

Impacts of weather events in the US

Sunday, December 21, 2014

1 Synopsis

The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm 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. This project analyze the impacts of those events and specify which event has the greatest damage.

The analysis start from get the row data from the internet, and then clean the data by grouping the weather events to **Convection, Flood, Winter, Extreme Temperatures and Others** groups. By using the tidy data, several descriptive graphs are made for explaining the impacts of the weather events.

The **Flood**-related events have the greatest damages in the economics point of view, while the **Convection**-related events threat people's health the most.

2 Data processing

2.1 Getting The Raw Data

We can get the raw data from the url on Coursera's website. First of all, we get the raw data for further analysis.

- Download the data.
- Unzip the data.
- Load the data to R and change it to *data.table* class.

```
library(R.utils)
library(data.table)
library(reshape2)
library(ggplot2)

## Getting Data
# Download data from url
data.url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
data.bz.file <- "G:/Coursera/Data Science/Github/RawData/repdata-data-StormData.csv.bz2"
download.file(data.url, data.bz.file, mode="wb")
# unzip the file
bunzip2(data.file)
data.csv.file <- "G:/Coursera/Data Science/Github/RawData/repdata-data-StormData.csv"
# read the file to R and change it to data.table format
rawdata <- read.csv(data.csv.file, stringsAsFactors=FALSE)
rawdata <- data.table(rawdata)
dim(rawdata)
```

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

2.2 Cleaning The Data

Since the raw data is hard to use, the following procedures are applied for cleaning and grouping it in a friendly-using way.

- Changing the names in a descriptive way.
- Normalize the unit of damage amount in one million of dollars.
- Group the event types to 5 groups: *Convection, Flood, Winter, Extreme Temperatures and Others*.
- Delete the redundant variables.

```
## Cleaning Data
# change name to lower case
setnames(rawdata,tolower(names(rawdata)))
## Set correct damage amount (in million)
DT <- rawdata[,.(state,evtype,fatalities,injuries,propdmg,propdmgexp,cropdmg,cropdmgexp)]
DT[grepl("[Bb]",propdmgexp),propdmg:=propdmg*1e9*1e-6]
DT[grepl("[Mm]",propdmgexp),propdmg:=propdmg*1e6*1e-6]
DT[grepl("[Kk]",propdmgexp),propdmg:=propdmg*1e3*1e-6]
DT[grepl("[Hh]",propdmgexp),propdmg:=propdmg*1e2*1e-6]

DT[grepl("[Bb]",cropdmgexp),cropdmg:=cropdmg*1e9*1e-6]
DT[grepl("[Mm]",cropdmgexp),cropdmg:=cropdmg*1e6*1e-6]
DT[grepl("[Kk]",cropdmgexp),cropdmg:=cropdmg*1e3*1e-6]

DT[,':='(propdmgexp=NULL,cropdmgexp=NULL)]
setnames(DT,c("State","EventType","Fatalities","Injuries","PropertyDamage","CropDamage"))
## Create total damage variable
DT[,TotalDamage:=PropertyDamage+CropDamage]

## Event Types
# Convection
Key.Convection <- c("TORNADO", "FUNNEL", "TSTM", "LIGHTNING",
                   "LIGHTING", "THUNDERSTORM", "WIND", "HAIL")
regex.Convection <- paste(Key.Convection, collapse = "|")
index.Convection <- grep(regex.Convection,DT$EventType)
DT[index.Convection,EventType:="Convection"]
# Extreme Temperatures
Key.ExTemperature <- c("COLD", "HEAT", "WARM", "COOL")
regex.ExTemperature <- paste(Key.ExTemperature, collapse = "|")
index.ExTemperature <- grep(regex.ExTemperature,DT$EventType)
index.ExTemperature <- setdiff(index.ExTemperature,index.Convection)
DT[index.ExTemperature,EventType:="Extreme Temperatures"]
# Flood
Key.Flood <- c("FLOOD", "RAIN", "PRECIP", "SHOWER")
regex.Flood <- paste(Key.Flood, collapse = "|")
index.Flood <- grep(regex.Flood,DT$EventType)
index.Flood <- setdiff(index.Flood,
                      unique(c(index.Convection,index.ExTemperature)))
DT[index.Flood,EventType:="Flood"]
# Winter
Key.Winter <- c("SNOW", "ICE", "ICY", "FREEZ", "WINT")
regex.Winter <- paste(Key.Winter, collapse = "|")
index.Winter <- grep(regex.Winter,DT$EventType)
```

```

index.Winter <- setdiff(index.Winter,
                        unique(c(index.Convection,index.ExTemperature,index.Flood)))
DT[index.Winter,EventType:="Winter"]
# Other
index.Other <- setdiff(1:nrow(DT),unique(c(index.Convection,index.ExTemperature,
                                           index.Flood,index.Winter)))
DT[index.Other,EventType:="Other"]

