Weather events’ influence on public health and economy in the United States

Soundouss Messoudi

September 27th, 2015

Synopsis

This report’s objective is to study the severe weather events that can cause both public health and economic problems for communities and municipalities in the US based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration’s (NOAA) from 1950 - 2011. Indeed, many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern. The report identifies the most harmful weather events that cause both public health (as measured by mean number of combined fatalities and injuries) and economic (as measured by the mean property damage and mean crop damage sustained during the event) problems. The results of this analysis show that convection and extreme temperatures are the weather events that have the greatest impact on public health. Whereas the extremely harmful events for economy are convection and flood.

About the Data

The weather events are divided into 13 groups:

- Convection (e.g. tornado, lightning, thunderstorm, hail)
- Flood (e.g. flash flood, river flood)
- Extreme temperatures (e.g. extreme cold, extreme hot)
- Marine (e.g. tsunami, coastal storm, rip current, high waves, high seas)
- Winter (e.g. avalanche, snow, blizzard, icy roads, freeze)
- Tropical Cyclones (e.g. tropical storm, hurricane)
- High Wind (e.g. winds, microburst)
- Fire
- Rain
- Drought/Dust (e.g. drought, dust storm, dust)
- Landslide
- Fog
- Others

Basic settings

```
echo = TRUE  # Always make code visible
options(scipen = 1)  # Turn off scientific notations for numbers
library(R.utils)
```

```
## Warning: package 'R.utils' was built under R version 3.2.2
```

```
## Loading required package: R.oo
```

```
## Warning: package 'R.oo' was built under R version 3.2.2
```

```
## Loading required package: R.methodsS3
```

```
## Warning: package 'R.methodsS3' was built under R version 3.2.2
```

```
## R.methodsS3 v1.7.0 (2015-02-19) successfully loaded. See ?R.methodsS3 for help.
## R.oo v1.19.0 (2015-02-27) successfully loaded. See ?R.oo for help.
##
## Attaching package: 'R.oo'
##
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
##
## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save
##
## R.utils v2.1.0 (2015-05-27) successfully loaded. See ?R.utils for help.
##
## Attaching package: 'R.utils'
##
## The following object is masked from 'package:utils':
##
##     timestamp
##
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.2
```

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
## Warning: package 'gridExtra' was built under R version 3.2.2
```

Data Processing

After downloading the data, unziping it and putting it in the working directory, we read the "repdata-data-StormData.csv" file.

```
if (!"stormData" %in% ls()) {
    stormData <- read.csv("repdata-data-StormData.csv", sep = ",")
}
dim(stormData)
```

