Untitled

gertlae

Saturday, September 26, 2015

# Impact of servere weather events on population health and the economy in the US

## Synopsis

This report aims to describe the impact of severe weather events in the US on population health and the economy by answering two questions. Across the United States, which types of events are most harmful with respect to population health? Across the United States, which types of events have the greatest economic consequences? Underlying analysis helps to anwser these two questions by exploring the data of the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database.  
Impact on population health is assessed by a Pareto analsysis on the weather event types with highest injury and fatality rates in this database.  
Economical impact is assessed by a Pareto analsyis on the weather event types causing the higest total cost for property and crop damage.

## Data Processing

We record the platform information on which this analysis is created.

sessionInfo()

## R version 3.1.2 (2014-10-31)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
##   
## locale:  
## [1] LC\_COLLATE=Dutch\_Belgium.1252 LC\_CTYPE=Dutch\_Belgium.1252   
## [3] LC\_MONETARY=Dutch\_Belgium.1252 LC\_NUMERIC=C   
## [5] LC\_TIME=Dutch\_Belgium.1252   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] digest\_0.6.8 evaluate\_0.7.2 formatR\_1.2 htmltools\_0.2.6  
## [5] knitr\_1.11 magrittr\_1.5 rmarkdown\_0.8 stringi\_0.5-5   
## [9] stringr\_1.0.0 tools\_3.1.2 yaml\_2.1.13

We download the file.

fileurl = "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
download.file(fileurl, destfile = "StormData.csv.bz2")  
date()

## [1] "Sat Sep 26 01:49:22 2015"

Then we unzip it and save the unzipped file.

library(R.utils)  
bunzip2("StormData.csv.bz2","StormData.csv",overwrite=TRUE, remove=FALSE)

We load the data set and convert it to a dplyr table for easier processing.

library(dplyr)  
StormData <- read.csv("StormData.csv", header = TRUE)  
StormData <- tbl\_df(StormData)

We check the amount of obervations loaded.

dim(StormData)

## [1] 902297 37

We will assess the impact on population heatlh by looking at the fatalities and injuries variables per event type over time across the US. From [www.ncdc.noaa.gov/stormevents/details.jsp](http://www.ncdc.noaa.gov/stormevents/details.jsp) we know that only from 1996 to present, all the 48 event types are recorded as defined in NWS Directive 10-1605. As the purpose is to compare between event types, we filter out all data that is from before 1996.

First convert to R date format and extract the year.

StormData$Year = strptime(StormData$BGN\_DATE, format ="%m/%d/%Y %H:%M:%S")  
StormData$Year = as.numeric(format(StormData$Year,'%Y'))  
table(StormData$Year)

##   
## 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961   
## 223 269 272 492 609 1413 1703 2184 2213 1813 1945 2246   
## 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973   
## 2389 1968 2348 2855 2388 2688 3312 2926 3215 3471 2168 4463   
## 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985   
## 5386 4975 3768 3728 3657 4279 6146 4517 7132 8322 7335 7979   
## 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997   
## 8726 7367 7257 10410 10946 12522 13534 12607 20631 27970 32270 28680   
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009   
## 38128 31289 34471 34962 36293 39752 39363 39184 44034 43289 55663 45817   
## 2010 2011   
## 48161 62174

Then keep only the data after 1995

StormData = StormData[StormData$Year > 1995,]  
table(StormData$Year)

##   
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007   
## 32270 28680 38128 31289 34471 34962 36293 39752 39363 39184 44034 43289   
## 2008 2009 2010 2011   
## 55663 45817 48161 62174

For the economical impact we need the amounts taking into consideration the exponent information.

summary(StormData$PROPDMGEXP)

## - ? + 0 1 2 3 4 5   
## 276185 0 0 0 1 0 0 0 0 0   
## 6 7 8 B h H K m M   
## 0 0 0 32 0 0 369938 0 7374

summary(StormData$CROPDMGEXP)

## ? 0 2 B k K m M   
## 373069 0 0 0 4 0 278686 0 1771

We will only convert the K (kilo) , M (million) and B(billion) exponents and leave other as is.