```

DT

```

##      State  EventType Fatalities Injuries PropertyDamage CropDamage
##    1:    AL Convection         0      15         0.0250         0
##    2:    AL Convection         0         0         0.0025         0
##    3:    AL Convection         0         2         0.0250         0
##    4:    AL Convection         0         2         0.0025         0
##    5:    AL Convection         0         2         0.0025         0
##    ---
## 902293:   WY Convection         0         0         0.0000         0
## 902294:   MT Convection         0         0         0.0000         0
## 902295:   AK Convection         0         0         0.0000         0
## 902296:   AK      Other         0         0         0.0000         0
## 902297:   AL      Winter         0         0         0.0000         0
##      TotalDamage
##    1:         0.0250
##    2:         0.0025
##    3:         0.0250
##    4:         0.0025
##    5:         0.0025
##    ---
## 902293:         0.0000
## 902294:         0.0000
## 902295:         0.0000
## 902296:         0.0000
## 902297:         0.0000

```

3 Results

3.1 Harms of Events on Population Health

In this section, the analysis of impacts of weather on population health is produced.

```

## events -- population health
DT.health <- DT[,.(Fatalities=sum(Fatalities),
                        Injuries=sum(Injuries)),
                by=EventType]
DT.health

```

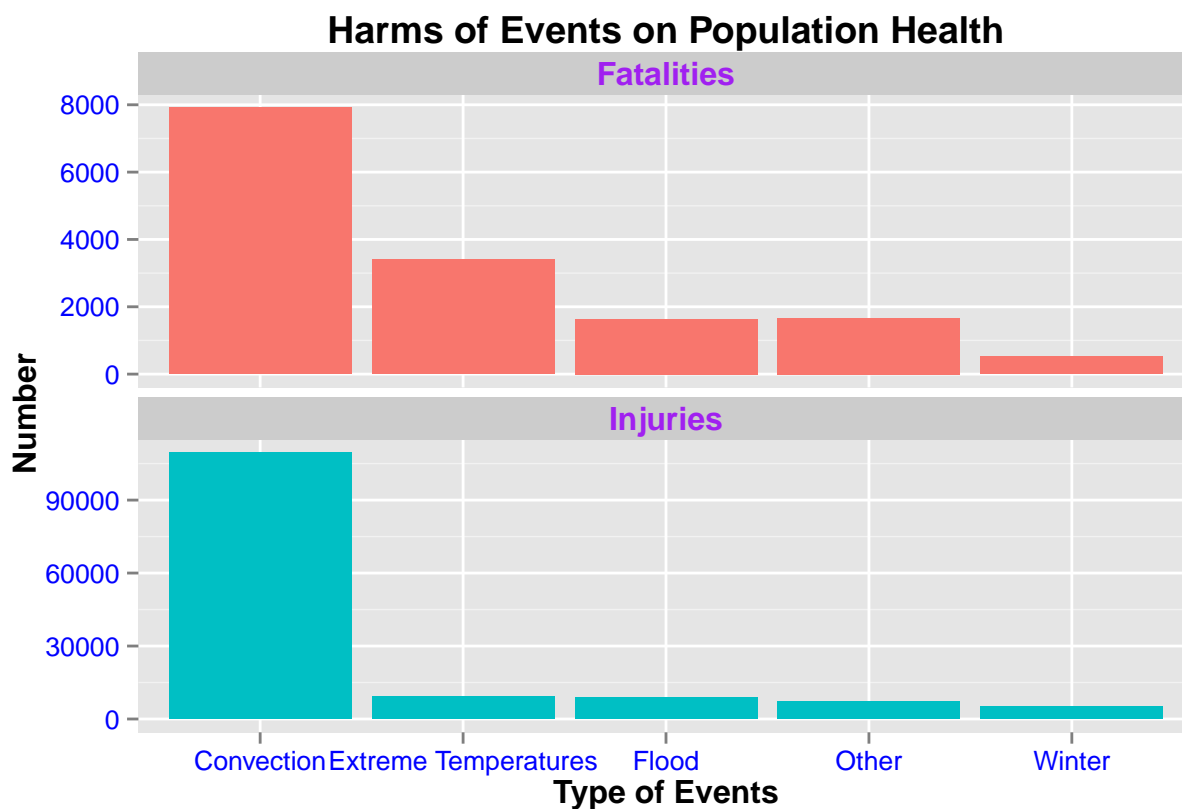
```

##      EventType Fatalities Injuries
## 1:    Convection      7915  109547
## 2:      Flood      1637    8930
## 3:    Winter       525    5267

```

```
## 4: Extreme Temperatures      3400      9453
## 5:                          Other      1668      7331
```

```
DT.health.melt <- melt(DT.health,measure=c("Fatalities","Injuries"),id="EventType")
ggplot(data=DT.health.melt) +
  geom_bar(aes(x=EventType,y=value,fill=variable),stat="identity",position="dodge") +
  facet_wrap(~variable, scales = "free_y",nrow=2) +
  labs(title="Harms of Events on Population Health", x= "Type of Events", y="Number") +
  theme(legend.position="none",
        strip.text=element_text(colour="purple",size = 12,face="bold"),
        title=element_text(size = 12,face = "bold"),
        axis.text=element_text(colour="blue"))
```



The **Convection**-related events threat people's health the most.

