Weather Impact Analisys

Marcelo Guimaraes

8 de setembro de 2016

Summary

We report the weather events impact on economic and population health in United States. The data collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database were analysed for this report. According to this analysis the greatest whether event affecting the population health is due to _____ which resulted in _____ deaths and _____ injuries each year. The _____ event caused the most economict impact where $ in damages were estimated.

Intro

There is great interest in the research of weather events. Since the resources provided to avert critical events are limited, the information of the events with the most impact on the population must be known in order to prioritize available resources to the most critical scenarios first, and to invest in the forecast and warning of those events. In this report data collected from the National Weather Service (NSW). The specification of the columns of the datasets are specified at this link. Due to time constraints many of the sources used to produce these data were unverified and we quote "Accordingly, the NWS does not guarantee the accuracy or validity of the information", read this report with that information in mind.

Data Processing

Gathering and exploring Data

In order to investigate the economic and health impact of the weather events the following columns were selected:

1. PROPDMG - Property Damage
2. PROPDMGEXP - Scale (Exponent)
3. CROPDMG - Crop Damage
4. CROPDMGEXP - Scale (Exponent)
5. FATALITIES - Number of deaths
6. INJURIES - Number of injuries

The weather events registred by the EVTYPE columns are redundant. For example, "HEAT", "EXCESSIVE HEAT", "HIGH TEMPERATURE", all represents the same event. It is necessary to agragate these events into the same category before begining the analysis.

```
source("scale.R")
library(data.table)
library(ggplot2)
storm.complete=data.table(read.csv("resources/repdata_data_StormData.csv"))
mapset=getEvents(storm.complete)
storm=applyMapset(storm.complete,mapset)
```

The events were reduced into the following categories: WIND, TORNADO, HURRICANE|TYPHOON, HEAT, STORM, COLD, FLOOD, TSUNAMI, FIRE, HAIL, VOLCANIC, SUMMAR, SLEET, SNOW, RAIN, SEA, LANDSLIDE, LIGHTNING, WIND, FUNNEL CLOUD, WATERSPOUT, WET, DRY, AVALANCHE, BLIZZARD, FOG|SMOKE, DROWNING, SURF, SWELL, GUST, DAM, DUST, OTHER

It can be seen that some exponents were not acurately specified ("?","+","-"). There are considered to be 0.

```
levels(storm$CROPDMGEXP)
```

```
## [1] ""  "?" "0" "2" "B" "k" "K" "m" "M"
```

```
levels(storm$PROPDMGEXP)
```

```
##  [1] ""  "-" "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8" "B" "h" "H" "K"
## [18] "m" "M"
```

These scales are applied in order to obtain the top events impacts based on FATALITIES, TOTAL PROPERTY DAMAGE defined as the sum of CROP DAMAGE and PROPERTY DAMAGE and the number of AFFECTED PEOPLE defined as the sum of the number of FATALITIES and the number of people with INJURIES. Each quantity is grouped by each event:

```
storm$CROPDMG=ApplyScale(storm$CROPDMGEXP,storm$CROPDMG)
storm$PROPDMG=ApplyScale(storm$PROPDMGEXP,storm$PROPDMG)
TotalStorm=storm[,.(FATALITIES=sum(FATALITIES),INJURIES=sum(INJURIES),CROP_DAMAGE=sum(CROPDMG),PROPERTY
total=getTop("TOTAL",TotalStorm,T,5)
totalv=getTopValues("TOTAL",TotalStorm,5)
total
```

```
##    SOURCE TOTAL_PROPERTY_DAMAGE AFFECTED_PEOPLE FATALITIES
## 1:  TOTAL                FLOOD         TORNADO    TORNADO
## 2:  TOTAL     HURRICANE|TYPHOON            WIND       HEAT
## 3:  TOTAL                STORM            HEAT      FLOOD
## 4:  TOTAL              TORNADO           FLOOD       WIND
## 5:  TOTAL                 WIND       LIGHTNING  LIGHTNING
```

```
totalv
```

```
##    SOURCE TOTAL_PROPERTY_DAMAGE AFFECTED_PEOPLE FATALITIES
## 1:  TOTAL          180529000734           97043       5636
## 2:  TOTAL           90762527810           12949       3178
## 3:  TOTAL           73537420711           12421       1557
## 4:  TOTAL           57418279946           10238       1451
## 5:  TOTAL           20110518812            6048        817
```

On a first glance it seems that the FLOOD is the main cause of property damage and TORNADO has the main effect on peoples health.

