Cost of weather events : Injuries, fatalities and economic loss

Peter Ho

August 22, 2015

1. Sypnosis

This is the HTML report produced as a submission for Assignment 2 of Reproducible Report. We are given the task of analysing the data acquired from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. The goal of the analysis is to present the severity of weather events, in terms of the afflicted victims' well-being, as well as the economic cost of weather events. Tornados are the most hazardous weather event, in terms of injuries and deaths. Floods causes the most property damage, while droughts are most hazardous to economic loss for crops. In terms of sheer economic damage to both property and crops, floods are the most damaging weather event.

2. Data Processing

In this section, we described how the required data set is read using R code. The process begins by first loading the necessary libraries.

```
library(knitr)
opts_chunk$set(echo = TRUE, results = 'hold')#load knitr library
library(data.table)#load data table library
library(ggplot2) # we shall use ggplot2 for plotting figures
library(zoo)#for filling in missing values
```

```
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Subsequently, the required data set is read using the *read.scv()* function, which can even read compressed data, such as the data set that we are using.

```
file_location <- "C:/Users/PeterYvonne/Documents/ReprodAss2/ReprodAss2/repdata-data-StormData.csv.bz2"
data <- read.csv(file_location)
save(data, file="oridata.saved")#toprevent long loading of original data should there be a need to roll
data$EVTYPE=toupper(data$EVTYPE) #set column headers to uppercase
head(data)#show first several rows of a dataset
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
```

```
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                    14.0   100 3   0          0
## 2         NA         0                     2.0   150 2   0          0
## 3         NA         0                     0.1   123 2   0          0
## 4         NA         0                     0.0   100 2   0          0
## 5         NA         0                     0.0   150 2   0          0
## 6         NA         0                     1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

3. Results

3.1 Most injuries and fatalities

In this section we will take a look at injuries and fatalities caused by weather events. The columns of interest to us at this point would be the EVTYTPE (Event type), INJURIES and FATALITIES.

We start off with injuries caused by weather event. Here, we sum the total number of injuries caused by a particular weather event, and we subsequently list the top-10 weather events that causes the most number of *injuries*. The code snippet below shows how this is done:

```
injury <- aggregate(INJURIES ~ EVTYPE, data = data, sum)
actualinjury <- injury[injury$INJURIES > 0, ]#events which caused at least one injury
orderedinjury<-actualinjury[order(actualinjury$INJURIES, decreasing = TRUE), ]
head(orderedinjury)
```

```
##             EVTYPE INJURIES
## 758        TORNADO    91346
## 779      TSTM WIND     6957
## 154          FLOOD     6789
## 116 EXCESSIVE HEAT     6525
## 418      LIGHTNING     5230
## 243           HEAT     2100
```

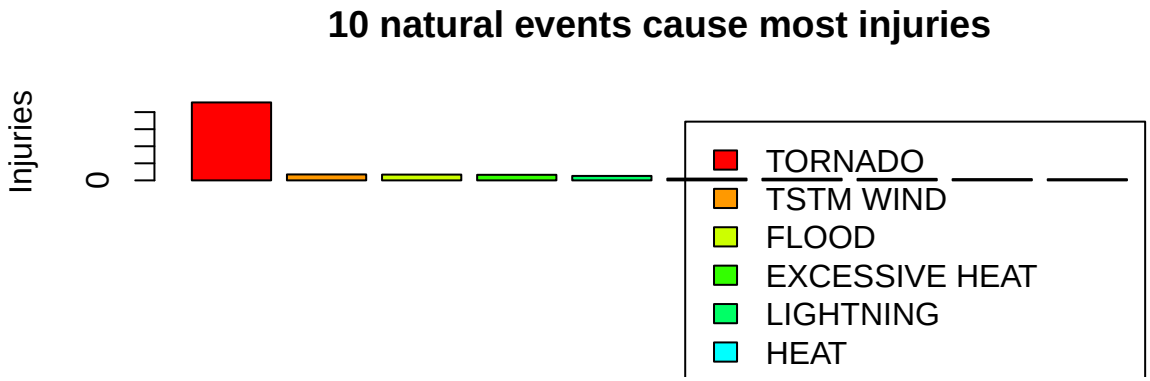
Subsequently, we move on to more serious effects of weather events, namely deaths caused by weather events. We give this the same treatment as the injury data, namely aggregating by summing up total number of deaths and ordering the list so that the top-10 weather events which causes *fatalities* are shown using the following code snippet.

```
fatality <- aggregate(FATALITIES ~ EVTYPE, data = data, sum)
actualfatality <- fatality[fatality$FATALITIES > 0, ]#events which caused at least one injury
orderedfatality <-actualfatality[order(actualfatality$FATALITIES, decreasing = TRUE), ]
head(orderedfatality)
```

```
##             EVTYPE FATALITIES
## 758        TORNADO       5633
## 116 EXCESSIVE HEAT       1903
## 138    FLASH FLOOD        978
## 243           HEAT        937
## 418      LIGHTNING        816
## 779      TSTM WIND        504
```

