Storms and damages: a brief report

Josep Anton Mir Tutusaus

Monday, May 18, 2015

1.Abstract

Here I present a data analysis performed in the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. Firstly, I read into R only the variables of interest (described below) and obtain its mean, grouped by type of event. Then I arrange the data so the higher mean is in the first row. It's those arranged "means by event" that I present in the form of graph (2 figures). The "Data processing" section describes how the data is read and processed prior to the analysis, which takes places in the section "Results", as well as the graphs.

2. Data Processing

Loading libraries and setting the default figure size to 10x12

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
knitr::opts_chunk$set(fig.width=10, fig.height=12)
```

The only variables of interest in my analysis are the "type of event" and those related to economic and population consequences. In order to spend the minimum RAM resources, only those variables (columns) are read. The data is then grouped by event type (variable EVTYPE) and its mean calculated by the "summarise" function in the "dplyr" package. All data discussed, then, are means of the raw data.

```
data <- read.csv("repdata_data_StormData.csv.bz2", colClasses=c("NULL","NULL","N
ULL","NULL","NULL","NULL","NULL", NA,"NULL","NULL","NULL","NULL","NULL","NULL","
NULL","NULL","NULL","NULL","NULL","NULL","NULL",NA,NA,NA,NA,NA,NA,"NULL",
"NULL","NULL","NULL","NULL","NULL","NULL","NULL","NULL"))

by_event_data <- group_by(data, EVTYPE)
```

a. Types of events most harmful with respect to population health.

Only the top ten events with the most fatalities, and the top ten events with the most injuries are selected in order to perform the analysis. Most of the events do not have any repercussion in the variables selected,

that is why only the top ten events are studied.

```
mean_by_event_data <- summarise(by_event_data, meanFatalities=mean(FATALITIES),
meanInjuries=mean(INJURIES))
topFatalities <- mean_by_event_data %>% filter(meanFatalities>0 | meanInjuries>0
) %>% arrange(desc(meanFatalities, meanInjuries)) %>% slice(1:10)
topInjuries <- mean_by_event_data %>% filter(meanFatalities>0 | meanInjuries>0)
%>% arrange(desc(meanInjuries, meanFatalities)) %>% slice(1:10)
```

b. Types of events with the greatest economic consequences.

Only the top ten events with regard to property and crop damage are selected, following the previous reasoning.