StormData$PROPDMG = ifelse(StormData$PROPDMGEXP == "K",  
 StormData$PROPDMG \* 10^3,  
 ifelse( StormData$PROPDMGEXP == "M",  
 StormData$PROPDMG \* 10^6,  
 ifelse (StormData$PROPDMGEXP== "B",  
 StormData$PROPDMG \* 10^9,  
 StormData$PROPDMG  
 )  
 )  
 )  
StormData$CROPDMG = ifelse(StormData$CROPDMGEXP == "K",  
 StormData$CROPDMG \* 10^3,  
 ifelse( StormData$CROPDMGEXP == "M",  
 StormData$CROPDMG \* 10^6,  
 ifelse( StormData$CROPDMGEXP == "B",  
 StormData$CROPDMG \* 10^9,  
 StormData$CROPDMG  
 )  
 )  
 )

## Results

### The health impact.

We group the data per event type across the US and over time and we calculate the total of the injuries and fatalities variables.

StormData$BGN\_DATE = as.character(StormData$BGN\_DATE) ## to avoid POSIX error in dplyr group\_by  
HealthImpact = StormData %>%  
 group\_by(EVTYPE) %>%  
 summarise(   
 totFatal = sum(FATALITIES, na.rm = T ),  
 totInjur = sum(INJURIES, na.rm = T)  
 )

Using Pareto analysis to find the most harmful event types by looking at the Fatalities and Injuries respectively. Preparing the vectors to create the Pareto chart

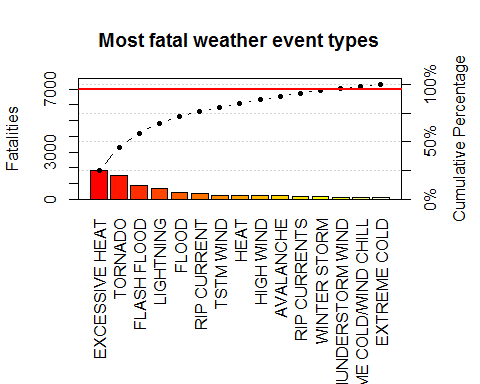
HI\_F = HealthImpact$totFatal  
HI\_I = HealthImpact$totInjur  
names(HI\_F) = HealthImpact$EVTYPE  
names(HI\_I) = HealthImpact$EVTYPE

Creating the Pareto charts for the two vriables focusing on most important event types. The red lines are at 80% of the total cumulative impact whereas the Pareto charts and tables are on the top 15 for readability.

Top\_HI\_F = head(sort(HI\_F, decreasing=TRUE), 15)  
Top\_HI\_I = head(sort(HI\_I, decreasing=TRUE), 15)  
library(qcc)  
pareto.chart(Top\_HI\_F, ylab = "Fatalities", main = "Most fatal weather event types")

##   
## Pareto chart analysis for Top\_HI\_F  
## Frequency Cum.Freq. Percentage Cum.Percent.  
## EXCESSIVE HEAT 1797 1797 24.626559 24.62656  
## TORNADO 1511 3308 20.707140 45.33370  
## FLASH FLOOD 887 4195 12.155680 57.48938  
## LIGHTNING 651 4846 8.921475 66.41085  
## FLOOD 414 5260 5.673564 72.08442  
## RIP CURRENT 340 5600 4.659449 76.74387  
## TSTM WIND 241 5841 3.302727 80.04659  
## HEAT 237 6078 3.247910 83.29450  
## HIGH WIND 235 6313 3.220502 86.51501  
## AVALANCHE 223 6536 3.056050 89.57106  
## RIP CURRENTS 202 6738 2.768261 92.33932  
## WINTER STORM 191 6929 2.617514 94.95683  
## THUNDERSTORM WIND 130 7059 1.781554 96.73839  
## EXTREME COLD/WIND CHILL 125 7184 1.713033 98.45142  
## EXTREME COLD 113 7297 1.548582 100.00000

