Storm Data Analysis: Top Health and Economic Risks

Tibor

23/08/2015

Synopsis

In the current analysis, I will show the most devastating types of weather events, that is, those that cause most damage in human life, property and crops. To do so, I will concentrate on four data types: 1. the number of fatalities, 2. the number of injuries, 3. the damage in property, and 4. the damage in crops the various event types cause. There will be no attempt to summarise the figures of the different types, only some verbal analysis will be performed to point out broader categories: air movement, percipitation, and surface water movements.

Data Processing

Data processing has the following major steps:

1. Downloading the data
2. Loading the data into a variable
3. Getting the data columns that will be used in analysis
4. Getting the top five causes from each type
5. Displaying the data

1-3. Getting the Data Used for Analysis

The following two functions will be used for this purpose:

- **get.the.data**: to download and load the data
- **get.processed.data**: to extract the data columns used for analysis

```
get.the.data <- function()
{
    fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    datapath <- file.path("./data/StormData.csv.bz2")
    data.filepath = "StormData.csv"

    if (!dir.exists(dirname(datapath)))
        dir.create(dirname(datapath))

    if (!file.exists(datapath))
        download.file(fileUrl,destfile=datapath,method="curl")
    else
        print(paste("File downloaded previously, using that: ",datapath))

    my.data <- read.csv(bzfile(datapath))
```

```
    return(my.data)
}

get.processed.data <- function()
    {
        saved.pruned.dat <- file.path("saved.pruned.dat")
        if(file.exists(saved.pruned.dat))
            {
                print(paste("Working from previously saved data:",saved.pruned.dat))
                load(saved.pruned.dat)
            }
        else
            {
                weather.data <- get.the.data()
                pruned.data <- data.frame(weather.data$EVTYPE,
                                          weather.data$FATALITIES,
                                          weather.data$INJURIES,
                                          weather.data$PROPDMG,
                                          weather.data$CROPDMG)
                names(pruned.data) = c("evtype","fatal","injury","prop.dmg","crop.dmg")
                save(pruned.data,file=saved.pruned.dat)
            }
        return(pruned.data)
    }
```

Executing the above functions, creates a variable called **the.data**:

```
the.data <- get.processed.data()
```

4. Getting the Top Five Causes

The top five causes are which cause the most fatalities, injuries, property damage, or crops damage.

The function used for this purpose:

```
get.top.causes <- function(the.values,the.type,the.num)
    {
        the.sums <- tapply(the.values,the.type,sum)
        top.sums <- head(sort(the.sums,decreasing=T,na.last=T),n=the.num)
        names(top.sums) = tolower(names(top.sums))
        return(top.sums)
    }
```

Four array variables will be used to store these data:

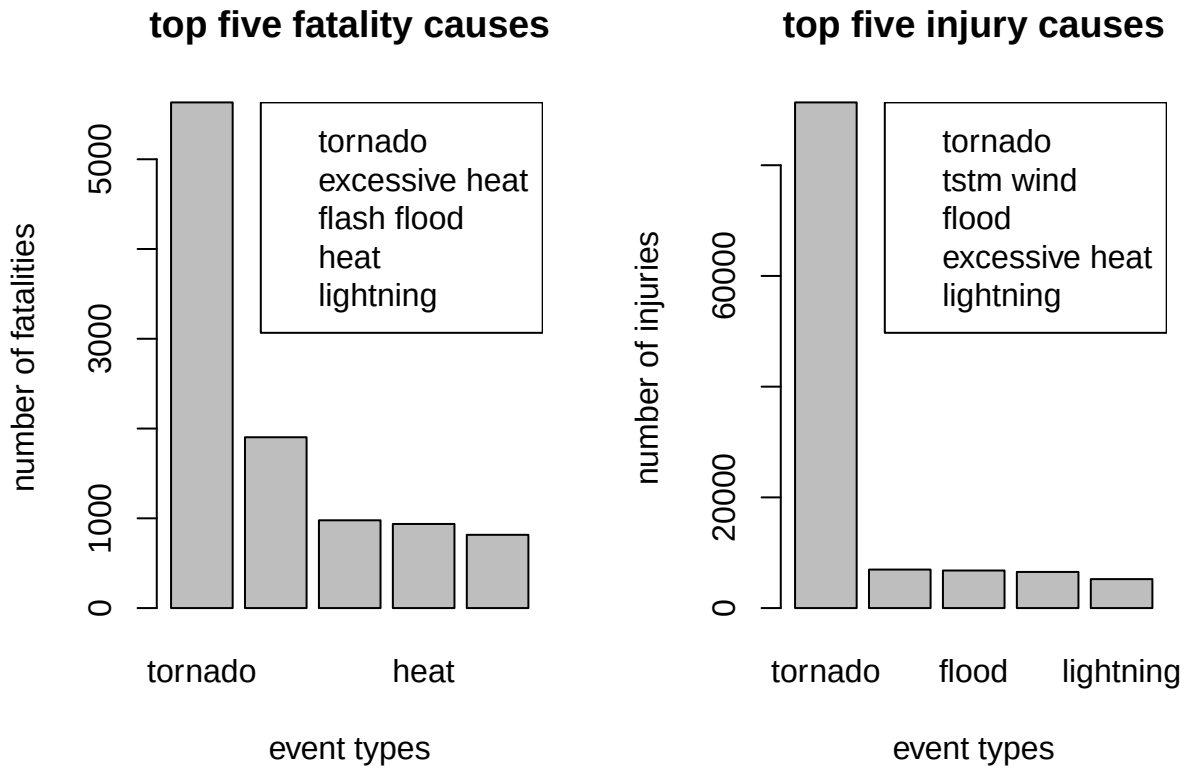
```
top.fatality.causes=get.top.causes(the.data$fatal,the.data$evtype,5)
top.injury.causes=get.top.causes(the.data$injury,the.data$evtype,5)
top.prop.dmg=get.top.causes(the.data$prop.dmg,the.data$evtype,5)
top.crop.dmg=get.top.causes(the.data$crop.dmg,the.data$evtype,5)
```

