Severe Weather Events Caused Hazard Analysis

Zhongyu

Tuesday, March 17, 2015

Severe Weather Events Caused Hazard Analysis

Synopsis

This report analyzes the Storm Dataset provided by NOAA to investigate the hazards caused by severe weather events with respect to population health and economy. To evaluate the harmful influences on population health, both fatality and injury counts are taken account of. The result shows that the top three most harmful weather events with respect to population health are tornado, heat, and thunderstorm. For evaluating economic impacts, the total damage is calculated by adding up the property damage and crop damage. The analysis shows that the top three weather events that bring significant economic consequences are tornado, flood and hail respectively.

Data Processing

This project is using R under Rstudio environment to perform all the analysis. The more detailed data processing steps is dissected in this section.

1. Loading data The Strom Dataset can be accessed in NOAA website, and it can be downloaded in CSV format. After setting up R working directory to the file where the downloaded data is stored, the following script can be executed for loading the data into R workspace.

```
storm <- read.csv(bzfile('repdata-data-StormData.csv.bz2', open='r'), header=TRUE)
```

The original storm dataset contains 902297 observations of 37 variables, the parameters that are needed for this project are: event type, fatality, injury, property damage, crop damage. In the dataset, the parameters which correctly reflect property/crop damage are splitted into two variables to store in the dataset: one (PROPDMG/CROPDMG) stores the base value, the other(PROPDMGEXP/CROPDMGEXP) stores the magnitudes/exponentials. And those associated variables are selected to form a new data frame specifically for perfoming this analysis.

```
varnames <- c('EVTYPE', 'FATALITIES', 'INJURIES', 'PROPDMG','PROPDMGEXP',
              'CROPDMG','CROPDMGEXP')
stormsel <- subset(storm,select=varnames)
colnames(stormsel) <- tolower(colnames(stormsel))
```

2. Cleaning and Consolidate data According to the Storm Data Event Table on page 6 of Storm Data Documentation, there are 48 types of storm events entried into the event type column of the dataset. However a quick investigation shows otherwise: event type variable has actually 985 categories.

```
library(dplyr)
TempData <- group_by(stormsel, evtype) %>% summarise(sum(fatalities, na.rm=TRUE), sum(injuries, na.rm=T
str(TempData$evtype)
```

```
##  Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 1 2 3 4 5 6 7 8 9 10 ...
```

This disagreement between documentation and dataset is most likely due to the wide range of data resources: some data might be recorded before standardizing event types in Storm Data Event Table. A closer gross look into the contents in event type column shows big variaty of descriptive entires, such as "record high", which shold be categorized as "Heat" accoording to Storm Data Event Table.

A great amount of work is devoted to address this issue to form a consolidated event type column and a cleaner dataset. The approach is a combination of manually classifying aliasing event types (i.e. high temperature = excessive warmth = HEAT; wintry mix = freezing drizzle = mixed precip = SLEET) and utilizing regular expression to replace the aliasing event type names in Storm Data Event Table. Besides the 48 types in Storm Data Documentations, two more types are added to serve as categorical options: land slide and others. The reason for this operation is due to the high frequency occurrence of land slides in the data, and whereas it is hard and ambiguous to simply equalize land slides to any of the 48 event types in the table. The others option is added for storing some entries, which are either misentry (i.e. summary of...), or extremely difficult to classify into any of the other 49 event types(i.e. simply denoted "high" without explicitly pointing out it is temperature high, or tidy high, etc). The code chunk for this section of processing is shown as following.

```
stormsolid <- mutate(stormsel, evtypesolid = as.character(tolower(evtype)))

evtable <- c('Astronomical Low Tide', 'Avalanche|avalance', 'Blizzard',
             'Coastal Flood|coastal surge',
             'Wind Chill|Cold Chill|cool|low temp|record low',
             'Debris Flow','Dense Fog|fog|vog', 'smoke|Dense Smoke',
             'Drought', 'Dust Devil|dust',
             'Dust Storm', 'Excessive Heat',
             'Extreme Cold|Wind Chill|cold', 'Flash Flood|flash floo',
             'Flood|stream',
             'Frost|Freeze|glaze|ice|icy', 'Funnel Cloud|Funne',
             'Freezing Fog|freezing spray','Hail',
             'Heat|hot|high temperature|record high|warm|dry',
             'Heavy Rain|wet|heavy shower|excessive precipitation|heavy precip|rain',
             'Heavy Snow|snow',
             'High Surf|surf','High Wind|wind|wnd','Hurricane|Typhoon', 'Ice Storm',
             'Lake Effect Snow',
             'Lakeshore Flood', 'Lightning|lighting|ligntning|lights','Marine Hail',
             'Marine High Wind',
             'Marine Strong Wind', 'Marine TSTM Wind|coastal storm','Rip Current',
             'Seiche',
             'Sleet|wintry mix|mix precip|freezing drizz|wintery mix|mixed precip
             |winter mix|heavy mix',
             'Storm Surge|high sea|high swell|high water|rough seas|wave|marine mishap
             |heavy sea|rising water|tide',
             'Strong Wind',
             'thunderstorm|TSTM Wind|tstm','Tornado|gustnado|wall cloud|torndao',
             'Tropical Depression',
             'Tropical Storm','Tsunami','Volcanic Ash|volcanic',
             'Waterspout|water spout|wayterspout',
             'Wildfire|fire|red flag critera',
             'Winter Storm|microburst|downburst','Winter Weather',
             'landslide|mudslide|rock slide|mud slide',
             'other|summary|none|temperature record|record temperature|monthly')
numbs <- seq(1,50,1)
```

```
i <- 1

for (event in evtable){
    indTemp <- grep(event, stormsolid$evtype, ignore.case=TRUE)
    stormsolid$evtypesolid[indTemp] <- i
    i <- i+1
}

markind <- stormsolid$evtypesolid %in% numbs
nonmarkind <- which(!markind)
others  <- stormsolid$evtypesolid[nonmarkind]
```