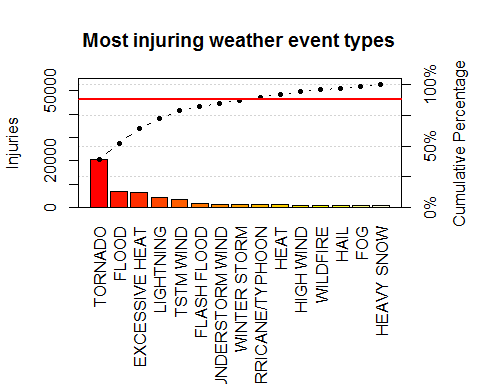
abline(h=(sum(HI\_F)\*.8),col="red",lwd=2)



pareto.chart(Top\_HI\_I, ylab = "Injuries", main = "Most injuring weather event types")

##   
## Pareto chart analysis for Top\_HI\_I  
## Frequency Cum.Freq. Percentage Cum.Percent.  
## TORNADO 20667 20667 39.316288 39.31629  
## FLOOD 6758 27425 12.856219 52.17251  
## EXCESSIVE HEAT 6391 33816 12.158049 64.33056  
## LIGHTNING 4141 37957 7.877716 72.20827  
## TSTM WIND 3629 41586 6.903702 79.11197  
## FLASH FLOOD 1674 43260 3.184568 82.29654  
## THUNDERSTORM WIND 1400 44660 2.663318 84.95986  
## WINTER STORM 1292 45952 2.457862 87.41772  
## HURRICANE/TYPHOON 1275 47227 2.425522 89.84324  
## HEAT 1222 48449 2.324697 92.16794  
## HIGH WIND 1083 49532 2.060267 94.22821  
## WILDFIRE 911 50443 1.733059 95.96127  
## HAIL 713 51156 1.356390 97.31766  
## FOG 712 51868 1.354488 98.67215  
## HEAVY SNOW 698 52566 1.327855 100.00000

abline(h=(sum(HI\_I)\*.8),col="red",lwd=2)



### The economical impact

We group the data per event type across the US and over time and we calculate the total dollar value of damage per event type to properties and crop. Inflation is not taken into consideration in this analysis.

EconImp = StormData %>%  
 group\_by(EVTYPE) %>%  
 summarise(totDMG = sum(PROPDMG) + sum(CROPDMG))

Using Pareto analysis to find the event types with highest economical impact in terms of total damage cost. Preparing the vector to create the pareto chart, and going to million USD values.

EI\_D = EconImp$totDMG / 10^6  
names(EI\_D) = EconImp$EVTYPE

Creating the Pareto chart for the economical impact. The red line is at 80% of the total cumulative impact whereas the Pareto chart and table are on the top 15 for readability.

Top\_EI\_D = head(sort(EI\_D, decreasing=TRUE), 15)  
pareto.chart(Top\_EI\_D, ylab = "Damage (in MUSD)", main = "Highest damage cost weather event types")

##   
## Pareto chart analysis for Top\_EI\_D  
## Frequency Cum.Freq. Percentage Cum.Percent.  
## FLOOD 148919.612 148919.6 38.3920265 38.39203  
## HURRICANE/TYPHOON 71913.713 220833.3 18.5396210 56.93165  
## STORM SURGE 43193.541 264026.9 11.1354545 68.06710  
## TORNADO 24900.371 288927.2 6.4194076 74.48651  
## HAIL 17071.173 305998.4 4.4010115 78.88752  
## FLASH FLOOD 16557.106 322555.5 4.2684830 83.15600  
## HURRICANE 14554.229 337109.7 3.7521340 86.90814  
## DROUGHT 14413.667 351523.4 3.7158966 90.62403  
## TROPICAL STORM 8320.187 359843.6 2.1449748 92.76901  
## HIGH WIND 5881.422 365725.0 1.5162522 94.28526  
## WILDFIRE 5054.140 370779.2 1.3029759 95.58824  
## TSTM WIND 5031.942 375811.1 1.2972532 96.88549  
## STORM SURGE/TIDE 4642.038 380453.1 1.1967346 98.08223  
## THUNDERSTORM WIND 3780.985 384234.1 0.9747520 99.05698  
## ICE STORM 3657.909 387892.0 0.9430224 100.00000

abline(h=(sum(EI\_D)\*.8),col="red",lwd=2)

