Weather events in the U.S.: Health and Economic Consequences

Author: Rodrigo Peres Ferreira Date: April 08, 2016

Synopsis

This analysis aims to answer two main questions: Which types of weather events are the most harmful to population health in the U.S., and which types have the greatest economic impacts. To address these questions, we have used the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database, which contains data about weather events between April 1950 and November 2011. The results were that the most harmful events for population health are Tornados, Excessive Heat, Thunderstorm Heat, Floods and Lightnings, and the ones that have the greatest economic impacts are Floods, Hurricanes, Tornados, Storm Surges and Hail.

Additional Libraries Required

```
library(plyr)
```

1. Data Processing

First, let us unzip and load the original CSV file from NOAA database (The read.csv() function is able to do both).

```
weather <- read.csv("repdata-data-StormData.csv.bz2")
```

For the purpose of this work, we do not need information about states or date and time, we only need need information about event types, fatalities/injuries and types of damages. We can filter the original data frame to achieve this.

```
weather_selected <- data.frame(weather$EVTYPE, weather$FATALITIES, weather$INJU
RIES, weather$PROPDMG, weather$PROPDMGEXP, weather$CROPDMG, weather$CROPDMGEXP)

names(weather_selected) <- c('EVTYPE','FATALITIES','INJURIES', 'PROPDMG', 'PROP
DMGEXP', 'CROPDMG', 'CROPDMGEXP')
```

We can also convert the columns "PROPDMG" and "PROPDMGEXP" into one column ("PROPDMG"), since the second column is the magnitude of the value in the first column.

```
weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'K'] <-
    weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'K']*1000

weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'M'] <-
    weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'M']*1000000

weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'B'] <-
    weather_selected$PROPDMG[weather_selected$PROPDMGEXP == 'B']*1000000000

weather_selected <- weather_selected[,-5]
```

We can do the same procedure with the columns "CROPDMG" and "CROPDMGEXP"

```
weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'K'] <-
    weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'K']*1000

weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'M'] <-
    weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'M']*1000000

weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'B'] <-
    weather_selected$CROPDMG[weather_selected$CROPDMGEXP == 'B']*1000000000

weather_selected <- weather_selected[,-6]
```

```
weather <- weather_selected
weather$EVTYPE <- as.factor(weather$EVTYPE)
```

Now we have a nice 5-variable data frame (weather) we can easily work with for the purposes of this analysis.

Results

Now we can answer the two main questions of this analysis report.

Let us first aggregate the data, to have the total sum of the number of injuries, fatalities and damages to generate our plots.

```
weather_aggregation <- ddply(weather,.(EVTYPE),summarize, total_harm=sum(FATALI
TIES)+sum(INJURIES), total_damage=sum(PROPDMG)+sum(CROPDMG))
```

With this aggregated data, we can answer the questions below.

Question 1. Across the United States, which types of events are most harmful with respect to population health?

```
Harms <- weather_aggregation$total_harm
names(Harms) <- weather_aggregation$EVTYPE

Harms <- rev(sort(Harms))

barplot(Harms[1:5], col = "red", main = "5 Most Harmful Events for Population H
ealth in the U.S.", ylab = "Total of Injuries/Fatalities")
```