Note: Each element in the evtable variable are corresponding to a number element in numbs. And for each element in evtable, the first word before "or"("|") sign is the exact event type name that can be matched to the Storm Data Event Table in Storm Data Documentation.

3. Severe Weather Events Caused Hazards Impacting Population Health Issues After consolidating event type column and obtaining a cleaner tidier dataset, the fatality counts and injury counts are computed for each category of severe weather events. Thereafter fatality counts and injury counts are added together to obtain the total counts. Hereafter the data is sorted primarily based on the total counts to seek out the top three harmful event types.

```
# group variables
health <- group_by(stormsolid, evtypesolid) %>% summarise(sum(fatalities), sum(injuries))
colnames(health) <- c('evetype', 'fatality', 'injury')

health <- mutate(health, fat_plus_inj = fatality + injury)
healthsort <- arrange(health,  desc(fat_plus_inj), desc(fatality), desc(injury))
evtable[as.numeric(healthsort$evetype[1:3])]
```

```
## [1] "Tornado|gustnado|wall cloud|torndao"
## [2] "Heat|hot|high temperature|record high|warm|dry"
## [3] "thunderstorm|TSTM Wind|tstm"
```

As the analysis reports, Tornado, Heat and Thunderstorm Wind are the top three severe weather event types that causes greatest population health issue. A bar plot is shown in resulst section to demonstrate the fatalities and injuries by each of them.

4. Severe Weather Events Caused Economic Consequences The following formula is taken to compute the property damage and crop damage:

property damage = PROPDMG * 10^PROPDMGEXP crop damage = CROPDMG * 10^CROPDMGEXP

The recording in PROPDMGEXP and CROPDMGEXp columns are not all consistent. Some of them are coded by letters: "H" denotes for hundred(x 100), "K" denotes for thousand(x 1000), "M" denotes for million(x 1000,000), "B" denotes for billion(x 1000,000,000); some of them are coded by numbers directly; and some of them has notations (i.e. "?", "+", "-") suspecting to be missing values or misentry. A quick consolidation is taken for clearing up contents in these two columns. All the uninterpretable entry contents are treated as (x 1), which will not magnify values stored in PROPDMG and CROPDMG columns.

After the further consolidation, property damage costs and crop damage costs are calculated for each category of event type, and then added up to compute the total damage costs.

```
stormclean <- stormsolid
expnote <- c('0','1','h|H|2','K|3','4','5','K|6','7','8','M','B')
expnum <- c('1','10','100','1000','10000','100000','1000000','10000000',
            '100000000','1000000000','1000000000')
i <- 1
stormclean$propdmgexp <- as.character(stormclean$propdmgexp)
stormclean$cropdmgexp <- as.character(stormclean$cropdmgexp)

for (exps in expnote){
    indTemp <- grep(exps, stormclean$propdmgexp, ignore.case=TRUE)
    stormclean$propdmgexp[indTemp] <- expnum[i]
    i = i+1
}

markind <- stormclean$propdmgexp %in% expnum
nonmarkind <- which(!markind)
stormclean$propdmgexp[nonmarkind] <- 1
stormclean$propdmgexp <- as.numeric(stormclean$propdmgexp)

cropexpnote <- c('0','1','h|H|2','K|3','4','5','K|k|6','7','8','M|m','B')
i <- 1
for (exps in cropexpnote){
    indTemp <- grep(exps, stormclean$cropdmgexp, ignore.case=TRUE)
    stormclean$cropdmgexp[indTemp] <- expnum[i]
    i=i+1
}

markind <- stormclean$cropdmgexp %in% expnum
nonmarkind <- which(!markind)
stormclean$cropdmgexp[nonmarkind] <- 1
stormclean$cropdmgexp <- as.numeric(stormclean$cropdmgexp)

stromclean <- mutate(stormclean, propdmgv = propdmg*propdmgexp, cropdmgv=cropdmg*cropdmgexp)
economy <- group_by(stromclean, evtypesolid) %>% summarise(sum(propdmgv,na.rm=TRUE),sum(cropdmgv,na.rm='
colnames(economy) <- c('evetype','propdmg', 'cropdmg')
economy <- mutate(economy, totaldmg=propdmg+cropdmg)
economysort <- arrange(economy, desc(totaldmg), desc(propdmg),desc(cropdmg))
evtable[as.numeric(economysort$evtype[1:3])]
```

