Analysis of U.S. National Oceanic and Atmospheric Administration's storm database

Reproducible Research: Peer Assessment 2

by Tomislav Vincelj

Abstract/Synopsis

Goal

To perform analysis on the U.S. National Oceanic and Atmospheric Administration storm database in order to find the answer to the following two questions: 1. Across the United states, which type of events are most harmufl with respect to population health? 2. Across the United States, which types of events have the greatest economic consequences?

Data Processing

I have downloaded the compressed database from here (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) and placed it into my working folder, and then uncompressed and loaded the data into an R data frame called *stormData*:

```
stormData<-read.csv("./repdata-data-StormData.csv.bz2")
```

In order to find out which types of events are most harmful with respect to human health as well as with respect to economic consequences, some data praparation is necessary.

I have summed the total number of injuries and fatalities across the US, and ordered them descendingly according to the injury/fatality count.

```
totalInjuriesPerEvent<-aggregate(INJURIES ~ EVTYPE,data=stormData,FUN=sum)
totalFatalitiesPerEvent<-aggregate(FATALITIES ~ EVTYPE,data=stormData,FUN=sum)

orderedInjuries<-totalInjuriesPerEvent[order(totalInjuriesPerEvent$INJURIES,decreasing=TRUE)
,]
orderedFatalities<-totalFatalitiesPerEvent[order(totalFatalitiesPerEvent$FATALITIES,decreasi
ng=TRUE),]
```

The property and crop damages required a different approach. Since the damage data is in two columns - one stating the damage in US Dollars, the other stating the exponent (K being 1000, M=1000K, B=1000M), true damages need to be calculated beforehand.

We can see that there are other exponents in both exponent columns.

```
table(stormData$PROPDMGEXP)
```

```
##
##             -       ?       +       0       1       2       3       4       5
## 465934      1       8       5     216      25      13       4       4      28
##      6      7       8       B       h       H       K       m       M
##      4      5       1      40       1       6  424665       7   11330
```

```
table(stormData$CROPDMGEXP)
```

```
##
##             ?       0       2       B       k       K       m       M
## 618413      7      19       1       9      21  281832       1    1994
```

However since these three (K, M, B) are most common, I will be using them only, since I have found an explanation for them in the database documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf), but no explanation for the other exponents.

So first we will aggregate the property and crop damage data by event and by exponent.

```
totalPropertyCost<-aggregate(PROPDMG ~ EVTYPE+PROPDMGEXP,data=stormData,sum)
totalCropCost<-aggregate(CROPDMG ~ EVTYPE+CROPDMGEXP,data=stormData,sum)
```

Then clean the data by keeping (subsetting to) only the K,M,B exponents.

```
cleanPropertyCost<-totalPropertyCost[totalPropertyCost$PROPDMGEXP %in% c("B","M","K"),]
cleanCropCost<-totalCropCost[totalCropCost$CROPDMGEXP %in% c("B","M","K"),]
```

Now I will calculate the true cost of each event by multiplying the property and crop damage with the appropriate exponent.

```r
#PROPERTY DAMAGE
for(i in 1:nrow(cleanPropertyCost)){
      if (cleanPropertyCost$PROPDMGEXP[i]=="K") {
      cleanPropertyCost$truePropDmg[i] <- cleanPropertyCost$PROPDMG[i] * 1000
      }
      else if (cleanPropertyCost$PROPDMGEXP[i]=="M") {
      cleanPropertyCost$truePropDmg[i] <- cleanPropertyCost$PROPDMG[i] * 1000000
      }
      else if (cleanPropertyCost$PROPDMGEXP[i]=="B") {
      cleanPropertyCost$truePropDmg[i] <- cleanPropertyCost$PROPDMG[i] * 1000000000
      }

}

#CROP DAMAGE
for(i in 1:nrow(cleanCropCost)){
      if (cleanCropCost$CROPDMGEXP[i]=="K") {
      cleanCropCost$trueCropDmg[i] <- cleanCropCost$CROPDMG[i] * 1000
      }
      else if (cleanCropCost$CROPDMGEXP[i]=="M") {
      cleanCropCost$trueCropDmg[i] <- cleanCropCost$CROPDMG[i] * 1000000
      }
      else if (cleanCropCost$CROPDMGEXP[i]=="B") {
      cleanCropCost$trueCropDmg[i] <- cleanCropCost$CROPDMG[i] * 1000000000
      }
}
```