5. Displaying the Data

Harm to Health

The first figure will help answer which types of events are most harmful with respect to population health.

```
par(mfrow=c(1,2))
barplot(top.fatality.causes,ylab="number of fatalities",
        xlab="event types",main="top five fatality causes")
legend("topright", legend = names(top.fatality.causes))
barplot(top.injury.causes,ylab="number of injuries",
        xlab="event types",main="top five injury causes")
legend("topright", legend = names(top.injury.causes))
```



The data in tabular form:

```
library(pander)
pander(top.fatality.causes,style="markdown")
```

tornado	excessive heat	flash flood	heat	lightning
5633	1903	978	937	816

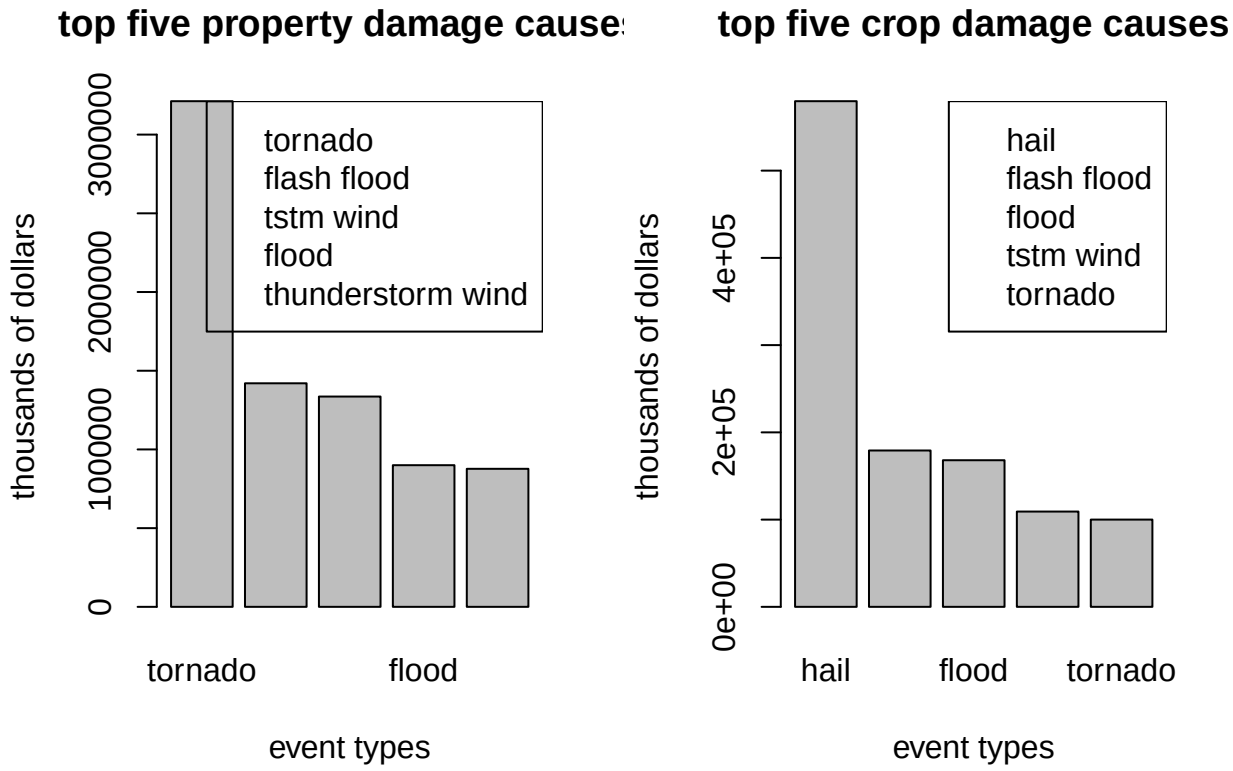
```
pander(top.injury.causes,style="markdown")
```

tornado	tstm wind	flood	excessive heat	lightning
91346	6957	6789	6525	5230

Greatest Economic Consequences

The second figure will help answer which types of events cause the most damage to property and crops.

```
par(mfrow=c(1,2))
barplot(top.prop.dmg,ylab="thousands of dollars",
        xlab="event types",main="top five property damage causes")
legend("topright", legend = names(top.prop.dmg))
barplot(top.crop.dmg,ylab="thousands of dollars",
        xlab="event types",main="top five crop damage causes")
legend("topright", legend = names(top.crop.dmg))
```



The data in tabular form:

```
pander(top.prop.dmg,style="markdown")
```

tornado	flash flood	tstm wind	flood	thunderstorm wind
3212258	1420125	1335966	899938	876844

```
pander(top.crop.dmg,style="markdown")
```

hail	flash flood	flood	tstm wind	tornado
579596	179200	168038	109203	1e+05

Results

Harm to Health

It can be seen that the most important damaging factor for both fatalities and injuries is tornadoes. When it comes to fatalities, we cannot ignore the effect of head-related deaths either: *excessive heat* and *heat* taken together, causes half as many deaths as tornadoes. Apart from winds and heat, the other key players seem to be floods (in the middle range) and lightning.

Greatest Economic Consequences

As for economic effects, there seems to be a clear distinction between property damage and crops damage. For property damage, the two major players are wind events (tornato, tsim wind, and thunderstorm wind) and flood events (flash floods and floods). However, the greatest enemy of the crops seems to be *hail*, and only after that come the flood events (flash floods and floods) and the wind events (tsim wind and tornado).