```
## [1] 902297     37
```

```
head(stormData, n = 2)
```

```
##   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
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14   100 3   0          0
## 2         NA         0                       2   150 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
```

There are 902297 rows and 37 columns in total. The events in the database start in the year 1950 and end in November 2011.

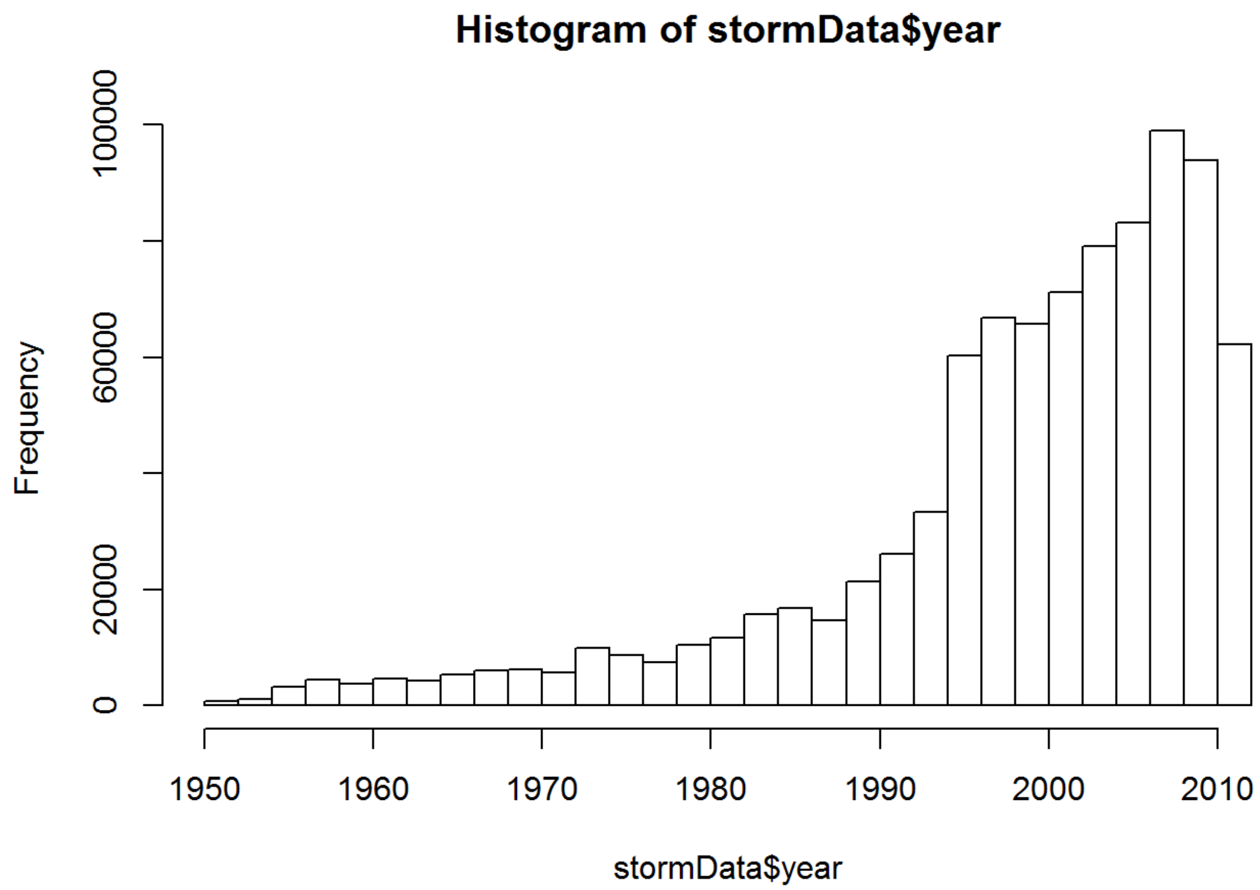
Select useful data

```
if (dim(stormData)[2] == 37) {
    stormData$year <- as.numeric(format(as.Date(stormData$BGN_DATE, format = "%m/%d/%Y
%H:%M:%S"), "%Y"))
}
```

```
## Warning in strptime(x, format, tz = "GMT"): unable to identify current timezone 'M':
## please set environment variable 'TZ'
```

```
## Warning in strptime(x, format, tz = "GMT"): unknown timezone 'localtime'
```

```
hist(stormData$year, breaks = 30)
```



Based on the above histogram, we see that the number of events tracked starts to significantly increase around 1995. So, we use the subset of the data from 1990 to 2011 to get most out of good records.

```
storm <- stormData[stormData$year >= 1995, ]
dim(storm)
```

```
## [1] 681500     38
```

Now, there are 681500 rows and 38 columns in total.

Analyse

Public Health Harmful Events

This histogram Shows fatalities and injuries for weather events.

```
sortHelper <- function(fieldName, top = 15, dataset = stormData) {
    index <- which(colnames(dataset) == fieldName)
    field <- aggregate(dataset[, index], by = list(dataset$EVTYPE), FUN = "sum")
    names(field) <- c("EVTYPE", fieldName)
    field <- arrange(field, field[, 2], decreasing = T)
    field <- head(field, n = top)
    field <- within(field, EVTYPE <- factor(x = EVTYPE, levels = field$EVTYPE))
    return(field)
}

fatalities <- sortHelper("FATALITIES", dataset = storm)
injuries <- sortHelper("INJURIES", dataset = storm)
```

Economic Harmful Events

This histograms show weather events' harm to the economy.

