

Reproducible Research: Peer Assessment 2

By: Huang Yinning

Impact of Severe Weather Events on Population Health and Economy in United States - An analysis based on NOAA Storm Database

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.

This report aims to conduct an exploratory analysis of the impact on population health and economy in United States by severe weather events. Data used in this analysis is taken from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. 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.

In particular, this report aims to address the following questions:

1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

Data Processing

Load required libraries

```
library(dplyr, warn.conflicts=FALSE)
library(ggplot2)
library(reshape2)
library(gridExtra)

## Loading required package: grid

library(R.utils, warn.conflicts=FALSE)

## Loading required package: R.oo
## Loading required package: R.methodsS3
## R.methodsS3 v1.7.0 (2015-02-19) successfully loaded. See ?R.methodsS3 for help.
## R.oo v1.19.0 (2015-02-27) 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 v2.0.0 (2015-02-28) successfully loaded. See ?R.utils for help.
```

Load the data

```
setwd("D:/Dropbox/Coursera/01 Data Science/05 Reproducible
Research/RepData_PeerAssessment2")

if (!"repdata_data_StormData.csv.bz2" %in% dir("./data")) {

download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormDat
a.csv.bz2",
  destfile = "./data/repdata_data_StormData.csv.bz2")
  bunzip2("./data/repdata_data_StormData.csv.bz2", overwrite=T, remove=F)
}

if ("repdata_data_StormData.csv" %in% dir("./data")) {
  storm <- read.csv("./data/repdata_data_StormData.csv")
}
```

Subset data to only contain relevant information

Since there are many variables that are irrelevant to this analysis, we create a subset of the data that only contain information on health and economic damages.

The relevant variables are: * EVTYPE (type of event) * FATALITIES * INJURIES * PROPDMG * PROPDMGEXP * CROPDGMG * CROPDGMGEXP

```
columns <-
c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDGMG", "CROPDGMGEXP")
data <- storm[,columns]
rm(storm)
```

Impact on Economy

We use the extend of property and crop damage as a measure of the impact on economy by severe weather events. We will first process data for property and crop damage

Replace symbols with actual numerical values

In variables "PROPDMGEXP" and "CROPDMGEXP", alphabetical characters are used as a multiplier to signify magnitude of variables "PROPDMG" and "CROPDMG" respectively. The labels used are: "H" for hundred, "K" for thousands, "M" for millions, and "B" for billions. So we replace the labels with its respective exponents where h|H = 2 k|K = 3 m|M = 6 and b|B = 9

For unknown characters, 0 is assigned.

PROPDMGEXP

```
table(data$PROPDMGEXP, exclude=NULL)

##
##           -      ?      +      0      1      2      3      4      5
## 465934      1      8      5     216     25     13      4      4     28
##          6      7      8      B      h      H      K      m      M    <NA>
##          4      5      1     40      1      6 424665      7  11330      0

data$PROPDMGEXP <- as.character(data$PROPDMGEXP)
data$PROPDMGEXP <- gsub("h|H", "2", data$PROPDMGEXP)
data$PROPDMGEXP <- gsub("k|K", "3", data$PROPDMGEXP)
data$PROPDMGEXP <- gsub("m|M", "6", data$PROPDMGEXP)
data$PROPDMGEXP <- gsub("b|B", "9", data$PROPDMGEXP)
data$PROPDMGEXP <- as.numeric(data$PROPDMGEXP)

## Warning: NAs introduced by coercion

data$PROPDMGEXP <- ifelse(is.na(data$PROPDMGEXP), 0, data$PROPDMGEXP)
table(data$PROPDMGEXP, exclude=NULL)

##
##          0      1      2      3      4      5      6      7      8      9
## 466164     25     20 424669      4     28  11341      5      1     40
##    <NA>
##          0
```

CROPDMGEXP

```
table(data$CROPDMGEXP, exclude=NULL)

##
##           ?      0      2      B      k      K      m      M    <NA>
## 618413      7     19      1      9     21 281832      1  1994      0

data$CROPDMGEXP <- as.character(data$CROPDMGEXP)
data$CROPDMGEXP <- gsub("h|H", "2", data$CROPDMGEXP)
data$CROPDMGEXP <- gsub("k|K", "3", data$CROPDMGEXP)
data$CROPDMGEXP <- gsub("m|M", "6", data$CROPDMGEXP)
data$CROPDMGEXP <- gsub("b|B", "9", data$CROPDMGEXP)
data$CROPDMGEXP <- as.numeric(data$CROPDMGEXP)
```

```
## Warning: NAs introduced by coercion

data$CROPDMGEXP <- ifelse(is.na(data$CROPDMGEXP), 0, data$CROPDMGEXP)
table(data$CROPDMGEXP, exclude=NULL)

##
##      0      2      3      6      9    <NA>
## 618439  1 281853 1995      9      0
```

Compute Property and Crop damage values.