To make sure that this analysis is realiable we visually explore the distribuition of those events on each year:

```
library("cowplot")
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

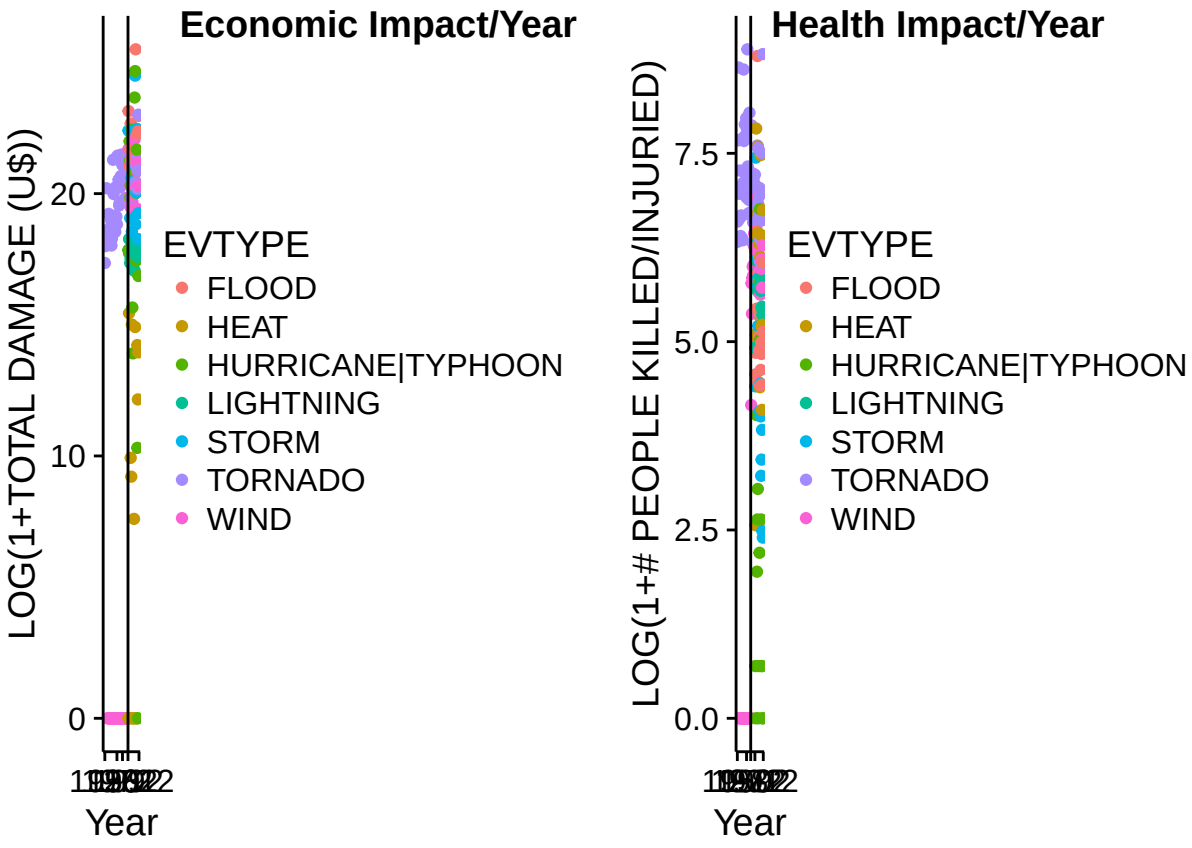
```
storm$Year=year(as.Date(gsub(" .*","",as.character(storm$BGN_DATE)),format="%m/%d/%Y"))
TotalStormYear=storm[,.(FATALITIES=sum(FATALITIES),INJURIES=sum(INJURIES),CROP_DAMAGE=sum(CROPDMG),PROPE
TotalStormYear[,TOTAL_DAMAGE:=CROP_DAMAGE+PROPERTY_DAMAGE]
TotalStormYear[,TOTAL_HEALTH_DAMAGE:= FATALITIES+INJURIES]