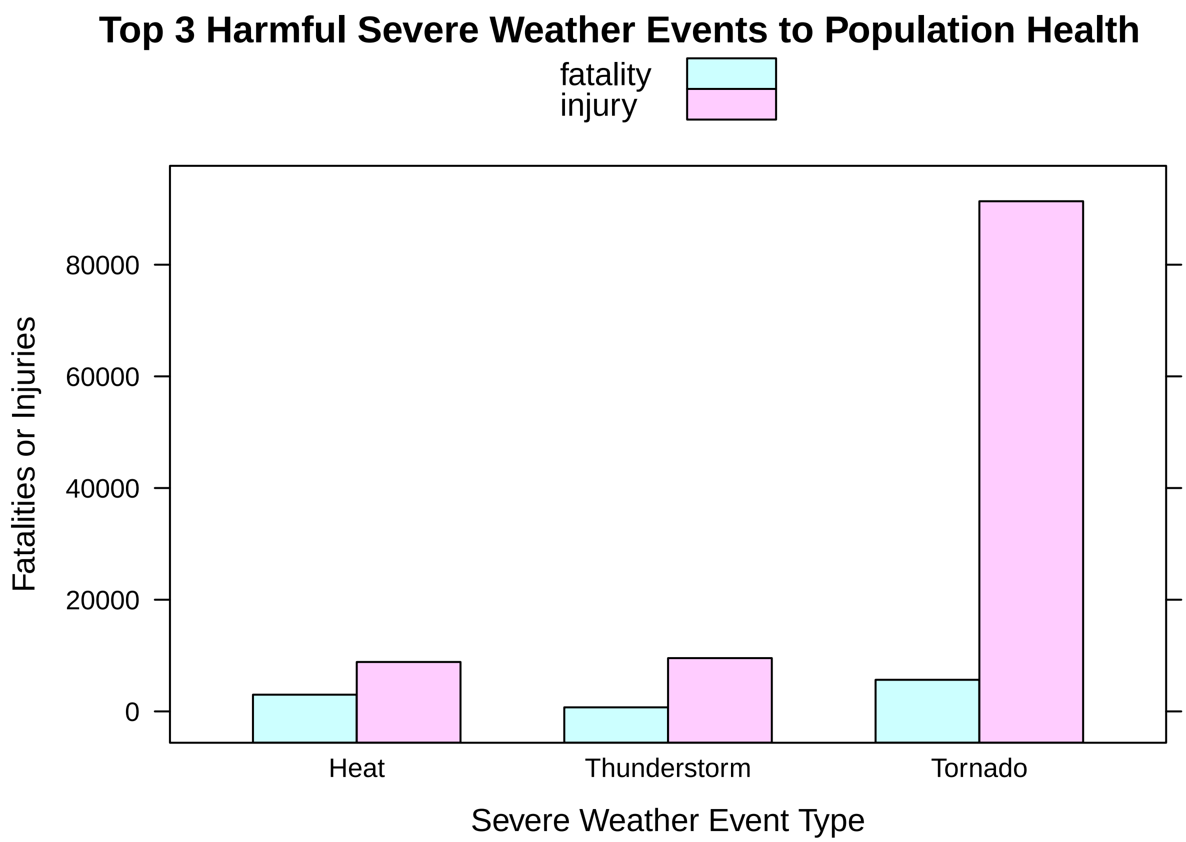
```
## character(0)
```

As the analysis reports, Tornado, Flood and Hail are the top three severe weather event types that causes greatest economic consequences. A bar plot is shown in results section to visually demonstrate the property damage and crop damage by each of them.

