Reproducible Research: Peer Assessment2

Created by Trent Lin sep.21 2014

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

Synonpsis

In this project we aim to analyze the impact of differenct weather events on public health and economy damage based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) from 1950 - 2011. We will use fatalities, injuries, property and crop damage to decide which types of event are most harmful to the population health and economy. We found that convection(Tornado, Wind) are the most harmful to people health. Storm(THUNDERSTORM, Winter STORM) and Winter(ICE, HAIL, SNOW) are the most harmful to economics.

Basic Setting

```
echo = TRUE # always make code visible
library(ggplot2)
library(plyr)
```

Data Processing

Load Data and convert data type

```
StormData <- read.csv("repdata-data-StormData.csv.bz2",header= TRUE,sep=",")
```

```
StormData$EVTYPE <- as.character(StormData$EVTYPE)
StormData$PROPDMGEXP <- as.character(StormData$PROPDMGEXP)
StormData$CROPDMGEXP <- as.character(StormData$CROPDMGEXP)
StormData$FATALITIES <- as.numeric(as.character(StormData$FATALITIES))
```

```
## Warning: 強制變更過程中產生了 NA
```

```
StormData$INJURIES <- as.numeric(as.character(StormData$INJURIES))
```

```
StormData$PROPDMG <- as.numeric(as.character(StormData$PROPDMG))
```

```
## Warning: 強制變更過程中產生了 NA
```

```
StormData$CROPDMG <- as.numeric(as.character(StormData$CROPDMG))
```

```
## Warning: 強制變更過程中產生了 NA
```

```
dim(StormData)
```

```
## [1] 692288     37
```

Srink the DataSet

Due to Dataset is too huge and may slow down the process. We exclude the data which fatalities = 0 or injuries = 0 or propdmg = 0 or cropdmg = 0

```
StormData <- StormData[StormData$FATALITIES > 0|StormData$INJURIES > 0|StormData$PROPDMG >0|StormData$CROPDMG >0,]
dim(StormData)
```

```
## [1] 292519     37
```

Classify Dataset

Due to EVTYPE has a lot of error and issue. We use grepl() to classify dataset in convection,cylones, flood,lighting,marine,storm,temperature, winter eight category.

```
convention <- c("?.*TORNADO|TORNDAO|WIND|TSTM?.*")
lighting <- c("?.*LIGHTING|LIGHTN|LIGNTNING?.*")
storm <- c("?.*THUNDERSTORM|STORM?.*")
temperature <-c("?.*COLD|HEAT|WARM|TEMPERATURE|THERMIA?.*")
flood <-c("?.*FLOOD|RISING|STREAM FLD?.*")
marine <-c("?.*COASTAL|TSUNAMI|CURRENT|SWELLS|TIDE|WAVE|SEAS|SURF|HIGH WATER?.*")
cylones <-c("?.*CYCLONE|HURRICANE|TYPHOON?.*")
winter <-c("?.*WINT|ICE|HAIL|AVALAN|SLEET|SNOW|FREEZ|BLIZZ|FROST|GLAZE|MIXED?.*")
StormData[grepl(convention,StormData$EVTYPE,ignore.case=TRUE),"category"]="Convection"
StormData[grepl(lighting,StormData$EVTYPE,ignore.case=TRUE),"category"]="Lighting"
StormData[grepl(storm,StormData$EVTYPE,ignore.case=TRUE),"category"]="Storm"
StormData[grepl(temperature,StormData$EVTYPE,ignore.case=TRUE),"category"]="Temperature"
StormData[grepl(flood,StormData$EVTYPE,ignore.case=TRUE),"category"]="Flood"
StormData[grepl(marine,StormData$EVTYPE,ignore.case=TRUE),"category"]="Marine"
StormData[grepl(cylones,StormData$EVTYPE,ignore.case=TRUE),"category"]="Cylones"
StormData[grepl(winter,StormData$EVTYPE,ignore.case=TRUE),"category"]="Winter"
StormData[is.na(StormData$category),"category"]="Others"
```

Make property damage and crop damage available to calculate

Because there are different unit type in PROPDMGEXP and CROPDMGEXP. We create four new columns PROPDMGCOV, NEWPROPDMG, CROPDMGCOV, NEWCROPDMG for unit conversion and damage calculation.

```
StormData$PROPDMGCOV <- 1
#StormData[grepl("h | H", StormData$PROPDMGEXP, ignore.case = TRUE),]
StormData[grepl("h|H", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 100
StormData[grepl("k|K", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1000
StormData[grepl("m|M", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1e+06
StormData[grepl("b|B", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1e+09
StormData$NEWPROPDMG <- (StormData$PROPDMG)*(StormData$PROPDMGCOV)

StormData$CROPDMGCOV <- 1
StormData[grepl("h|H", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 100
StormData[grepl("k|K", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1000
StormData[grepl("m|M", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1e+06
StormData[grepl("b|B", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1e+09
StormData$NEWCROPDMG <- (StormData$CROPDMG)*(StormData$CROPDMGCOV)
```

Result

Make property damage and crop damage available to calculate

Because there are different unit type in PROPDMGEXP and CROPDMGEXP. We create four new columns PROPDMGCOV, NEWPROPDMG, CROPDMGCOV, NEWCROPDMG for unit conversion and damage calculation.