events=unique(as.array(as.matrix(total[,-1,with=FALSE])))[1:15])
datas=c("1950","1972","1982", "1992","2012")
ndatas=as.numeric(datas)

plot_dmg=ggplot(data=TotalStormYear[EVTYPE %in% events],aes(x=Year,y=log(1+TOTAL_DAMAGE),colour=EVTYPE))
        geom_vline(xintercept=1992)


plot_health=ggplot(data=TotalStormYear[EVTYPE %in% events],aes(x=Year,y=log(1+TOTAL_HEALTH_DAMAGE),colou
        geom_vline(xintercept=1982)+ylab("LOG(1+# PEOPLE KILLED/INJURIED)")


plot_grid(plot_dmg, plot_health, labels=c("Economic Impact/Year", "Health Impact/Year"), ncol = 2, nrow
```



It is clear from the data that the properties damages was recorded only for TORNADOS until 1992. Similarly weather related fatalities/injuries other than from TORNADOS were included only after 1982. Lets cut the data before 1992:

```
TotalStorm83=storm[Year >1982,.(FATALITIES=sum(FATALITIES),INJURIES=sum(INJURIES),CROP_DAMAGE=sum(CROPDM
post82=getTop("POST 82 TOTAL",TotalStorm83,T,5)
post82v=getTopValues("POST 82 TOTAL",TotalStorm83,5)

TotalStorm93=storm[Year >1992,.(FATALITIES=sum(FATALITIES),INJURIES=sum(INJURIES),CROP_DAMAGE=sum(CROPDM
post92=getTop("POST 92 TOTAL",TotalStorm93,T,5)
post92v=getTopValues("POST 92 TOTAL",TotalStorm93,5)

post82.92=post82
post82.92$TOTAL_PROPERTY_DAMAGE=post92$TOTAL_PROPERTY_DAMAGE
post82.92v=post82v
post82.92v$TOTAL_PROPERTY_DAMAGE=post92v$TOTAL_PROPERTY_DAMAGE
post82.92$SOURCE="POST 82/92"
post82.92v$SOURCE="POST 82/92"
post82.92
```

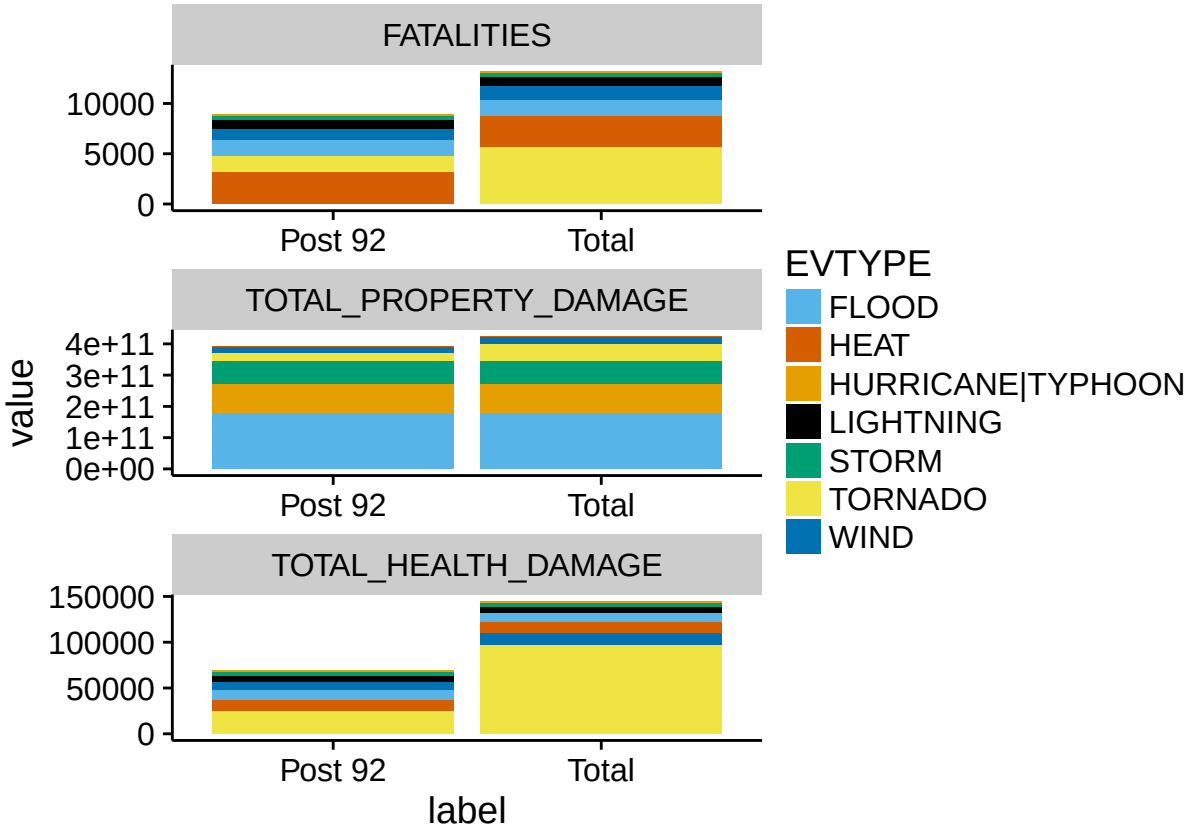
```
##        SOURCE TOTAL_PROPERTY_DAMAGE AFFECTED_PEOPLE FATALITIES
## 1: POST 82/92                FLOOD         TORNADO       HEAT
## 2: POST 82/92     HURRICANE|TYPHOON            WIND    TORNADO
## 3: POST 82/92                STORM            HEAT      FLOOD
## 4: POST 82/92              TORNADO           FLOOD       WIND
## 5: POST 82/92                 WIND       LIGHTNING  LIGHTNING
```