Then we compute the Property or Crop numerical damages.

```
data <- mutate(data, PROPERTY = PROPDMG*10^PROPDMGEXP, CROP =
CROPDMG*10^CROPDMGEXP, ECONOMY = PROPERTY + CROP)
```

Find the top 10 severe weather events for Property and Crop damage, and economy as a whole.

```
dataProperty <- data %>%
  group_by(EVTYPE) %>%
  summarise(totPropertyDamage = sum(PROPERTY)) %>%
  arrange(desc(totPropertyDamage)) %>%
  head(n=10)

dataCrop <- data %>%
  group_by(EVTYPE) %>%
  summarise(totCropDamage = sum(CROP)) %>%
  arrange(desc(totCropDamage)) %>%
  head(n=10)

dataEconomy <- data %>%
  group_by(EVTYPE) %>%
  summarise(totPropertyDamage=sum(PROPERTY),
totCropDamage=sum(CROP), totEconomy=totPropertyDamage+
totCropDamage) %>%
  arrange(desc(totEconomy)) %>%
  head(n=10) %>%
  melt(id.vars="EVTYPE")
```

Impact on Population Health

We use the total number of fatalities and injuries as a measure of impact on population health during severe weather events.

```
dataFatalities <- data %>%
  group_by(EVTYPE) %>%
  summarise(totFatalities = sum(FATALITIES)) %>%
  arrange(desc(totFatalities)) %>%
  head(n=10)
```

```
dataInjuries <- data %>%
  group_by(EVTYPE) %>%
  summarise(totInjuries = sum(INJURIES)) %>%
  arrange(desc(totInjuries)) %>%
  head(n=10)
```

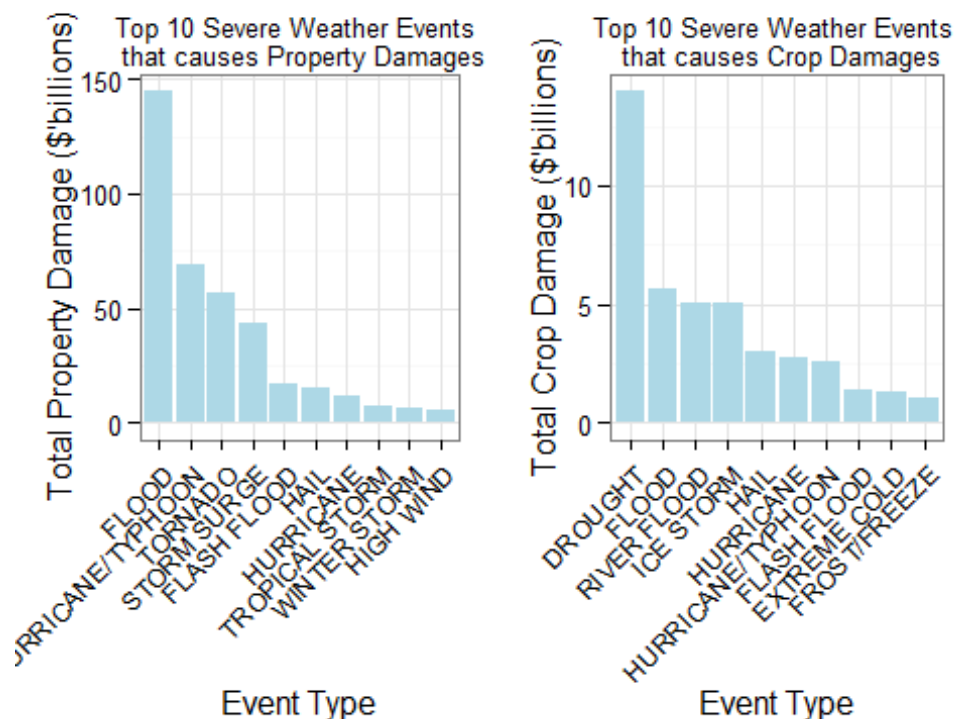
Results

Impact on Property and Crop (Economic)

```
plotProperty <- ggplot(dataProperty, aes(x=reorder(EVTYPE, -
totPropertyDamage), y=totPropertyDamage/10^9)) +
  geom_bar(stat="identity", fill="lightblue") +
  labs(x="Event Type", y="Total Property Damage ($'billions)",
       title="Top 10 Severe Weather Events \n that causes
Property Damages") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(size = rel(0.8)))

plotCrop <- ggplot(dataCrop, aes(x=reorder(EVTYPE, -totCropDamage),
y=totCropDamage/10^9)) +
  geom_bar(stat="identity", fill="lightblue") +
  labs(x="Event Type", y="Total Crop Damage ($'billions)",
       title="Top 10 Severe Weather Events \n that causes Crop Damages")
+
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(size = rel(0.8)))

grid.arrange(plotProperty, plotCrop, ncol=2)
```



Impact on Fatalities and Injuries (Population Health)

```
plotFatalities <- ggplot(dataFatalities, aes(x=reorder(EVTYPE, -
totFatalities), y=totFatalities/10^3)) +
  geom_bar(stat="identity", fill="lightgreen") +
  labs(x="Event Type", y="Total Fatalities ('thousands)",
       title="Top 10 Severe Weather Events \n that causes Fatalities") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(size = rel(0.8)))

plotInjuries <- ggplot(dataInjuries, aes(x=reorder(EVTYPE, -totInjuries),
y=totInjuries/10^3)) +
  geom_bar(stat="identity", fill="lightgreen") +
  labs(x="Event Type", y="Total Injuries ('thousands)",
       title="Top 10 Severe Weather Events \n that causes Injuries") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(size = rel(0.8)))

grid.arrange(plotFatalities, plotInjuries, ncol=2)
```