To make a comparison visually, a 2-panel graph plot is shown using the next block of code. This graph plot shows visually which weather events causes the most injuries and fatalities. From the graph plot, it is clear that tornados are the single most dangerous weather event that one can be exposed to.

```
par(mfrow = c(2, 1))
barplot(orderedinjury[1:10, 2], col = rainbow(10), legend.text = orderedinjury[1:10,
                                                                               1], ylab = "Injuries", main =
barplot(orderedfatality[1:10, 2], col = rainbow(10), legend.text = orderedfatality[1:10,
                                                                                 1], ylab = "Fatalities", mai
```



3.2 Most economic loss

We would next like to examine which weather events causes the most economic loss. Before we can aggregate the data, we need to do some preprocessing.

A quick inspection of the supporting document for the data set reveals that data found in the PROPDMG and CROPDMG columns has different scales, and the scales are shown in both the PROPDMGEXP and CROPDMGEXP. As such, we need to transform the value of the PROPDMG AND CROPDMGS into unit values of 1 USD.

The following code snippets accomplishes this purpose.

```
data[data$PROPDMGEXP == "K", ]$PROPDMG <- data[data$PROPDMGEXP == "K", ]$PROPDMG *
  1000
data[data$PROPDMGEXP == "M", ]$PROPDMG <- data[data$PROPDMGEXP == "M", ]$PROPDMG *
  1e+06
data[data$PROPDMGEXP == "m", ]$PROPDMG <- data[data$PROPDMGEXP == "m", ]$PROPDMG *
  1e+06
data[data$PROPDMGEXP == "B", ]$PROPDMG <- data[data$PROPDMGEXP == "B", ]$PROPDMG *
  1e+09
data[data$CROPDMGEXP == "K", ]$CROPDMG <- data[data$CROPDMGEXP == "K", ]$CROPDMG *
  1000
data[data$CROPDMGEXP == "k", ]$CROPDMG <- data[data$CROPDMGEXP == "k", ]$CROPDMG *
  1000
data[data$CROPDMGEXP == "M", ]$CROPDMG <- data[data$CROPDMGEXP == "M", ]$CROPDMG *
  1e+06
```

```
data[data$CROPDMGEXP == "m", ]$CROPDMG <- data[data$CROPDMGEXP == "m", ]$CROPDMG *
  1e+06
data[data$CROPDMGEXP == "B", ]$CROPDMG <- data[data$CROPDMGEXP == "B", ]$CROPDMG *
  1e+09
```

Once the amount of damages is standardized, we then perform the same aggregation as was done in Section 3.1. This is similar in effect as a SQL GROUPBY and ORDERBY operation and can be seen in the following chunk of code. We first focus on *property* damage.

```
damage <- aggregate(PROPDMG ~ EVTYPE, data = data, sum)
actualdamage <- damage[damage$PROPDMG > 0, ]#events which caused at least one USD worth of damage
ordereddamage<-actualdamage[order(actualdamage$PROPDMG, decreasing = TRUE), ]
head(ordereddamage)
```

```
##                EVTYPE      PROPDMG
## 154             FLOOD 144657709807
## 372 HURRICANE/TYPHOON  69305840000
## 758           TORNADO  56937160779
## 599       STORM SURGE  43323536000
## 138       FLASH FLOOD  16140812067
## 212              HAIL  15732267048
```

Floods causes the most damage by far, followed by hurricanes and typhons.

