Analysing severe weather events from NOAA Database

Christian

17/3/2018

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.
The following report tries to answer 2 important questions in that regard:

1. Across the United States, which types of events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

Data Processing

Importing the data

The data used for this analysis is downloaded from (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) In the following code I assume the data has been downloaded and stored in the project reopsitory.

import libraries and Download the data:

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.4.3
```

```
url <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
download.file(url, destfile = 'storm_data.csv.bz2')
data <- read.csv('storm_data.csv.bz2', header = TRUE)
```

In order to see the data structure we will examine it a bit

```
str(data)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
```

```
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","  Christiansburg",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","  CANTON","  TULIA",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","  CI","%SD",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Looking at the documentation from, it seems the columns we are interested in are the 'PROPDMG' and 'CROPDMG' columns, since they describe the economic cost of events. But they don't represent the actual value, since the 'PROPDMGEXP' and 'CROPDMGEXP' columns contains the factor to multiply on the values. We need to align the different rows, so we wil make the factor columns lowercase, apply the a factorlogic and multiply the columns together and

```
#Making the al lowercase
data$PROPDMGEXP <- tolower(data$PROPDMGEXP)
data$CROPDMGEXP <- tolower(data$CROPDMGEXP)

#Creating the multiplication factor for PROPDMG
data$prop_factor <- 1 # default one, so we only have to translate the actual fctors
data$prop_factor[data$PROPDMGEXP == "h" || data$PROPDMGEXP == "2"] <- 10^2
data$prop_factor[data$PROPDMGEXP == "k" || data$PROPDMGEXP == "3"] <- 10^3
data$prop_factor[data$PROPDMGEXP == "4"] <- 10^4
data$prop_factor[data$PROPDMGEXP == "5"] <- 10^5
data$prop_factor[data$PROPDMGEXP == "m" || data$PROPDMGEXP == "6"] <- 10^6
data$prop_factor[data$PROPDMGEXP == "7"] <- 10^7
data$prop_factor[data$PROPDMGEXP == "8"] <- 10^8
```

```
data$prop_factor[data$PROPDMGEXP == "b"] <- 10^9

#Creating the multiplication factor for CROPDMG
data$crop_factor <- 1 # default one, so we only have to translate the actual fctors
data$crop_factor[data$CROPDMGEXP == "2"] <- 10^2
data$crop_factor[data$CROPDMGEXP == "k"] <- 10^3
data$crop_factor[data$CROPDMGEXP == "m"] <- 10^6
data$crop_factor[data$CROPDMGEXP == "b"] <- 10^9

# Now we can create the actual cost columns
data$prop_cost <- data$PROPDMG * data$prop_factor
data$crop_cost <- data$CROPDMG * data$crop_factor
```

Results

In this section I will analyse the data and present the answer to the two questions ### Events that are most harmful with respect to population health I will start by calculating the total injuries and fatalities

```
Total_injurie <- data %>%
                  group_by(EVTYPE) %>%
                  summarise(INJURIES = sum(INJURIES), FATALITIES = sum(FATALITIES), total_injurie = INJ
                  arrange(desc(total_injurie))
Total_injurie <- head(Total_injurie, 10)
```

