Severe weather Effects on Economy and Publick Health in the US

Patrick L

October 22, 2016

Synopsis.

Did you know that every year in the US, Drought and Flood alone cause upward to **27.5 billions** in damages? This is a finding from the analysis of 61 years of reported economic damages, as shown in this report. For the policy maker, understanding the effects of these meteorological events on the economy and on public health, could guide the allocation of emergency ressources, which could be the difference between a resilient community with a vibrant economy and a catastrophe further compounded by the lack of response from appropriate authorities. This reports attempts to find, from a historical perspective, what type of weather events has caused, over the last 61 years period in the US:

- the most harmful effect on population health
- the greatest economic impacts

Mining of the NOAA Storm Database found that over 2 decades starting in 1996:

- Tornadoes were a danger to public health. They caused the highest level of fatalities and injuries with well over 85000 recorded cases,
- Drought and floods were the most disastrous to the economy, with a combined $27.5bn loss on average annually[1].

These results do not account for all possible weather events, hence could be considered conservative.

The analysis is based on NOAA's events classification method documented here and the data from the Storm Database.

Before cleaning and processing, it's very important to research the raw data under investigation. In this particular case web research revealed what preliminary exploratory analysis did not: different types of weather event have been recorded for different numbers of years. Only since January 1996 have all event types been recorded each year. For that reason, this report only considers the last ~20 year period (Jan 1996-July 2016 inclusive).

Data Processing

Downloading data from the Web

The first thing needed is to download the weather database file fron the interwebs. A local data folder is created for that purpose and will contain the zipped file.

```r
#Check if a data folder exists, if not create
if(!file.exists("data")){
  dir.create("data")
}
#setwd("data")
```

```r
# get data from the web
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
destinationFile="./data/StormData.csv.bz2"

if(!file.exists(destinationFile)){
  download.file(fileUrl,destfile=destinationFile,method="curl")
}
```

Loading and Processing the Raw Data

The raw data is loaded from the database into the *weatherData* data frame.

```r
# load data from file if not loaded already
if (!exists("weatherData"))
  weatherData <- read.csv(destinationFile)
```

Let's peek into the data with the str() command.

```r
str(weatherData)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "10/10/1954 0:00:00",..: 6523 6523 4213 11116 1426 1426 1462 28
##  $ BGN_TIME  : Factor w/ 3608 levels "000","0000","00:00:00 AM",..: 212 257 2645 1563 2524 3126 122 1
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 1
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "?","ABNORMALLY DRY",..: 830 830 830 830 830 830 830 830 830 830
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","E","Eas","EE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","?","(01R)AFB GNRY RNG AL",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","10/10/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","?","0000",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","(0E4)PAYSON ARPT",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","2","43","9V9",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
```

```
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels ""," ","  ","   ",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Couple of observations can be made:

1. Even though there is a 'BGN TIME' field, the BGN_DATE field includes the time.
2. The output seems to suggest (from the STATE attribute) that there is data for 72 differents US STATES in the database. Since we do care much about localizing weather events, we'll ignore it.
3. The Storm Data Event Table (Table 2.1.1 in Instruction Document, Page 6) alludes to the existence of 48 types of events. However, the EVTYPE attribute identifies 985 unique type of events.

```r
names(weatherData)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

```r
# head(weatherData)
```

Further exploration of the data might be useful. But before, let's subset the data and extract only attributes that are meaningful.

```r
weatherData.subset <- weatherData [,c("BGN_DATE","EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP
```

From observation #1, we can fix the date attribute to reflect only the date and not the time. Going even further, we'll keep only the year the event was recorded.

```r
weatherData.subset$BGN_DATE = as.numeric(format(as.Date(weatherData$BGN_DATE,format = "%m/%d/%Y"),"%Y"))
summary(weatherData.subset$BGN_DATE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1950    1995    2002    1999    2007    2011
```

as shown in teh summary above, we have 61 years worth of recorded events, from 1950 to 2011.

```r
levels(weatherData.subset$PROPDMGEXP)
```

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

```r
levels(weatherData.subset$CROPDMGEXP)
```

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

As seen above, couple of attributes (PROPDMGEXP and PROPDMGEXP) are labelled for the multiplier to the recorded property (PROPDMG) and crop (CROPDMG) damage amount. Some labels are letters that need to be changed to their equivalent exponent values. For example K or k will be replaced by 3 (i.e: k stands for 10e3) and "" - ? + will be replaced by 0.

```r
levels(weatherData.subset$PROPDMGEXP) <- c(0,0,0,0,0,1,2,3,4,5,6,7,8,9,2,2,3,6,6)
weatherData.subset$PROPDMGEXP <- as.numeric(weatherData.subset$PROPDMGEXP)

levels(weatherData.subset$CROPDMGEXP) <- c(0,0,0,2,9,2,2,6,6)
weatherData.subset$CROPDMGEXP <- as.numeric(weatherData.subset$CROPDMGEXP)
```