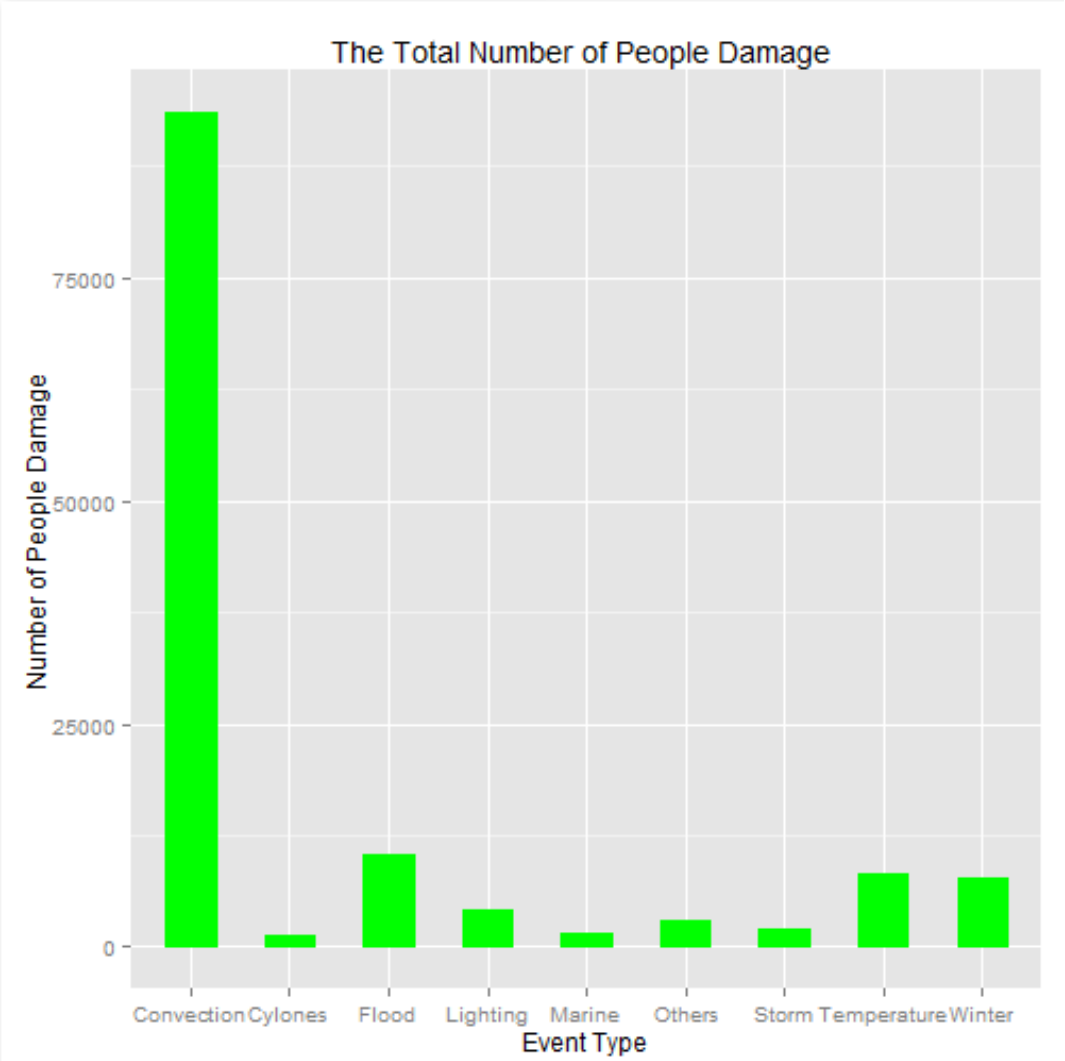
```
StormData$PROPDMGCOV <- 1
#StormData[grepl("h | H", StormData$PROPDMGEXP, ignore.case = TRUE),]
StormData[grepl("h|H", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 100
StormData[grepl("k|K", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1000
StormData[grepl("m|M", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1e+06
StormData[grepl("b|B", StormData$PROPDMGEXP, ignore.case = TRUE),"PROPDMGCOV"] = 1e+09
StormData$NEWPROPDMG <- (StormData$PROPDMG)*(StormData$PROPDMGCOV)

StormData$CROPDMGCOV <- 1
StormData[grepl("h|H", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 100
StormData[grepl("k|K", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1000
StormData[grepl("m|M", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1e+06
StormData[grepl("b|B", StormData$CROPDMGEXP, ignore.case = TRUE),"CROPDMGCOV"] = 1e+09
StormData$NEWCROPDMG <- (StormData$CROPDMG)*(StormData$CROPDMGCOV)
```