Finally, I can now sum (aggregate) and order the atmospheric events by their true cost.

```r
truePropertyCost<-aggregate(truePropDmg ~ EVTYPE, data=cleanPropertyCost,sum)
trueCropCost<-aggregate(trueCropDmg ~ EVTYPE, data=cleanCropCost,sum)

orderedPropertyCost<-truePropertyCost[order(truePropertyCost$truePropDmg,decreasing=TRUE),]
orderedCropCost<-trueCropCost[order(trueCropCost$trueCropDmg,decreasing=TRUE),]
```

Since I plan to be showing top 10 graphs, I have also subset the four main variables into top ten lists of events by fatalities, injuries, property and crop damage.

```r
top10fatality<-head(orderedFatalities,10)
top10injury<-head(orderedInjuries,10)
top10property<-head(orderedPropertyCost,10)
top10crop<-head(orderedCropCost,10)
```

And I will clear up (recalculate) the excess factors.

```r
top10fatality$EVTYPE<-factor(top10fatality$EVTYPE)
top10injury$EVTYPE<-factor(top10injury$EVTYPE)
top10property$EVTYPE<-factor(top10property$EVTYPE)
top10crop$EVTYPE<-factor(top10crop$EVTYPE)
```

Results

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

In order to answer the question, let’s see the top 10 events summed by total injuries caused from the entire database.

```
top10injury
```

```
##                EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

As we can see, tornadoes cause the most injuries by far, followed by thunderstorm winds and floods. Excessive heat takes fourth place, and lighning comes fifth.

And now the top 10 events which have caused the most fatalities.

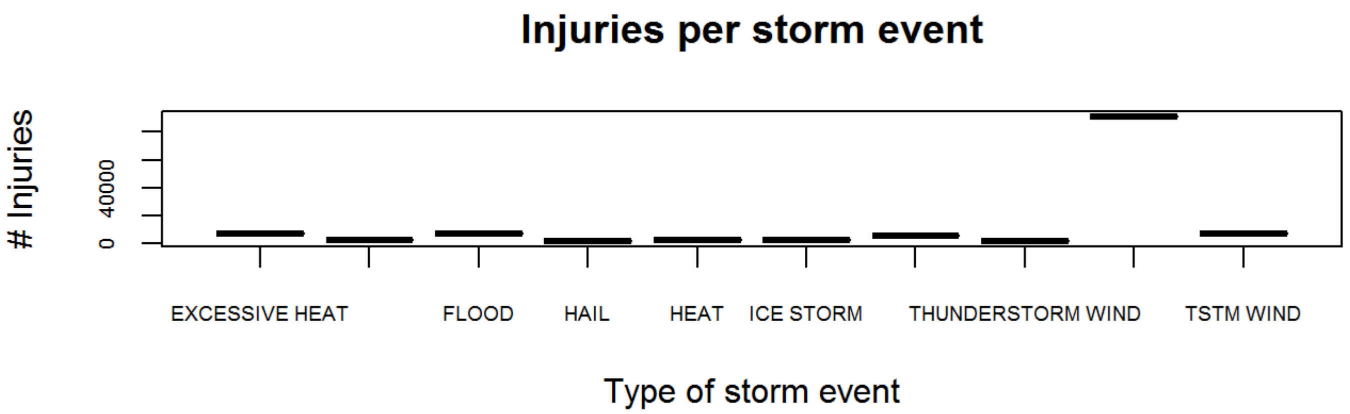
```
top10fatality
```

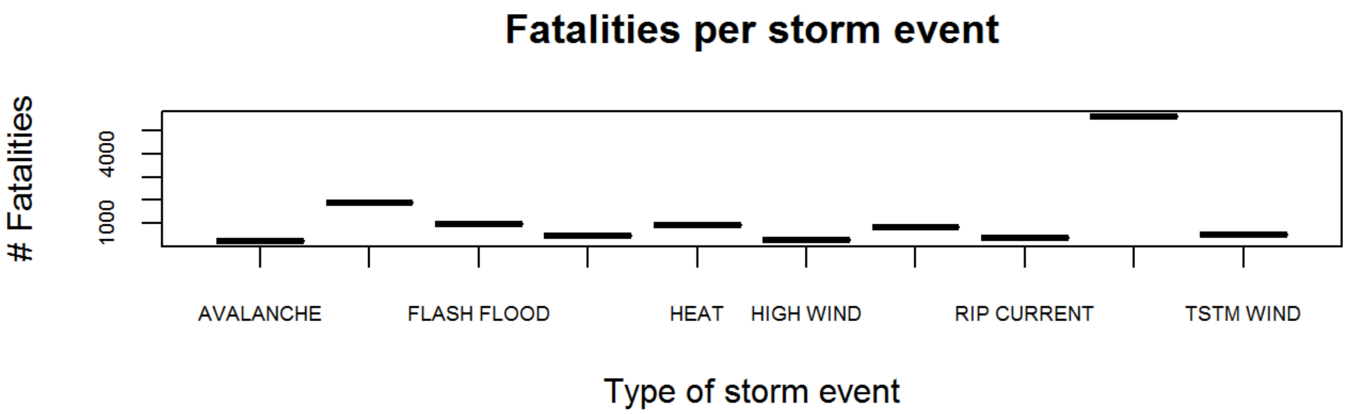
```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

