Severe Weather Envents and Their Impact on Public Health and Economy in the U.S.A

Ron Ferens

July 25, 2015

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

Storm Database is an official publication of the National Oceanic and Atmospheric Administration (NOAA) which tracks characteristics of major storms and weather events in the United States.

In this report, based on the *Storm Database*, we will identify the severe weather events which have the most most harmful with respect to population health (fatalities and injuries) and events which have the greatest economic consequences (damage estimation)

Data Processing

Loading the Data

From the U.S. National Oceanic and Atmospheric Administration's (NOAA), we obtained the Storm Database (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2). This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

```
setwd("C:/Users/rferens/Git/datasciencecoursera/Reproducible.Research/Project.02/")
dataDir <- "./Data"
destFileZip <- paste(dataDir, sep="/", "repdata-data-StormData.csv.bz2")

if(!file.exists(destFileZip)) {
  url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(url, dest=destFileZip, mode="wb")
}

stormData <- read.csv(destFileZip)
```

Examining the Data

Once the data have been loaded properly, we can now examine the column headers and make sure that they are properly formated for R data frames:

```
names(stormData)
```

```
##  [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"
```

The main goal of this report is to explore the NOAA *Storm Database*, and evaluate which types of events are most harmful with respect to population health and which types of events have the greatest economic consequences. Hence, the columns we are interested in are:

1. **FATALITIES** and **INJURIES** - Those columns contain the direct causes of weather-related fatalities or injuries.
2. **PROPDMG** and **CROPDMG** - Those columns contain the damage estimates in actual dollar amounts. Those columns refer to property and crop related damage estimations.

Results

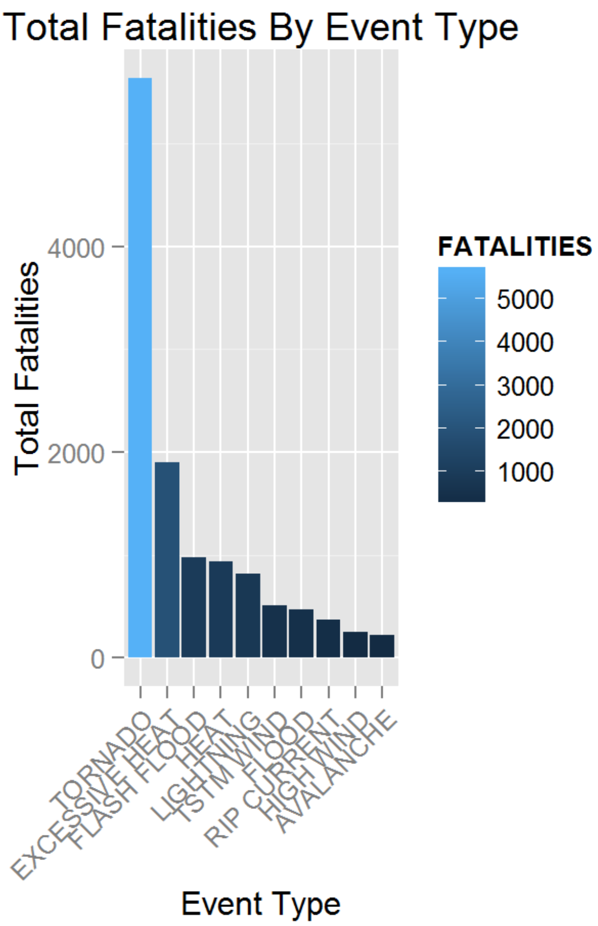
Which types of events are most harmful with respect to population health?