From observation #3, another foray into the values recorded in EVTYPE showed that recordings were made using both upper and lower case characters. This could lead to a possible separation of event types that are actually identical. Same could be said of the multiple spelling of the same type of events. The event table identifies only one type of thunderstom event but the data actually contains a multitude of variations. The attemps at a solution to the aforementioned issues consist in turning all EVTYPE to upper case and then correcting few of the one that could be identified.

```r
# Uppercase all EVTYPES
weatherData.subset$EVTYPE = toupper(as.character(weatherData.subset$EVTYPE))

# this function replaces an occurence of the word that is submitted by the equivalent type in the Event
fixEventType <- function(txt){
       if( sum(grepl("HAIL",txt)) )
    return ("HAIL")
  else if ( sum(grepl("TSTM",txt)))
    return ("THUNDERSTORM WIND")
  else if ( sum(grepl("THUNDER",txt)))
    return ("THUNDERSTORM WIND")
  else if ( sum(grepl("FLOOD",txt)))
    return ("FLASH FLOOD")
  else if ( sum(grepl("TORN",txt)))
    return ("TORNADO")
  else if ( sum(grepl("FLASH",txt)))
    return ("FLASH FLOOD")
  else if ( sum(grepl("HEAT",txt)))
    return ("EXCESSIVE HEAT")
  else
    return (txt)

}
weatherData.subset$EVTYPE <- sapply(weatherData.subset$EVTYPE, FUN=fixEventType)
# Convert back to factor
weatherData.subset$EVTYPE <- as.factor(weatherData.subset$EVTYPE)
# Count the numbers of unique types of events
length(unique(weatherData.subset$EVTYPE))
```

```
## [1] 624
```

The correction in EVTYPE attribute fixed 348 types of mislabeled records and brought the numer of unique type of events to 637. This is still far from the 48 types in the documentation but would provide us with a more accurate figure than before.

Now we need to evaluete the total cost of each weather occurence. For that we'll add two columns in the data: one for the total Property damage cost and another for the total crop damage.

```
weatherData.subset$PROPDMGTOT <- weatherData.subset$PROPDMG*10^weatherData.subset$PROPDMGEXP
weatherData.subset$CROPDMGTOT <- weatherData.subset$CROPDMG*10^weatherData.subset$CROPDMGEXP
```

Based on that, we can summarize the costs of each event type, sorted in decreasing order of cost, and get the 5 most impactful.

```
top5.propdmg <- sort(tapply(weatherData.subset$PROPDMGTOT, weatherData.subset$EVTYPE, sum), decreasing =
top5.cropdmg <- sort(tapply(weatherData.subset$CROPDMGTOT, weatherData.subset$EVTYPE, sum), decreasing =
top5.fatalities <- sort(tapply(weatherData.subset$FATALITIES, weatherData.subset$EVTYPE, sum), decreasin
top5.injuries <- sort(tapply(weatherData.subset$INJURIES, weatherData.subset$EVTYPE, sum), decreasing =
```