```
eco_mean_by_event_data <- summarise(by_event_data, meanPROPDMG=mean(PROPDMG), me
anCROPDMG=mean(CROPDMG))
topProperty <- eco_mean_by_event_data %>% filter(meanPROPDMG>0 | meanCROPDMG>0)
%>% arrange(desc(meanPROPDMG, meanCROPDMG)) %>% slice(1:10)
topCrop <- eco_mean_by_event_data %>% filter(meanPROPDMG>0 | meanCROPDMG>0) %>%
arrange(desc(meanCROPDMG, meanPROPDMG)) %>% slice(1:10)
```

3. Results

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

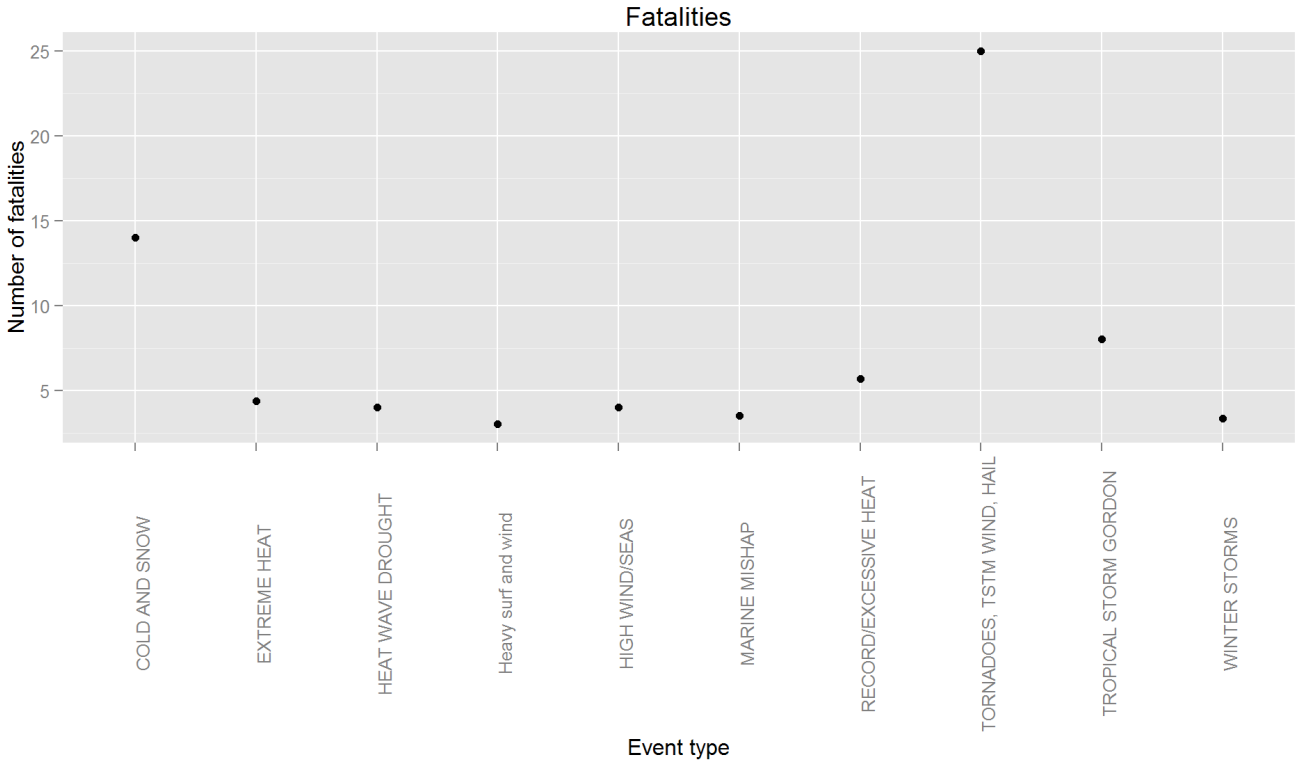
```
## Warning: package 'gridExtra' was built under R version 3.1.3
```

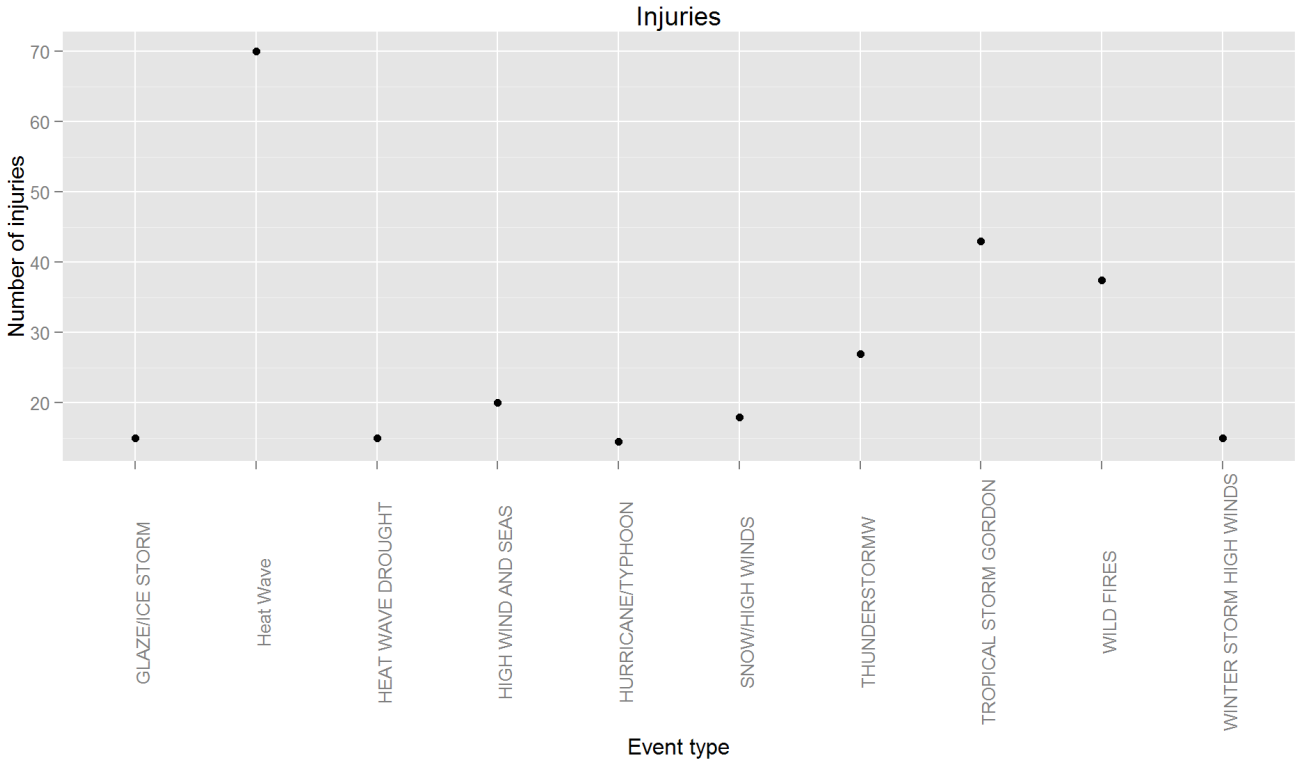
```
## Loading required package: grid
```