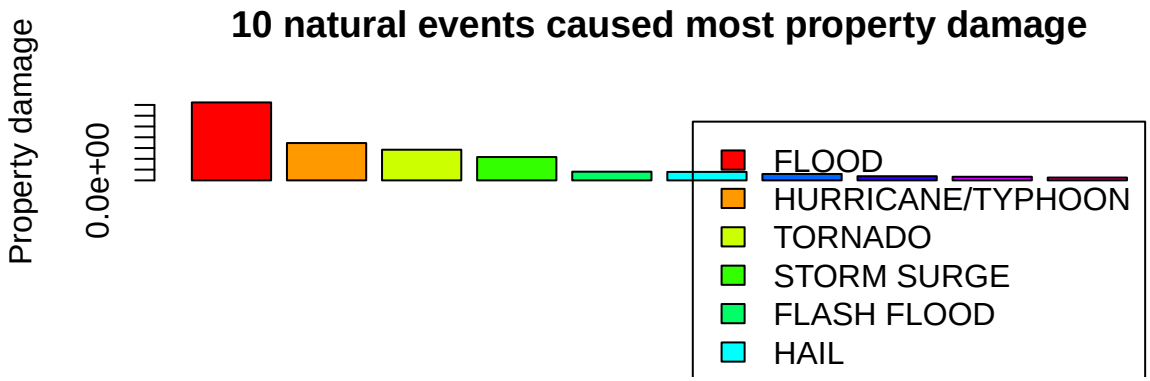
We then move our focus to economic damage caused by weather events to crops. The same set of procedures is applied to *crops*, using the following code.

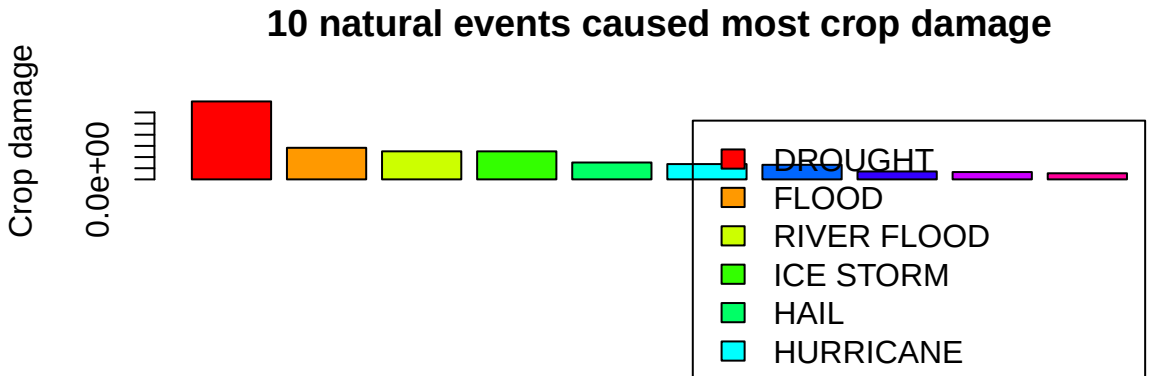
```
damagecrop <- aggregate(CROPDMG ~ EVTYPE, data = data, sum)
actualdamagecrop <- damagecrop[damagecrop$CROPDMG > 0, ]#events which caused at least one USD worth of
ordereddamagecrop<-actualdamagecrop[order(actualdamagecrop$CROPDMG, decreasing = TRUE), ]
head(ordereddamagecrop)
```

```
##          EVTYPE     CROPDMG
## 84      DROUGHT 13972566000
## 154       FLOOD  5661968450
## 529 RIVER FLOOD  5029459000
## 387   ICE STORM  5022113500
## 212        HAIL  3025954473
## 363   HURRICANE  2741910000
```

Another plot shows the a comparison of damage caused to property and crops, according to different weather events.

```
par(mfrow = c(2, 1))
barplot(ordereddamage[1:10, 2], col = rainbow(10), legend.text = ordereddamage[1:10,
                                                                                1], ylab = "Property damage"
barplot(ordereddamagecrop[1:10, 2], col = rainbow(10), legend.text = ordereddamagecrop[1:10,
                                                                                  1], ylab = "Crop damage", m
```