5 Most Harmful Events for Population Health in the U.S.
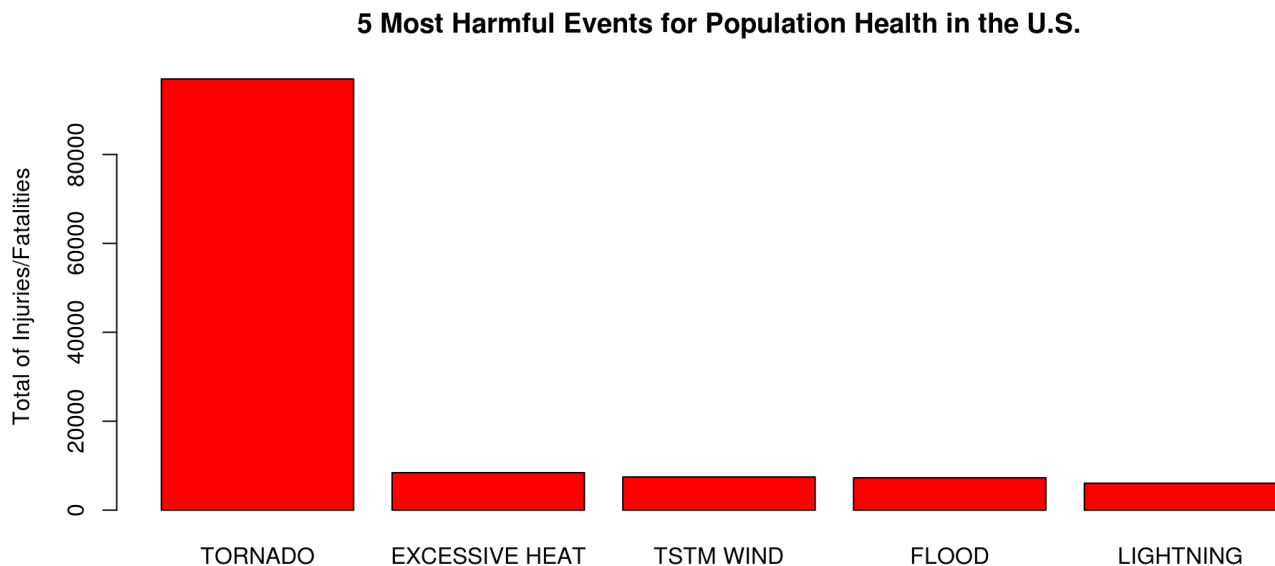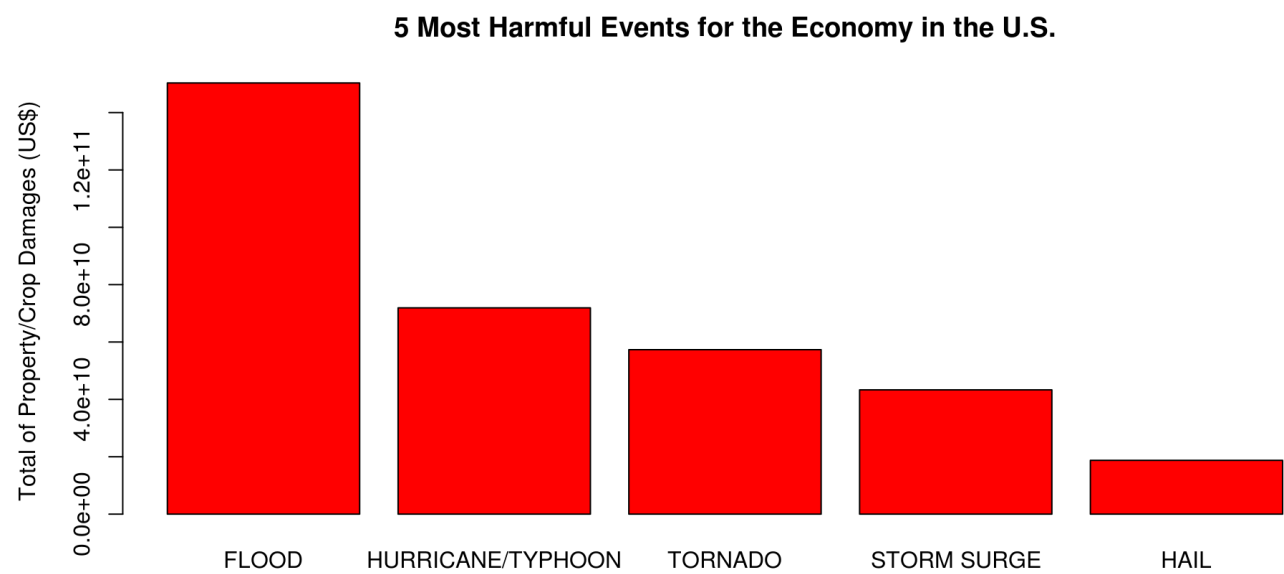

As we can see from the plot above, the most harmful events for population health across the U.S. are Tornados, Excessive Heat, Thunderstorm Heat (TSTM HEAT), Floods and Lightnings.

Question 2. Across the United States, which types of events have the greatest economic consequences?

```
Damages <- weather_aggregation$total_damage
names(Damages) <- weather_aggregation$EVTYPE

Damages <- rev(sort(Damages))

barplot(Damages[1:5], col = "red", main = "5 Most Harmful Events for the Econom
y in the U.S.", ylab = "Total of Property/Crop Damages (US$)")
```

5 Most Harmful Events for the Economy in the U.S.

As we can see from the plot above, the most harmful events for the economy across the U.S. are Floods, Hurricanes, Tornados, Storm Surges (Tides) and Hail.