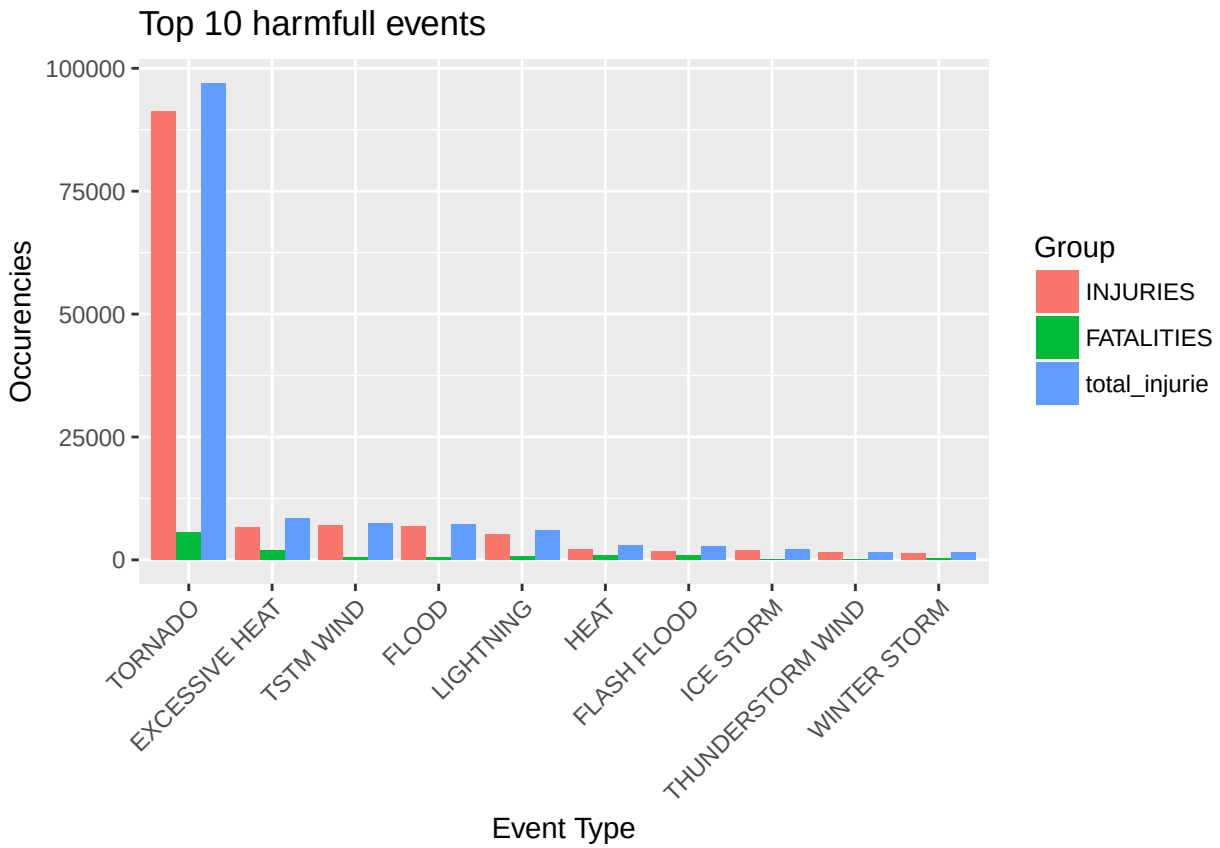
In order to use the ggplot package, I need to melt the data

```
inj_chart_data <- melt(Total_injurie, id.vars = "EVTYPE", variable.name = "Group")
head(inj_chart_data)
```

```
##           EVTYPE    Group value
## 1        TORNADO INJURIES 91346
## 2 EXCESSIVE HEAT INJURIES  6525
## 3      TSTM WIND INJURIES  6957
## 4          FLOOD INJURIES  6789
## 5      LIGHTNING INJURIES  5230
## 6           HEAT INJURIES  2100
```

Charting the data for a more visual inspaection

```
ggplot(inj_chart_data, aes(x = reorder(EVTYPE, -value), y = value)) +
  geom_col(aes(fill = Group), position = 'dodge') +
  ylab('Occurencies') +
  xlab('Event Type') +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  ggtitle('Top 10 harmfull events')
```



Events that have the greatest economic consequences I will start with calculating the total costs

```
Total_cost <- data %>%
              group_by(EVTYPE) %>%
              summarise(prop_cost = sum(prop_cost), crop_cost = sum(crop_cost), total_cost = prop_cost +
              arrange(desc(total_cost))
Total_cost <- head(Total_cost, 10)
```