Which types of events are most harmful to polulation health ?

we create new column peopledmg to combine fatalitels and injuries. Then draw a picture to show the relationship between weather events and polulation health

```
StormData$peopledmg <- StormData$FATALITIES + StormData$INJURIES
ggplot(StormData, aes(category,peopledmg))+geom_bar(stat = "identity", colour = "green", fill = "green", width = 0.5)+
  labs(title= "The Total Number of People Damage ", x="Event Type", y="Number of People Damage ")
```



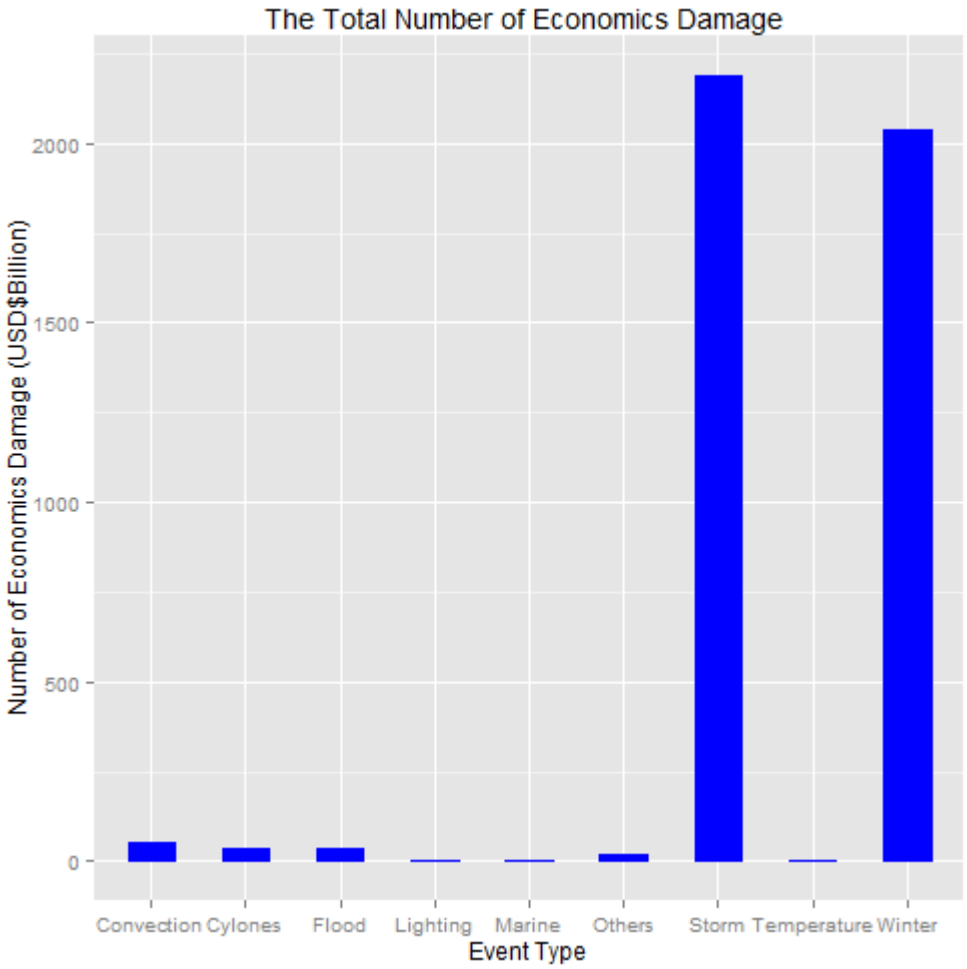
Conclusion1

As the picture we found that convetion(Tornado, Wind) are the most harmful to people health.

Which type of event are most harmful to economics

we create new column economicsdmg to calculate the total economics damage which combine property damage and crop damage. Then draw a picture to showthe relationship between weather events and economics damage.

```
StormData$economicsdmg <- ((StormData$NEWPROPDMG + StormData$NEWCROPDMG)/1e+09)
ggplot(StormData, aes(category,economicsdmg))+geom_bar(stat = "identity", colour = "blue", fill = "blue", width = 0.5)+
  labs(title= "The Total Number of Economics Damage ", x="Event Type", y="Number of Economics Damage (USD$Billion)")
```



Conclusion2

As the picture we found that Storm(THUNDERSTORM, Winter STORM) and Winter(ICE, HAIL, SNOW) are the most harmful to economics.