```
convertHelper <- function(dataset = storm, fieldName, newFieldName) {
    totalLen <- dim(dataset)[2]
    index <- which(colnames(dataset) == fieldName)
    dataset[, index] <- as.character(dataset[, index])
    logic <- !is.na(toupper(dataset[, index]))
    dataset[logic & toupper(dataset[, index]) == "B", index] <- "9"
    dataset[logic & toupper(dataset[, index]) == "M", index] <- "6"
    dataset[logic & toupper(dataset[, index]) == "K", index] <- "3"
    dataset[logic & toupper(dataset[, index]) == "H", index] <- "2"
    dataset[logic & toupper(dataset[, index]) == "", index] <- "0"
    dataset[, index] <- as.numeric(dataset[, index])
    dataset[is.na(dataset[, index]), index] <- 0
    dataset <- cbind(dataset, dataset[, index - 1] * 10^dataset[, index])
    names(dataset)[totalLen + 1] <- newFieldName
    return(dataset)
}

storm <- convertHelper(storm, "PROPDMGEXP", "propertyDamage")
```

```
## Warning in convertHelper(storm, "PROPDMGEXP", "propertyDamage"): NAs
## introduits lors de la conversion automatique
```

```
storm <- convertHelper(storm, "CROPDMGEXP", "cropDamage")
```

```
## Warning in convertHelper(storm, "CROPDMGEXP", "cropDamage"): NAs introduits
## lors de la conversion automatique
```

```
names(storm)
```

```
##  [1] "STATE__"        "BGN_DATE"       "BGN_TIME"       "TIME_ZONE"
##  [5] "COUNTY"         "COUNTYNAME"     "STATE"          "EVTYPE"
##  [9] "BGN_RANGE"      "BGN_AZI"        "BGN_LOCATI"     "END_DATE"
## [13] "END_TIME"       "COUNTY_END"     "COUNTYENDN"     "END_RANGE"
## [17] "END_AZI"        "END_LOCATI"     "LENGTH"         "WIDTH"
## [21] "F"              "MAG"            "FATALITIES"     "INJURIES"
## [25] "PROPDMG"        "PROPDMGEXP"     "CROPDMG"        "CROPDMGEXP"
## [29] "WFO"            "STATEOFFIC"     "ZONENAMES"      "LATITUDE"
## [33] "LONGITUDE"      "LATITUDE_E"     "LONGITUDE_"     "REMARKS"
## [37] "REFNUM"         "year"           "propertyDamage" "cropDamage"
```

```
options(scipen=999)
property <- sortHelper("propertyDamage", dataset = storm)
crop <- sortHelper("cropDamage", dataset = storm)
```

Results

Public Health Harmful Events

We have two sorted lists of severe weather events below by the number of people badly affected.

