## Reproducible Research: Course Project 2

## Author: Shayan (Sean) Taheri

## Impact of Severe Weather Events on Public Health and Economy in the United States

### Synonpsis

Our goal in this project is analysis of the impact of different weather events on public health and economy based on the storm database cllected from U.S. National Occeanic and Atmospheric Administration’s (NOAA) from 1950 - 2011.

The estimates of fatalities, injuries, property, and crop damage are used in cropping damage to harmful to the population healthy and economy. From these data, we found that excessive heat and tornado are most harmful with respect to population health, while flood, drought, and hurricane/typhoon have the greatest economic consequences.

### 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 3.5.3

## Loading required package: R.oo

## Loading required package: R.methodsS3

## R.methodsS3 v1.7.1 (2016-02-15) successfully loaded. See ?R.methodsS3 for help.

## R.oo v1.22.0 (2018-04-21) 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 v2.8.0 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 3.5.3

library(plyr)  
require(gridExtra)

## Loading required package: gridExtra

### Data Processing

Let’s download the data file and uncompress it in the first step.

getwd()

## [1] "C:/Users/shaya/Desktop/Reprocible\_Research\_Course\_Project\_2"

setwd("C:/Users/shaya/Desktop/Reprocible\_Research\_Course\_Project\_2")  
  
if (!"stormData.csv.bz2" %in% dir("./data/")) {  
 print("hhhh")  
 download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile = "data/stormData.csv.bz2")  
 bunzip2("data/stormData.csv.bz2", overwrite=T, remove=F)  
}

Then, we read the generated csv file. If the data already exists in the working environment, we do not need to load it again. Otherwise, we read the csv file.

Next, the generated CSV file is read if it does not exist in the working environment.

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

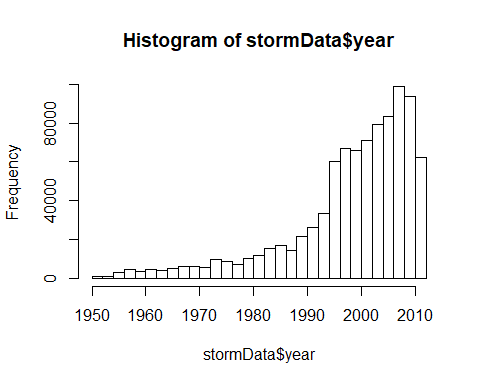
## [1] 902297 37

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 database start is in the year of 1950 and ends in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

if (dim(stormData)[2] == 37) {  
 stormData$year <- as.numeric(format(as.Date(stormData$BGN\_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"))  
}  
hist(stormData$year, breaks = 30)



Based on the delievered histogram, it can be seen that the number of events tracked starts to increase significantly specifically 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.

#### Impact on Public Health

The number of **fatalities** and **injuries** are reviewd.They are caused by the sever weather events. We would like to get the first 15 most severe types of 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)

#### Impact on Economy

The **property damage** and **crop damage** data are converted into comparable numerical forms. This conversion is in accordance with the meaning of units described in the code book ([Storm Events](http://ire.org/nicar/database-library/databases/storm-events/)).

Both PROPDMGEXP and CROPDMGEXP columns record a multiplier for each observation where we have Hundred (H), Thousand (K), Million (M) and Billion (B).

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  
## introduced by coercion

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

## Warning in convertHelper(storm, "CROPDMGEXP", "cropDamage"): NAs introduced  
## by coercion

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

As for the impact on public health, we have got two sorted lists of severe weather events below by the number of people badly affected.

fatalities

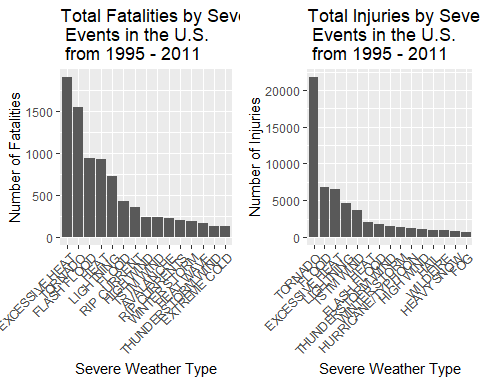
## EVTYPE FATALITIES  
## 1 EXCESSIVE HEAT 1903  
## 2 TORNADO 1545  
## 3 FLASH FLOOD 934  
## 4 HEAT 924  
## 5 LIGHTNING 729  
## 6 FLOOD 423  
## 7 RIP CURRENT 360  
## 8 HIGH WIND 241  
## 9 TSTM WIND 241  
## 10 AVALANCHE 223  
## 11 RIP CURRENTS 204  
## 12 WINTER STORM 195  
## 13 HEAT WAVE 161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

A pair of graph of total fatalities nand total injuries affected by these severe weather events is shown in the following.

fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", binwidth = NULL) +  
 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 = NULL) +   
 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)



The histogram represented in above demonstrates the **excessive heat** and **tornado** cause most fatalities; **excessive heat** and **tornado** cause most fatalities; **tornato** causes most injuries in the United States from 1995 to 2011.

Two sorted lists by the amount of money cost by damages are provided in order to show the impact on economy.

property

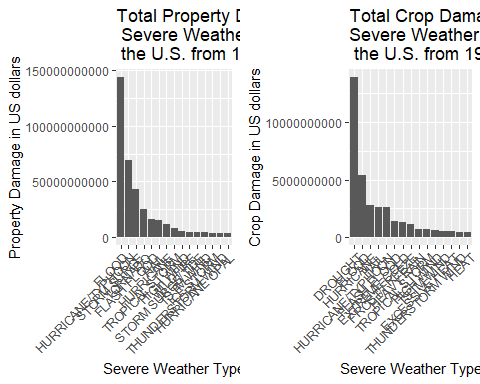
## EVTYPE propertyDamage  
## 1 FLOOD 144022037057  
## 2 HURRICANE/TYPHOON 69305840000  
## 3 STORM SURGE 43193536000  
## 4 TORNADO 24935939545  
## 5 FLASH FLOOD 16047794571  
## 6 HAIL 15048722103  
## 7 HURRICANE 11812819010  
## 8 TROPICAL STORM 7653335550  
## 9 HIGH WIND 5259785375  
## 10 WILDFIRE 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

A pair of graphs of total property damage and total crop damage affected by these severe weather events is given in the following.

propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, geom = "bar", binwidth = NULL) +   
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + scale\_y\_continuous("Property Damage in US dollars")+   
 xlab("Severe Weather Type") + ggtitle("Total Property Damage by\n Severe Weather Events in\n the U.S. from 1995 - 2011")  
  
cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, geom = "bar", binwidth = NULL) +   
 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)



For the goal of finding out the **flood** and **hurricane/typhoon** cause most property damage; **drought** and **flood** causes most crop damage in the United States from 1995 to 2011, two histogram plots are created and displayed side-by-side.

### Conclusion

The interpretation from this study is described as **excessive heat** and **tornado** are most harmful with respect to population health, while **flood**, **drought**, and **hurricane/typhoon** have the greatest economic consequences.