```
q1 <- qplot(topFatalities$EVTYPE,topFatalities$meanFatalities, data=topFatalitie
s, xlab="Event type", ylab="Number of fatalities", main="Fatalities")
q1 <- q1 + theme(axis.text.x=element_text(angle=+90))

q2 <- qplot(topInjuries$EVTYPE,topInjuries$meanInjuries, data=topInjuries, xlab=
"Event type", ylab="Number of injuries", main="Injuries")
q2 <- q2 + theme(axis.text.x=element_text(angle=+90))

grid.arrange(q1, q2, nrow=2)
```





This figure shows the ten most dangerous events in regards to the mean of fatalities and injuries. Tornadoes/hail and cold/snow are the events with the most casualties, with 25 and 14 fatalities, respectively. In contrast, it is the heat wave the event with more injuries, 70, followed by tropical storm Gordon and wild fires, with 44 and 38 injured pepople, respectively.

```
filter(data, FATALITIES==max(FATALITIES))
```

```
##   EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1   HEAT        583        0       0                  0
```

```
filter(data, INJURIES==max(INJURIES))
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO         42     1700     250          M       0
```

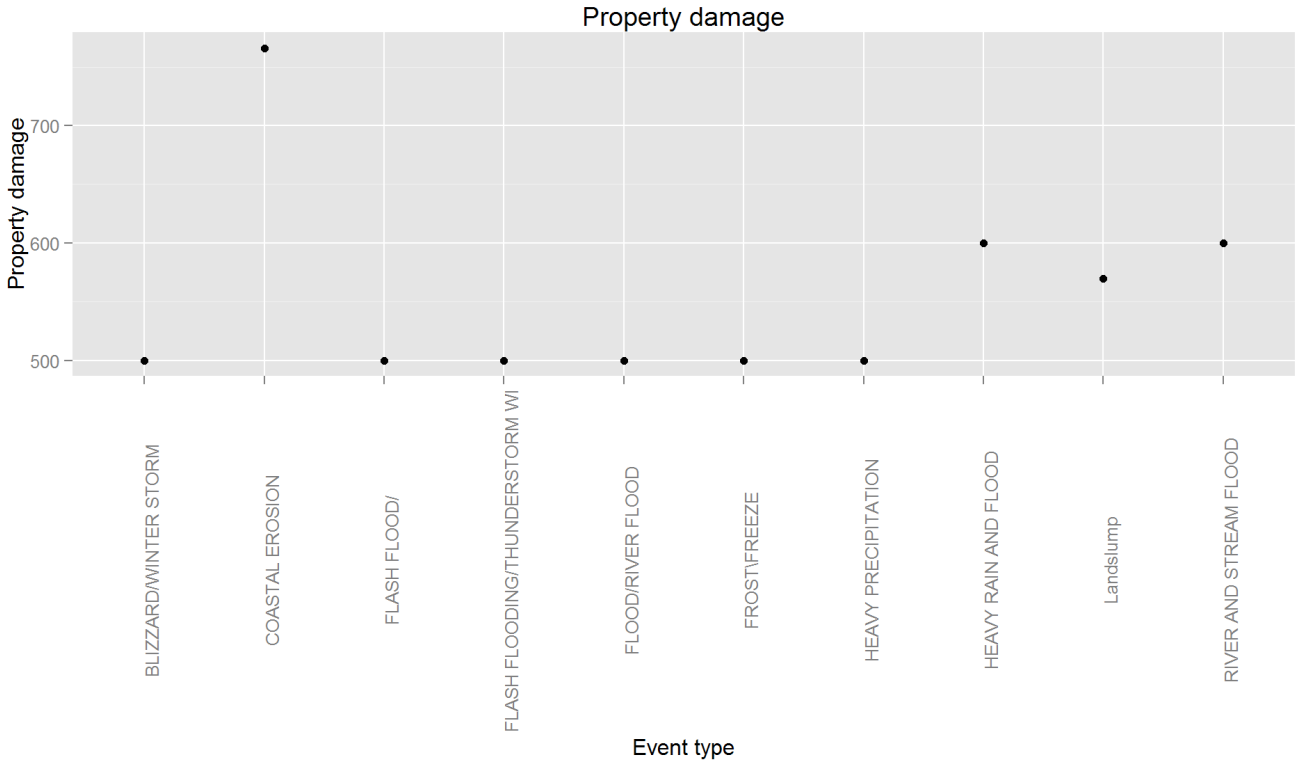
As you can see in these tables, heat is the most fatal event, and tornado is the worst event in reagards to injured pepole.

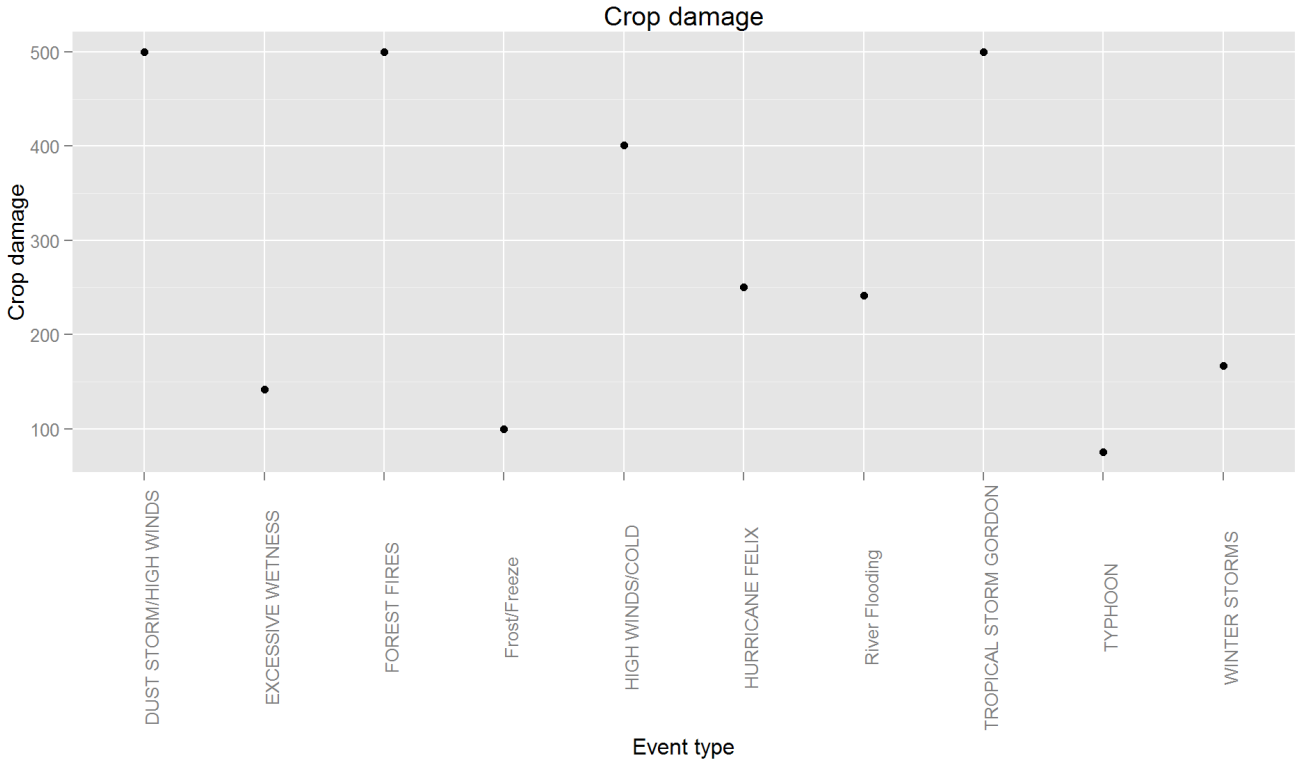
3.2. Which types of events have the greatest economic consequences?