```
fatalities
```

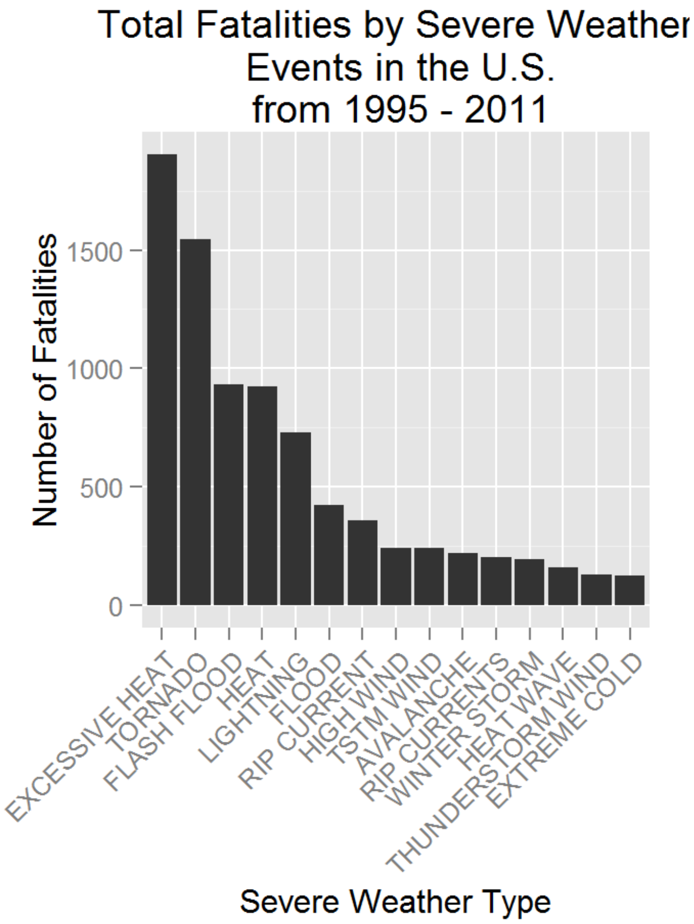
```
##               EVTYPE FATALITIES
## 1     EXCESSIVE HEAT       1903
## 2            TORNADO       1545
## 3        FLASH FLOOD        934
## 4               HEAT        924
## 5          LIGHTNING        729
## 6              FLOOD        423
## 7        RIP CURRENT        360
## 8          HIGH WIND        241
## 9          TSTM WIND        241
## 10         AVALANCHE        223
## 11      RIP CURRENTS        204
## 12      WINTER STORM        195
## 13         HEAT WAVE        161
## 14 THUNDERSTORM WIND        131
## 15      EXTREME COLD        126
```

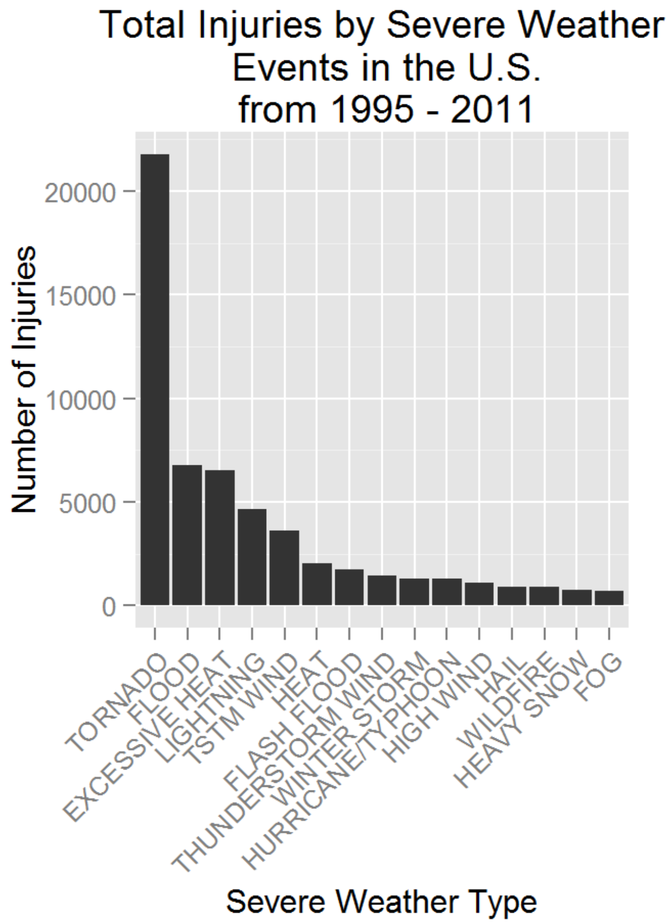
```
injuries
```