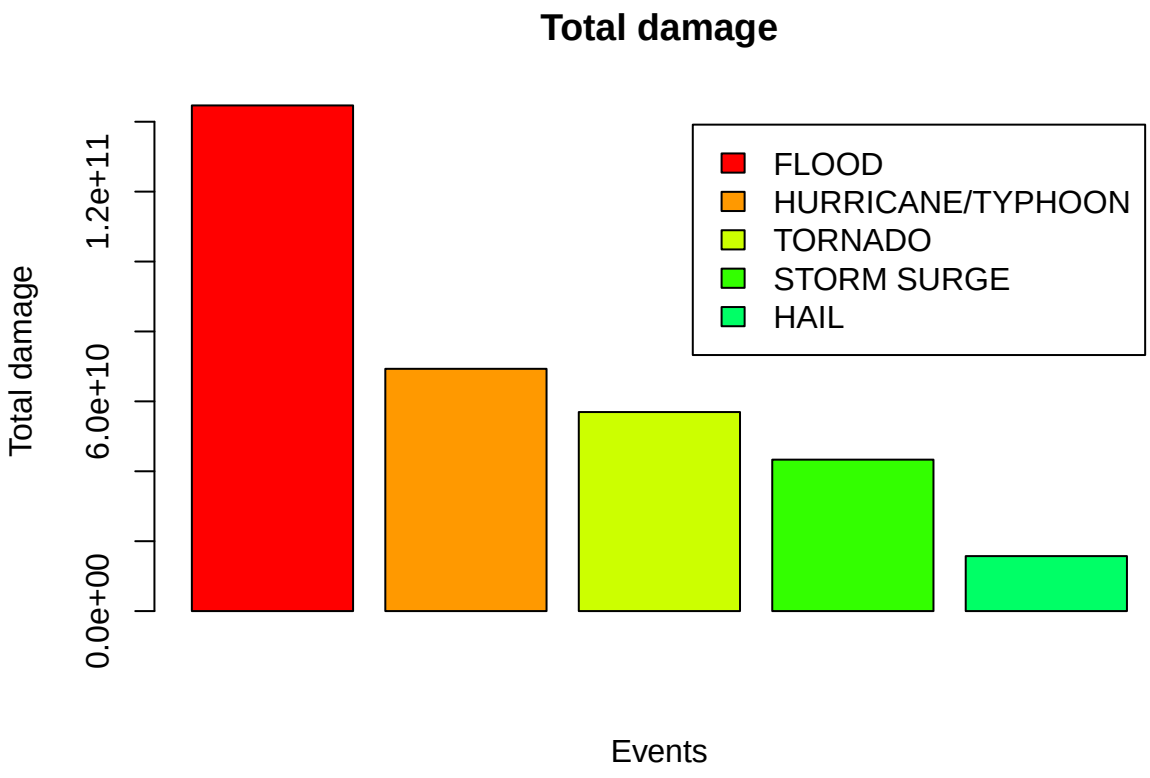




Drought causes the most damage to crops, follwed by flood. A simple deduction can be made that in terms of economic damage, floods have the most effect on property and crop damages.

This deduction is proven by our final analysis. We merge the two data objects containing the ordered list of property and crop damage, and add the total amount of damage. The outcome is shown on a plot using the following code snippet.

```
totaldmg <- merge(ordereddamage, ordereddamagecrop, by = "EVTYPE")
totaldmg$total = totaldmg$PROPDMG + totaldmg$CROPDMG
totaldmgorder <- totaldmg[order(totaldmg$total, decreasing = TRUE), ]
totaldmgorder[1:5, ]
barplot(totaldmgorder[1:5, 2 ], col = rainbow(10), legend.text = totaldmgorder[1:5, 1], ylab = "Total da
```




```
##                EVTYPE      PROPDMG    CROPDMG        total
## 19              FLOOD 144657709807 5661968450 150319678257
## 53 HURRICANE/TYPHOON  69305840000 2607872800  71913712800
## 85            TORNADO  56937160779  414953270  57352114049
## 68        STORM SURGE  43323536000       5000  43323541000
## 31               HAIL  15732267048 3025954473  18758221521
```