We can see that tornadoes win again with 5933 fatalities, followed by excessive heat, flash floods, head and lightning in the fifth place again.

And now, a graph for comparison.

```
par(mfrow=c(2,1))
with(top10injury,plot(EVTYPE,INJURIES,type="h",main="Injuries per storm event",xlab="Type of
 storm event",ylab="# Injuries",cex.axis=0.6))
with(top10fatality,plot(EVTYPE,FATALITIES,type="h",main="Fatalities per storm event",xlab="T
ype of storm event",ylab="# Fatalities", cex.axis=0.6))
```





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

In order to answer the question, let’s see the top 10 events summed by total property damages caused from the entire database.

```
top10property
```

```
##                EVTYPE truePropDmg
## 62               FLOOD   1.447e+11
## 178 HURRICANE/TYPHOON   6.931e+10
## 331            TORNADO   5.693e+10
## 280        STORM SURGE   4.332e+10
## 50         FLASH FLOOD   1.614e+10
## 103               HAIL   1.573e+10
## 170          HURRICANE   1.187e+10
## 339     TROPICAL STORM   7.704e+09
## 397       WINTER STORM   6.688e+09
## 155          HIGH WIND   5.270e+09
```

We can see that floods have caused the most property damage so far with over 144 BN $ in damage, followed by hurricanes with 69 BN damage and tornadoes being “only” third with 59 BN $ in damages.

And now the top 10 events which have caused the most crop damages.