```
post82.92v
```

```
##        SOURCE TOTAL_PROPERTY_DAMAGE AFFECTED_PEOPLE FATALITIES
## 1: POST 82/92          180529000734           36962       3178
## 2: POST 82/92           90762527810           12949       2161
## 3: POST 82/92           73537420711           12421       1557
## 4: POST 82/92           26820081376           10238       1451
## 5: POST 82/92           20110518812            6048        817
```

Now it appears that HEAT is the main cause of fatalities.

```
avgstorm=storm[Year>1992,.(FATALITIES=mean(FATALITIES),INJURIES=mean(INJURIES),CROP_DAMAGE=mean(CROPDMG)
avg=getTop("POST 92 MEAN",avgstorm,F)
avgv=getTopValues("POST 92 MEAN",avgstorm)
results=rbind(post82.92,avg,total)
resultsv=rbind(post82.92v,avgv,totalv)
resultsv$TOTAL_PROPERTY_DAMAGE=resultsv$TOTAL_PROPERTY_DAMAGE/10^9
colnames(resultsv)[2]="PROPERTY DAMAGE (BILLIONS OF US$)"
events=unique(as.array(as.matrix(results[,-1,with=FALSE]))[1:33])
short=TotalStorm[as.character(EVTYPE) %in% events,c(1,2,6,7),with=F]
short$label="Total"
short93=TotalStorm93[as.character(EVTYPE) %in% events,c(1,2,6,7),with=F]
short93$label="Post 92"
shorts=rbind(short,short93)
mr=melt(shorts,id.vars=c("label", "EVTYPE"))
setorder(mr,-value)
cbbPalette <- c("#56B4E9","#D55E00","#E69F00", "#000000",  "#009E73", "#F0E442", "#0072B2",  "#CC79A7")
ggplot(data=mr,aes(x=label,y=value,fill=EVTYPE,order=value)) + geom_bar(stat="identity")+facet_wrap(~va
```


FATALITIES
10000
5000
0
Post 92
Total
TOTAL_PROPERTY_DAMAGE
4e+11
3e+11
2e+11
1e+11
0e+00
value
Post 92
Total
TOTAL_HEALTH_DAMAGE
150000
100000
50000
0
Post 92
Total
label
EVTYPE
FLOOD
HEAT
HURRICANE|TYPHOON
LIGHTNING
STORM
TORNADO
WIND

Results