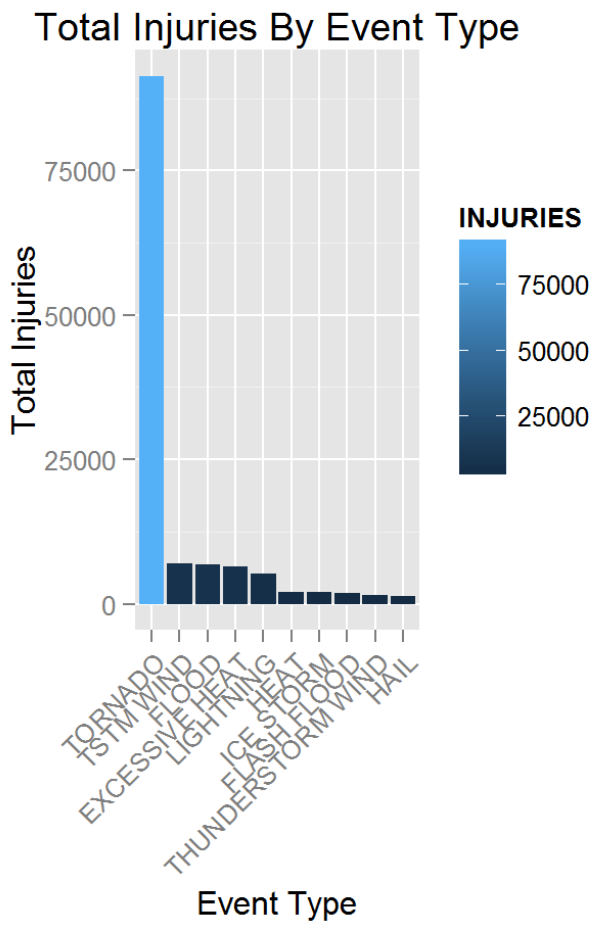
In order to find the most harmful events with respect to population health, we will first aggragate the total number of fatalities and injuries for each event type. Once calculated, we will extract the top ten event types with the largest number of fatalities and injuries:

```
sortEvntTypeByColumn <- function(dataset, colName, threhold) {
  colIndex <-  which(colnames(dataset) == colName)
  aggData <- aggregate(dataset[,colIndex], by=list(dataset$EVTYPE), sum, na.rm=TRUE)
  colnames(aggData) <- c("EVTYPE", colName)
  aggData <- aggData[order(aggData[[colName]], decreasing = T),]
  return(aggData[1:threhold,])
}

evntByFatalities <- sortEvntTypeByColumn(stormData, "FATALITIES", 10)
evntByInjuries <- sortEvntTypeByColumn(stormData, "INJURIES", 10)
```

The graphs below list the event types with the largest number of fatalities (right graph) and injuries (left graph):

```
library(ggplot2)
library(gridExtra)
evntByFatalitiesPlot <- ggplot( data = evntByFatalities,
                                aes(x=reorder(EVTYPE, -FATALITIES), y=FATALITIES, fil
l=FATALITIES)) +
  geom_bar(stat="identity") +
  labs(x="Event Type", y="Total Fatalities", title="Total Fatalities By Event Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

evntByInjuriesPlot <- ggplot(data = evntByInjuries,
                                aes(x=reorder(EVTYPE, -INJURIES), y=INJURIES, fill=INJU
RIES)) +
  geom_bar(stat="identity") +
  labs(x="Event Type", y="Total Injuries", title="Total Injuries By Event Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

grid.arrange(evntByFatalitiesPlot, evntByInjuriesPlot, ncol=2, nrow=1)
```





As seen above, **tornados** cause the most number of fatalities and injuries among all listed event types in the United State from 1950-2011.

Which types of events have the greatest economic consequences?

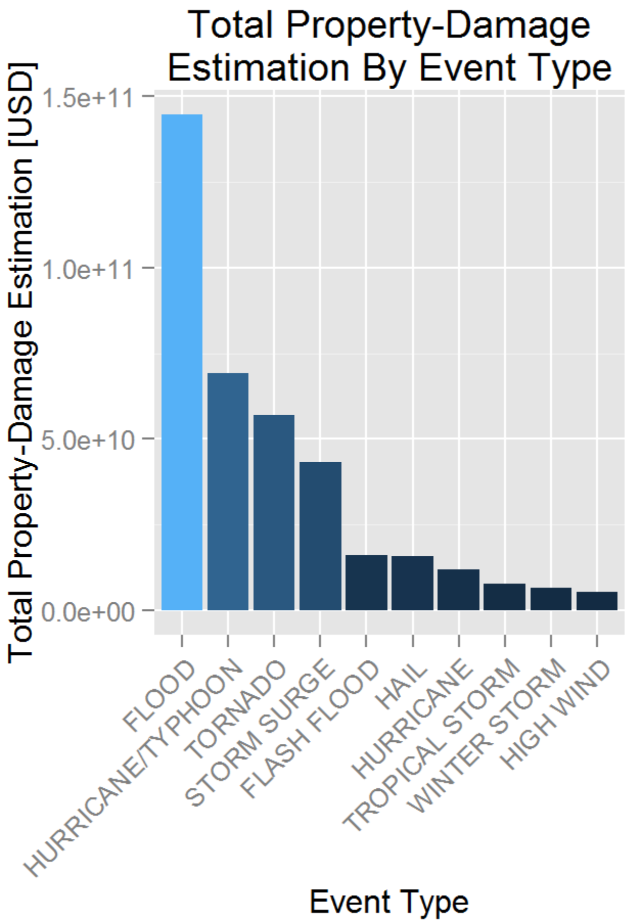
In order to find the event types which have the greatest economic consequences, we will first aggragate the total cost of property and crop damage for each event type. Once calculated, we will extract the top ten event types which cause most property and crop damages.