Results

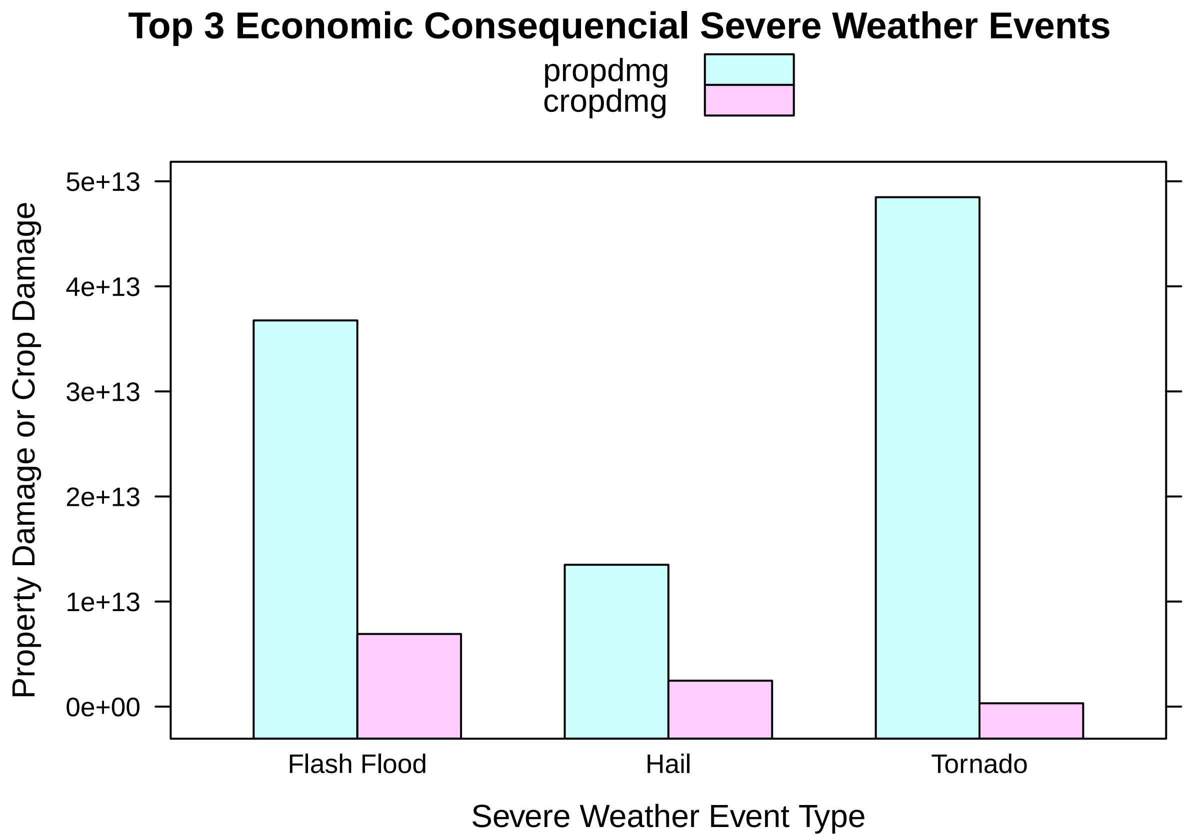
As the barcharts showing below, tornado brings the most fatalities and a lot of injuries. Heat brings more fatality counts than thunerstorm, however thunderstorm brings slightly more injuries than heat. Overall, heat has a slightly bigger hazard on population health than thunderstorm

```r
healthplot <- healthsort[1:3,]
healthplot$evetype <- c("Tornado","Heat", "Thunderstorm")
library(lattice)
barchart(fatality+injury~evetype, data=healthplot, auto.key=TRUE,
         xlab="Severe Weather Event Type",
         ylab="Fatalities or Injuries",
         main="Top 3 Harmful Severe Weather Events to Population Health")
```



As for the economic hazards bringing by severe weather events, tornado also ranks at the top for a substantial property damages it brings, and a fairly significant amount of crop damages. Flood outbeats hail with a significantly larger property damaging and a lightly smaller crop damaging, and ranks at the second place. Even though hail brings smaller economical consequences as compared to the other two events, it still has a pretty significant economic hazard as shown in the barchart.

```r
economyplot <- economysort[1:3,]
economyplot$evetype <- c("Tornado","Flash Flood", "Hail")
barchart(propdmg+cropdmg~evetype, data=economyplot, auto.key=TRUE,
         xlab="Severe Weather Event Type",
         ylab="Property Damage or Crop Damage",
         main="Top 3 Economic Consequencial Severe Weather Events")
```



Note that depending on different methods for categorizing the descriptive event types (which can not be directly find in the Storm Data Table) and different methods for determining the hazard ranking, the final concluded ranking order might variate sligtly. However the top 3 population health and economy harmful severe weather event types concluded from this analysis - Tornado, Heat, Thunderstorm, Flood, Hail - should be on the list of severe weather event types which brings significant population health hazard and causes substantial economic consequences. Associated insurance or risk manangement policy should be built to minimize the consequences. And the public should be alerted about the potential hazards these server weather event types could bring.