```
##              EVTYPE INJURIES
## 1           TORNADO    21765
## 2             FLOOD     6769
## 3    EXCESSIVE HEAT     6525
## 4         LIGHTNING     4631
## 5         TSTM WIND     3630
## 6              HEAT     2030
## 7       FLASH FLOOD     1734
## 8  THUNDERSTORM WIND     1426
## 9      WINTER STORM     1298
## 10 HURRICANE/TYPHOON     1275
## 11         HIGH WIND     1093
## 12              HAIL      916
## 13          WILDFIRE      911
## 14        HEAVY SNOW      751
## 15               FOG      718
```

These are graphs of total fatalities and total injuries affected by these severe weather events.

```
fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", binwidth = 1) + scale_y_continuous("Number of Fatalities") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + xlab("Severe Weather Type") + ggtitle("Total Fatalities by Severe Weather\n Events in the U.S.\n from 1995 - 2011")
injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, geom = "bar", binwidth = 1) + scale_y_continuous("Number of Injuries") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + xlab("Severe Weather Type") + ggtitle("Total Injuries by Severe Weather\n Events in the U.S.\n from 1995 - 2011")
grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)
```





Based on the above histograms, we find that Excessive heat and Tornado cause most fatalities. Moreover, Tornado causes most injuries in the U.S. from 1995 to 2011.

Economic Harmful Events

We have two sorted lists of severe weather events below by the amount of money cost by damages.