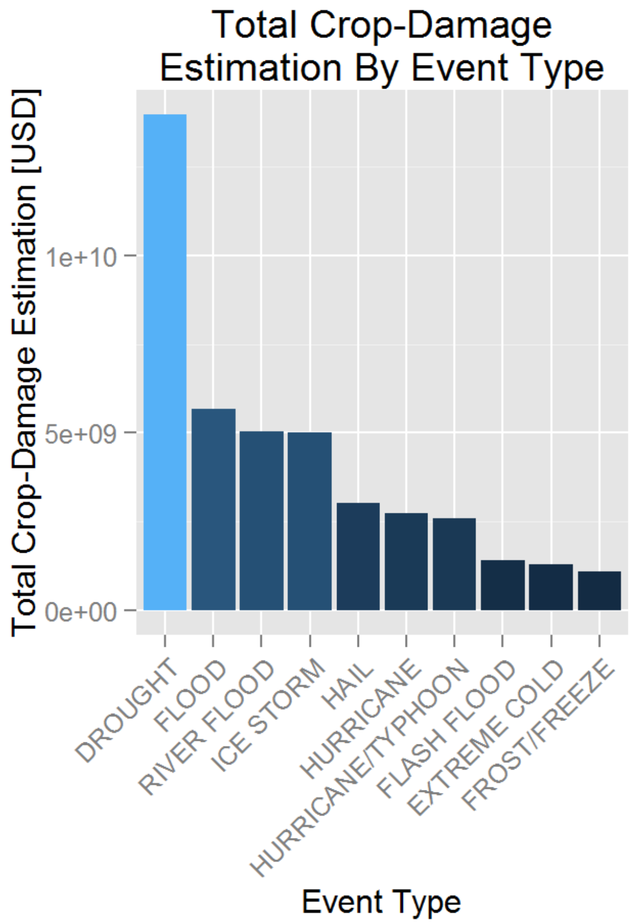
The storm data's damage estimation is represented in two seperate fields. The first indicates the damage value in USD and the second is the damage magnatude (Hundreds, Thousands, Millions, Billions). Hence, as a pre-processing step, we will first have to convert the damage estimation values (in the database) into total cost of damage. This process will be done for both property-damage and crop-damage:

```
damageEstimation <- function(damageValue, damageMagnitude) {
  if(damageMagnitude %in% c('h','H'))
    damageEstimation <- damageValue * (10 ** 2)
  else if(damageMagnitude == 'K')
    damageEstimation <- damageValue * (10 ** 3)
  else if(damageMagnitude %in% c('m','M'))
    damageEstimation <- damageValue * (10 ** 6)
  else if(damageMagnitude == 'B')
    damageEstimation <- damageValue * (10 ** 9)
  else
    damageEstimation <- damageValue
  return(damageEstimation)
}

stormData$PROPDMG_VAL <- mapply(damageEstimation, stormData$PROPDMG, stormData$PROPDMGE
XP)
evntByPropDmg <- sortEvntTypeByColumn(stormData, "PROPDMG_VAL", 10)

stormData$CROPDMG_VAL <- mapply(damageEstimation, stormData$CROPDMG, stormData$CROPDMGE
XP)
evntByCropDmg <- sortEvntTypeByColumn(stormData, "CROPDMG_VAL", 10)
```

The graphs below list the event types with with the largest property-damage estimation (right graph) and crop-damage extimation (left graph):

```
evntByPropDmgPlot <- ggplot(data = evntByPropDmg,
                            aes(x=reorder(EVTYPE, -PROPDMG_VAL), y=PROPDMG_VAL, fill=PR
OPDMG_VAL)) +
  geom_bar(stat="identity") +
  labs(x="Event Type", y="Total Property-Damage Estimation [USD]", title="Total Propert
y-Damage\nEstimation By Event Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position="none")

evntByCropDmgPlot <- ggplot(data = evntByCropDmg,
                            aes(x=reorder(EVTYPE, -CROPDMG_VAL), y=CROPDMG_VAL, fill=CR
OPDMG_VAL)) +
  geom_bar(stat="identity") +
  labs(x="Event Type", y="Total Crop-Damage Estimation [USD]", title="Total Crop-Damag
e\nEstimation By Event Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position="none")

grid.arrange(evntByPropDmgPlot, evntByCropDmgPlot, ncol=2, nrow=1)
```





As seen above, **flood** is the event type that causes most property-damage and **drought** causes the most crop-damage in the United State from 1950-2011.