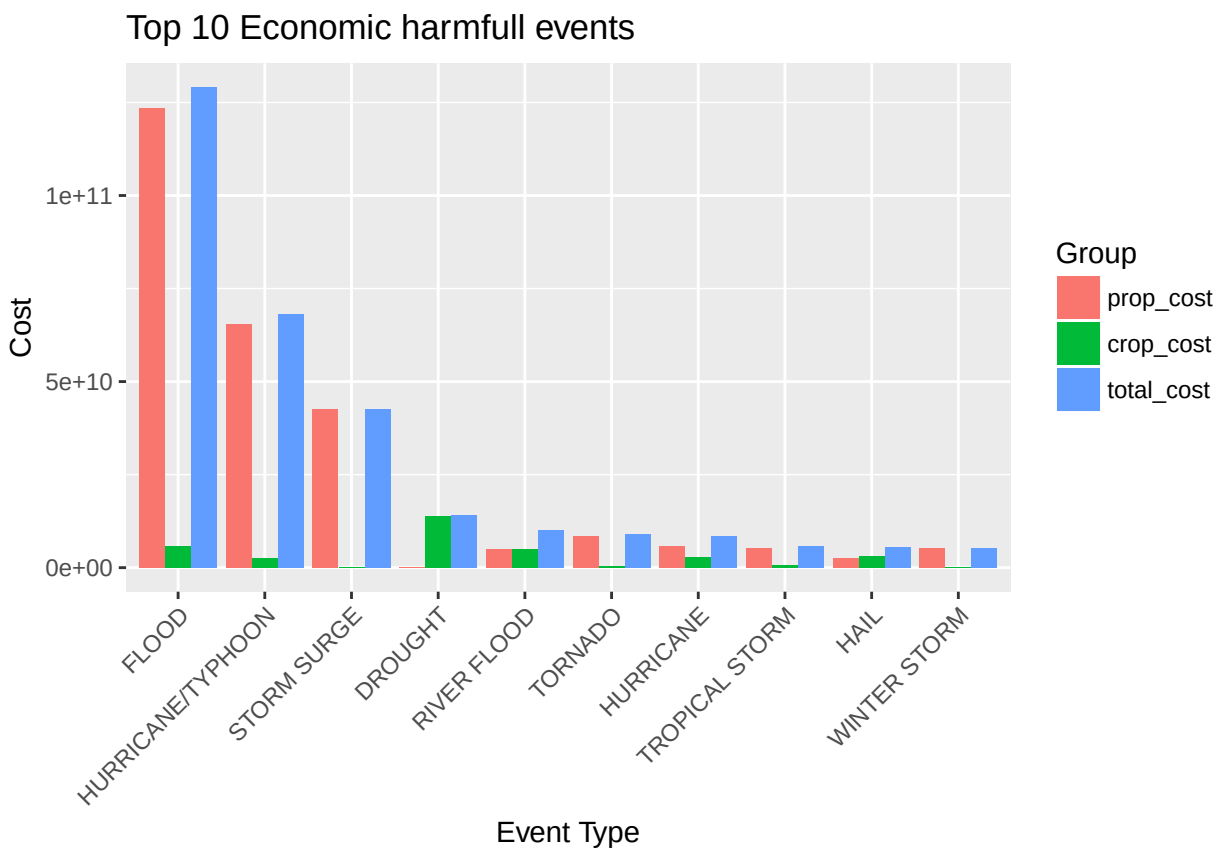
Again in order to easily visualise in ggplot I will melt the dataset

```
cost_chart_data <- melt(Total_cost, id.vars = 'EVTYPE', variable.name = 'Group')
head(cost_chart_data)
```

```
##              EVTYPE     Group        value
## 1             FLOOD prop_cost 123399815980
## 2 HURRICANE/TYPHOON prop_cost  65505773870
## 3       STORM SURGE prop_cost  42579350930
## 4           DROUGHT prop_cost      4099050
## 5       RIVER FLOOD prop_cost   5013850700
## 6           TORNADO prop_cost   8522370660
```

creating the chart for a more visual inspaection

```
ggplot(cost_chart_data, aes(x = reorder(EVTYPE, -value), y = value)) +
  geom_col(aes(fill = Group), position = 'dodge') +
  ylab('Cost') +
  xlab('Event Type') +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  ggtitle('Top 10 Economic harmfull events')
```


Top 10 Economic harmfull events
Cost
1e+11
5e+10
0e+00
FLOOD
HURRICANE/TYPHOON
STORM SURGE
DROUGHT
RIVER FLOOD
TORNADO
HURRICANE
TROPICAL STORM
HAIL
WINTER STORM
Event Type
Group
prop_cost
crop_cost
total_cost