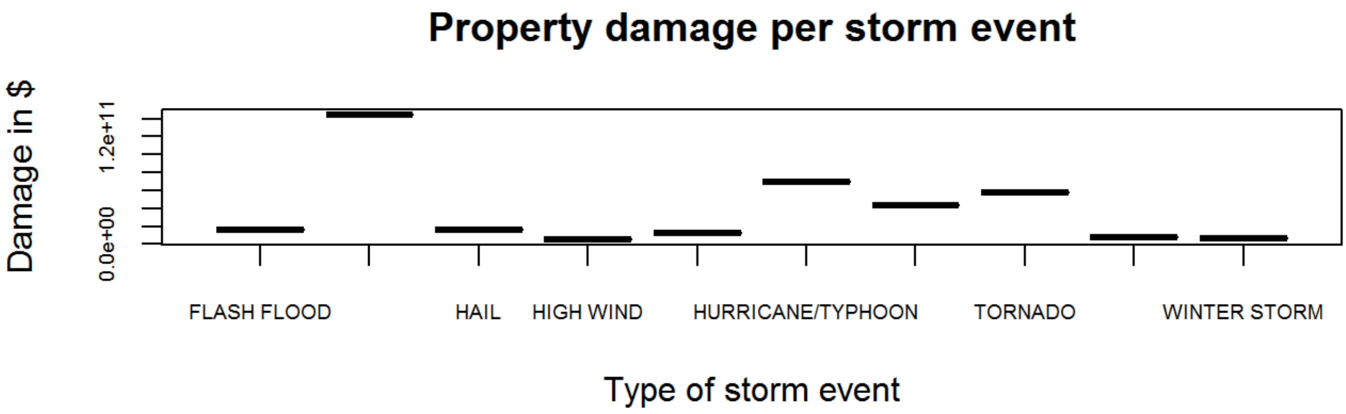
```
top10crop
```

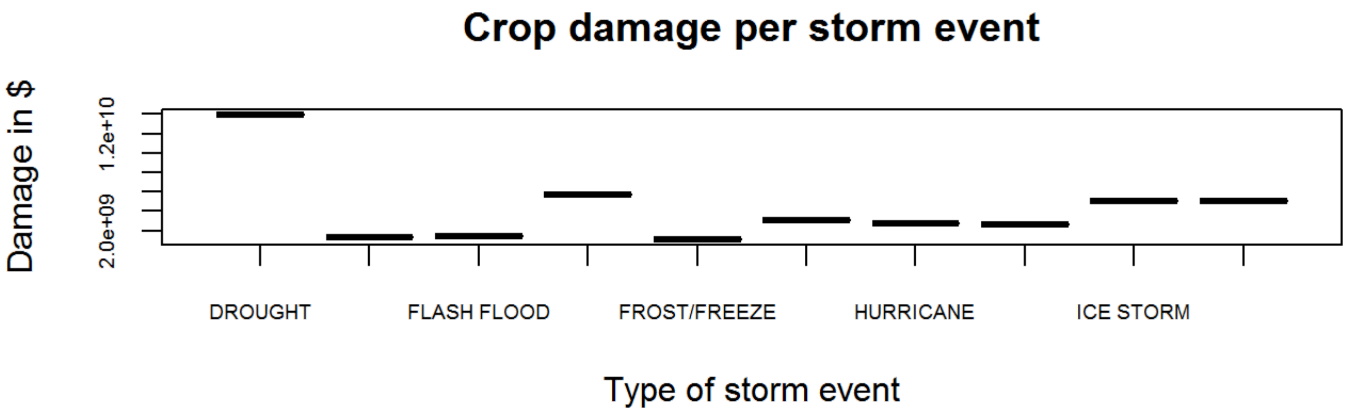
```
##               EVTYPE trueCropDmg
## 16           DROUGHT   1.397e+10
## 34             FLOOD   5.662e+09
## 97       RIVER FLOOD   5.029e+09
## 84         ICE STORM   5.022e+09
## 52              HAIL   3.026e+09
## 76         HURRICANE   2.742e+09
## 81 HURRICANE/TYPHOON   2.608e+09
## 30       FLASH FLOOD   1.421e+09
## 26      EXTREME COLD   1.293e+09
## 46      FROST/FREEZE   1.094e+09
```

While floods may have taken the lead in property damage, droughts “win” with most damage done to crops, with floods in second and third place.

Again, a graphic representation:

```
par(mfrow=c(2,1))
with(top10property,plot(EVTYPE,truePropDmg,type="h",main="Property damage per storm event",x
lab="Type of storm event",ylab="Damage in $",cex.axis=0.6))
with(top10crop,plot(EVTYPE,trueCropDmg,type="h",main="Crop damage per storm event",xlab="Typ
e of storm event",ylab="Damage in $", cex.axis=0.6))
```


Property damage per storm event
Damage in $
0.0e+00
1.2e+11
FLASH FLOOD
HAIL
HIGH WIND
HURRICANE/TYPHOON
TORNADO
WINTER STORM
Type of storm event


Crop damage per storm event
Damage in $
2.0e+09
1.2e+10
DROUGHT
FLASH FLOOD
FROST/FREEZE
HURRICANE
ICE STORM
Type of storm event