```
q3 <- qplot(topProperty$EVTYPE,topProperty$meanPROPDMG, data=topProperty, xlab="
Event type", ylab="Property damage", main="Property damage")
q3 <- q3 + theme(axis.text.x=element_text(angle=+90))

q4 <- qplot(topCrop$EVTYPE,topCrop$meanCROPDMG, data=topCrop, xlab="Event type",
 ylab="Crop damage", main="Crop damage")
q4 <- q4 + theme(axis.text.x=element_text(angle=+90))

grid.arrange(q3, q4, nrow=2)
```





This figure shows the ten most dangerous events in regards to economic consequences; more precisely, to property damage and crop damage. Coastal erosion is the event with the most cost, nearly 800. When speaking about crop damage, dust storm/high winds, forest fires and the tropical storm Gordon are tied with 500 worth of crop damage.

```
filter(data, PROPDMG==max(PROPDMG))
```

```
##              EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG
## 1 THUNDERSTORM WIND          0        0    5000          K       0
## 2       FLASH FLOOD          0        0    5000          K       0
## 3       FLASH FLOOD          0        0    5000          K       0
## 4        WATERSPOUT          0        0    5000          K       0
##   CROPDMGEXP
## 1          K
## 2          K
## 3          K
## 4          K
```

```
filter(data, CROPDMG==max(CROPDMG))
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 DROUGHT          0        0       0                990          K
```

As you can see in these tables, thunderstorm wind, flash flood and watersprout are the events with the most property damage, while drought (as we can expect) is the event with the most crop damage.