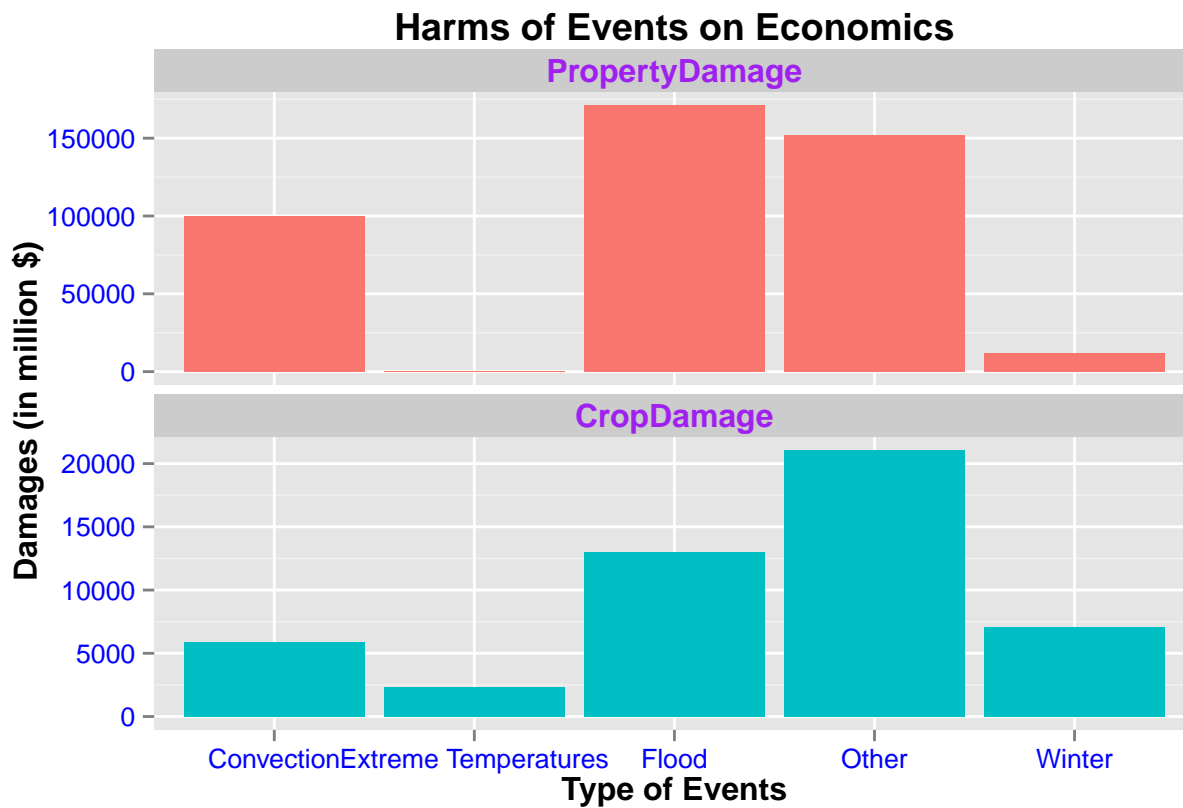
3.2 Harms of Events on Economics

In this section, the analysis of impacts of weather on economics is produced.

```
## events -- economic consequences
DT.eco <- DT[,.(PropertyDamage=sum(PropertyDamage),
                                   CropDamage=sum(CropDamage)),by=EventType]
DT.eco
```

```
##           EventType PropertyDamage CropDamage
## 1:      Convection    100204.1741    5922.450
## 2:        Flood     171213.1560   13043.892
## 3:        Winter     11772.9375    7043.552
## 4: Extreme Temperatures      145.6131    2293.495
## 5:           Other    152154.4741   21071.802
```

```
DT.eco.melt <- melt(DT.eco,measure=c("PropertyDamage","CropDamage"),id="EventType")
ggplot(data=DT.eco.melt) +
  geom_bar(aes(x=EventType,y=value,fill=variable),stat="identity",position="dodge") +
  facet_wrap(~variable, scales = "free_y",nrow=2) +
  labs(title="Harms of Events on Economics", x= "Type of Events", y="Damages (in million $)") +
  theme(legend.position="none",
        strip.text=element_text(colour="purple",size = 12,face="bold"),
        title=element_text(size = 12,face = "bold"),
        axis.text=element_text(colour="blue"))
```



The **Flood**-related events have the greatest damages in the economics point of view.