We can look at the trend over time for each of the metrics

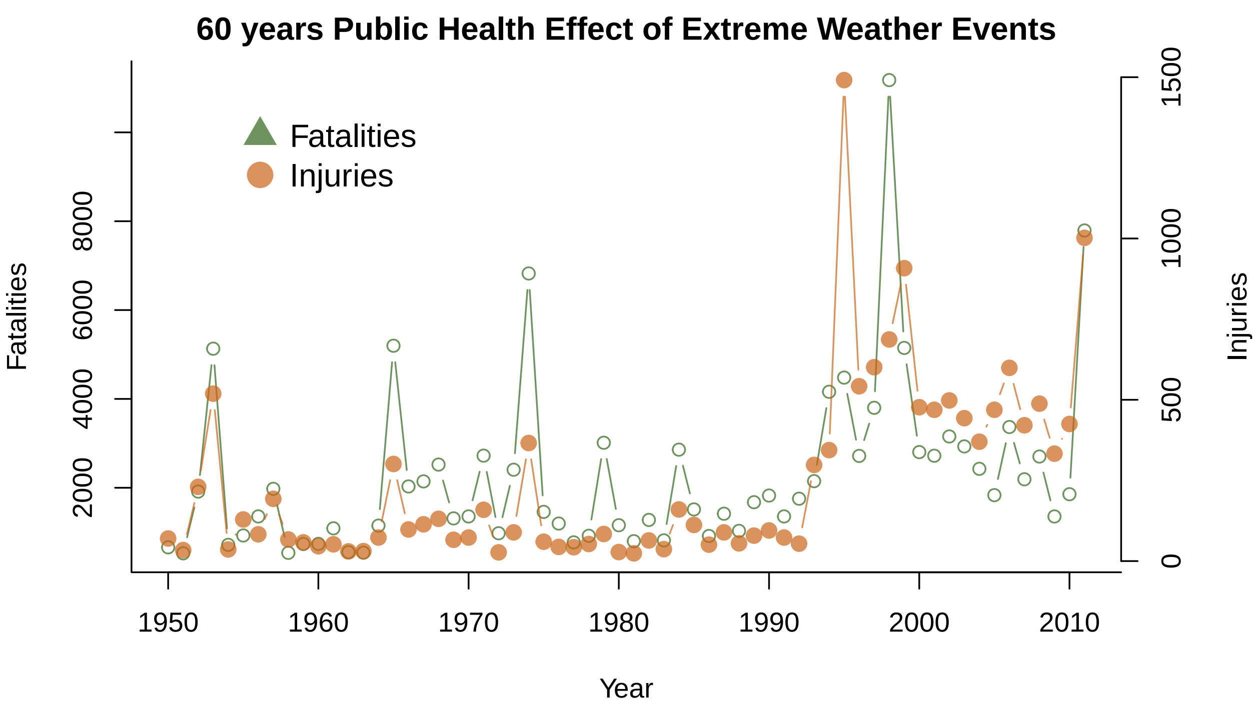
```
yty.propdmg <- tapply(weatherData.subset$PROPDMGTOT, weatherData.subset$BGN_DATE, sum)
yty.cropdmg <- tapply(weatherData.subset$CROPDMGTOT, weatherData.subset$BGN_DATE, sum)
yty.fatalities <- tapply(weatherData.subset$FATALITIES, weatherData.subset$BGN_DATE, sum)
yty.injuries <- tapply(weatherData.subset$INJURIES, weatherData.subset$BGN_DATE, sum)

# Normalize the data given the range of values
yty.injuries.norm <- (yty.injuries - min(yty.injuries))/(max(yty.injuries)-min(yty.injuries))
yty.fatalities.norm <- (yty.fatalities - min(yty.fatalities))/(max(yty.fatalities)-min(yty.fatalities))
```