```
property
```

```
##               EVTYPE propertyDamage
## 1              FLOOD   144022037057
## 2  HURRICANE/TYPHOON    69305840000
## 3        STORM SURGE    43193536000
## 4            TORNADO    24935939545
## 5        FLASH FLOOD    16047794571
## 6               HAIL    15048722103
## 7          HURRICANE    11812819010
## 8     TROPICAL STORM     7653335550
## 9          HIGH WIND     5259785375
## 10          WILDFIRE     4759064000
## 11  STORM SURGE/TIDE     4641188000
## 12         TSTM WIND     4482361440
## 13         ICE STORM     3643555810
## 14 THUNDERSTORM WIND     3399282992
## 15    HURRICANE OPAL     3172846000
```

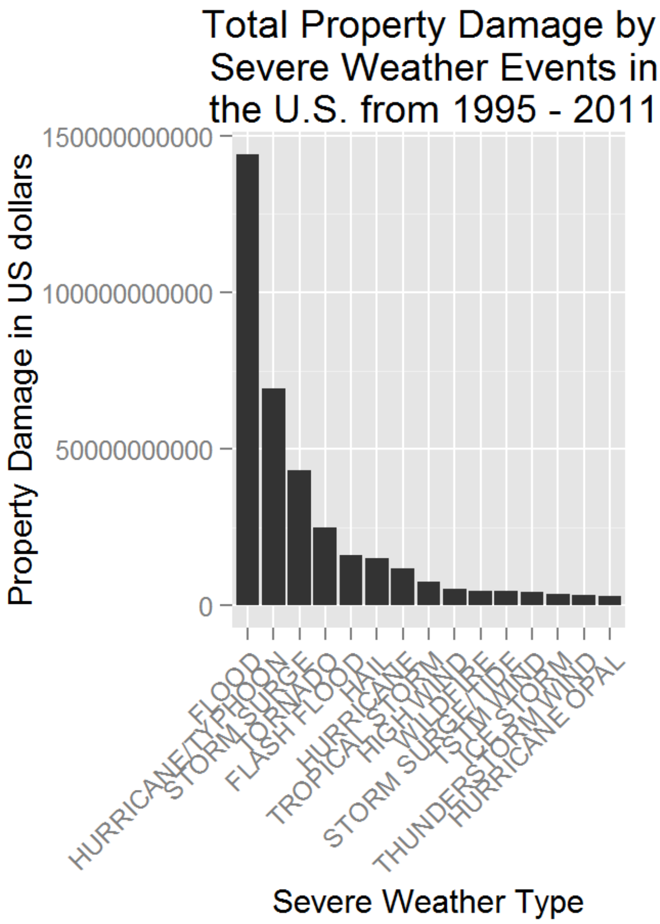
```
crop
```

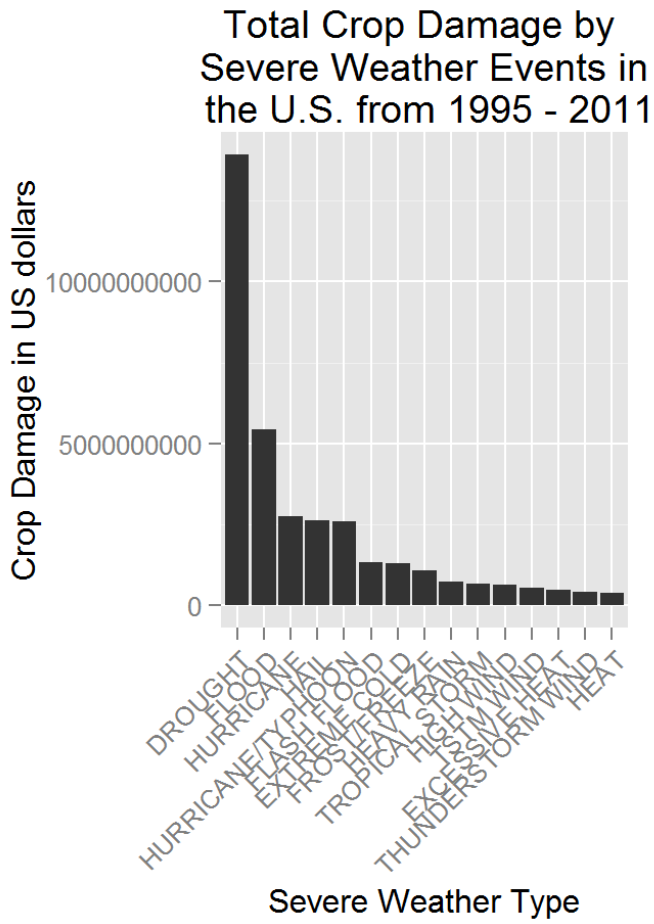
```
##              EVTYPE  cropDamage
## 1           DROUGHT 13922066000
## 2             FLOOD  5422810400
## 3         HURRICANE  2741410000
## 4              HAIL  2614127070
## 5 HURRICANE/TYPHOON  2607872800
## 6       FLASH FLOOD  1343915000
## 7      EXTREME COLD  1292473000
## 8      FROST/FREEZE  1094086000
## 9        HEAVY RAIN   728399800
## 10   TROPICAL STORM   677836000
## 11        HIGH WIND   633561300
## 12        TSTM WIND   553947350
## 13   EXCESSIVE HEAT   492402000
## 14 THUNDERSTORM WIND   414354000
## 15             HEAT   401411500
```

These are graphs of of total property damage and total crop damage affected by these severe weather events.

```
propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, geom = "bar", b
inwidth = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Prop
erty Damage in US dollars")+
    xlab("Severe Weather Type") + ggtitle("Total Property Damage by\n Severe Weather Ev
ents in\n the U.S. from 1995 - 2011")

cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, geom = "bar", binwidth = 1)
+
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Crop
Damage in US dollars") +
    xlab("Severe Weather Type") + ggtitle("Total Crop Damage by \nSevere Weather Events
in\n the U.S. from 1995 - 2011")
grid.arrange(propertyPlot, cropPlot, ncol = 2)
```





Based on the above histograms, we find that Flood and hurricane/typhoon cause most property damage. Furthermore, drought and flood cause most crop damage in the U.S. from 1995 to 2011.

Conclusion

The results of this analysis show that convection and extreme temperatures are the weather events that have the greatest impact on public health. Whereas the extremely harmful events for economy are convection and flood.