impacted") +
coord_flip()

```



## Economical Impacts

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

We made a plot indicates the economic impacts in dollars by event types, using ggplot2 package. We draw to a conclusion that **FLOOD CONTRIBUTES MOST** in the economic impacts.

```
ggplot(eco2, aes(x = EVTYPE, y = value, fill = variable), color = variable) + geom_bar(stat = "identity") + ggtitle("Economic impacts caused by Top10 Events") + labs(x = "", y = "damages in dollars") + coord_flip()
```