Results

60 years Impact of Extreme Weather Events

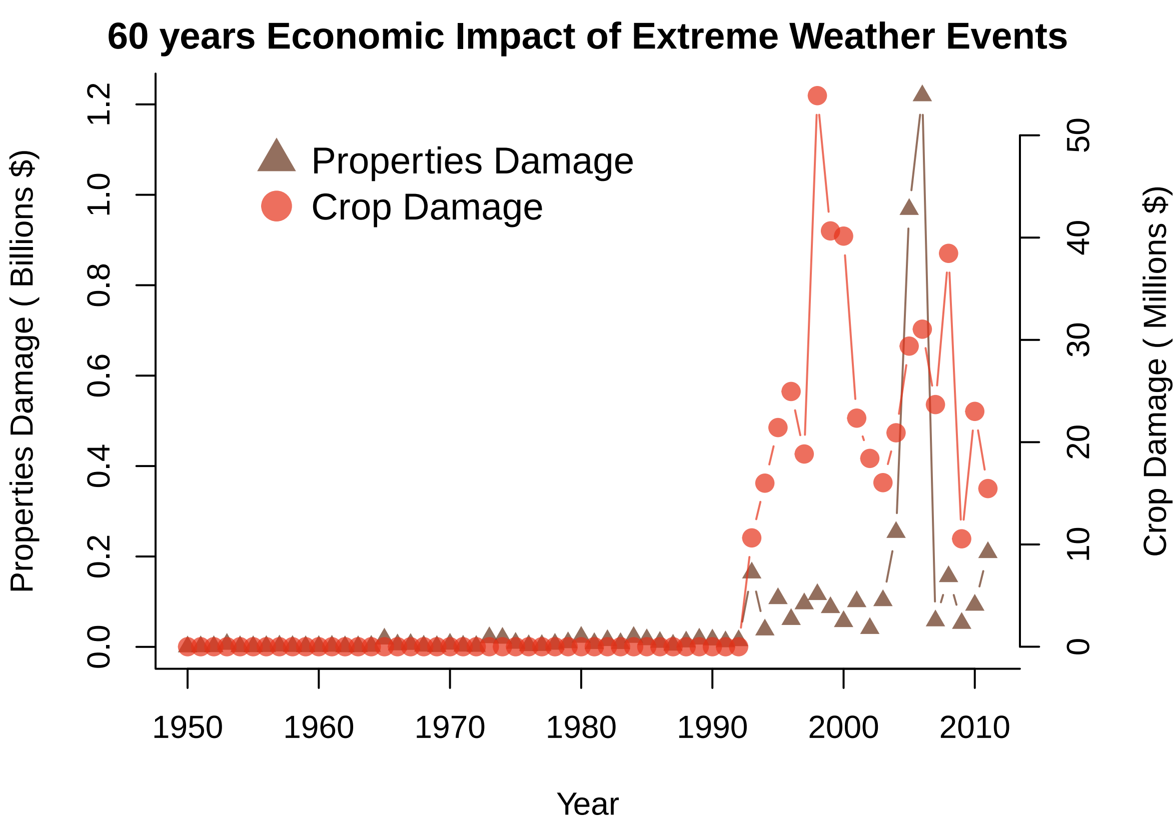
```
par(mar = c(5,5,2,5))
plot( names(yty.injuries),
      yty.injuries,
      type = "b",
      col=rgb(0.2,0.4,0.1,0.7) ,
      xlab = "Year",
      ylab = "Fatalities",
      bty="l"
      )

par(new = T)
plot( names(yty.fatalities),
      yty.fatalities,
      type = "b",
      pch=19,
      col=rgb(0.8,0.4,0.1,0.7),
      axes = F,
      xlab=NA, ylab=NA, cex=1.2
```

```
      )

axis(side = 4)
mtext(side = 4, line = 3, 'Injuries')

# Add a legend
legend("bottomleft",
  legend = c("Fatalities", "Injuries"),
  col = c(rgb(0.2,0.4,0.1,0.7),
  rgb(0.8,0.4,0.1,0.7)),
  pch = c(17,19),
  bty = "n",
  pt.cex = 2,
  cex = 1.2,
  text.col = "black",
  horiz = F ,
  inset = c(0.1, 0.7))
title("60 years Public Health Effect of Extreme Weather Events")
```




```
par(mar = c(5,5,2,5))
plot( names(yty.propdmg),
      yty.propdmg/1e12,
      type = "b",
      pch=17,
      col=rgb(0.4,0.2,0.1,0.7) ,
      xlab = "Year",
      ylab = "Properties Damage ( Billions $)",
      bty="l"
      )

par(new = T)
```

```
plot( names(yty.cropdmg),
      yty.cropdmg/1e6,
      type = "b",
      pch=19,
      col=rgb(0.9,0.2,0.1,0.7),
      axes = F,
      xlab=NA, ylab=NA, cex=1.2
      )

axis(side = 4)
mtext(side = 4, line = 3, 'Crop Damage ( Millions $)')

# Add a legend
legend("bottomleft",
  legend = c("Properties Damage", "Crop Damage"),
  col = c(rgb(0.4,0.2,0.1,0.7),
  rgb(0.9,0.2,0.1,0.7)),
  pch = c(17,19),
  bty = "n",
  pt.cex = 2,
  cex = 1.2,
  text.col = "black",
  horiz = F ,
  inset = c(0.1, 0.7))

title("60 years Economic Impact of Extreme Weather Events")
```



Impact on the Economy

Property Damage from Weather Events

```
print("Property Damage in Trillions of Dollars from 1950 to 2011")
```

```
## [1] "Property Damage in Trillions of Dollars from 1950 to 2011"
```

```
top5.propdmg / 1e12
```

```
##       FLASH FLOOD HURRICANE/TYPHOON           TORNADO       STORM SURGE
##         1.6821169         0.6930584         0.5700332         0.4332354
##              HAIL
##         0.1762299
```

Flash Flood had the greatest impact by causing *1.6 Trillions* of dollars in property damages

Agricultural impact of Weather Events

```
top5.cropdmg/1e6
```

```
##           DROUGHT       FLASH FLOOD              HAIL THUNDERSTORM WIND
##         126.65730         105.97621          83.64862          29.51205
##         HURRICANE
##          27.65310
```

The agricultural impact was paradoxically felt greatly when caused first by *Drought* and then by *Floods*

Weather Events - Impact on Public Health

Fatalities from Weather Events

```
top5.fatalities
```

```
##           TORNADO    EXCESSIVE HEAT       FLASH FLOOD         LIGHTNING
##              5636              3138              1525               816
## THUNDERSTORM WIND
##               726
```

The greater number of fatalities are caused first by *Tornadoes* and second by *Excessive Heat*

Injuries from Weather Events

```
top5.injuries
```

```
##           TORNADO THUNDERSTORM WIND    EXCESSIVE HEAT       FLASH FLOOD
##             91407              9449              9224              8604
##         LIGHTNING
##              5230
```

The cause of most injuries are manly *tornadoes* again and then by *Thunderstorm winds*

Conclusion

Tornadoes have had the greatest impact on public health over the 61 years period from 1950 to 2011 with an average of 92 deaths and 1500 injuries every year. On the other hand, the biggest economical impact was caused by both *Drought* and *Floods*

[1] Annual average based on data from 1950 to 2011. Recents years not included.