Evaluating Economic and Safety Effects of Severe Weather Events in the United States

Synopsis

In this document, we look at the effects of severe weather events in the United States, in regards to both public health and economic damage. For this analysis we use the data from the NOAA Storm Database, which includes records from 1950 through November 2011...

Data Processing

We begin by downloading our data to our working directory and reading in our .csv file. We include the output of the str() function to get a first look at the data

```
filename <- "repdata_data_StormData.csv.bz2"
if(!file.exists(filename)){
    dataurl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(dataurl,destfile=filename)
}
NOAA <- read.csv(filename)
```

We want to evaluate the health impacts of our data, so we will subset our data to only include events that resulted in nonzero fatalities or injuries.

```
fatalevents <- subset(NOAA, FATALITIES >= 1)
injuryevents <- subset(NOAA, INJURIES >= 1)
```

Next, we want to group similar events together since the ENVTYPE factor variable has over 900 levels. We begin by storing all the various event types in a vector. We remove some the nondescript events which are summaries of days/months and dont lend themselves to easily to categorization.

```
eventnames <- levels(NOAA$EVTYPE)
regex <- "SUMMARY|Summary|MONTHLY|Monthly"
eventnames <- eventnames[!grepl(regex, eventnames)]
length(eventnames)
```

```
## [1] 912
```

We want to group similar event together (e.g. Snow, Blizzard, Ice are all "winter" events). To achieve this we use regular expressions to extract the relevant event names, which we will later use to subset and summarize the data. We start with all events corresponding to wild fires:

```
regex <- "FIRE|Fire|fire|Smoke|SMOKE"
wildfire <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

For our tornado group, we consder all events involving "funnels", "spouts" and "whirl winds":

```
regex <- "TORNADO|[Tt]ornado|FUNNEL|[Ff]unnel|SPOUT|[Ss]pout|DEVIL|[Dd]evil|[Ww]hirl|WHIRL|TORNDAO"
tornado <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

We use several phrases to match winter events, all of which involve cold temperatures and/or freezing precipitation:

```r
regex <- "M[Ii][Xx]|WINT|[Ww]int|SNOW|[Ss]now|SLEET|I[Cc][Ee]|[Cc]old|COLD|BLIZZARD|[Bb]lizzard|FREEZ|[
winter <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

Next, we want to include all hurricanes, tropical storm and typhoon events:

```r
regex <- "[Hh]urricane|HURRICANE|[Tt]ropical|TROPICAL|DEPRESSION|TYPHOON|[Tt]yphoon"
tropical <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

For thunder storms, we match "thunder", "lightning" and the commonly used abbreviation "tstm":

```r
regex <- "THUNDER|[Tt]hunder|LIGHTNING|[Ll]ightning|TSTM|[Tt]stm"
tstorm <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

Next, we want to capture all the flooding events that aren't already included in the thunder storm or tropical storm categories:

```r
regex <- "WET|[Ww]et|[Ff]lood|FLOOD|[Uu]rban|URBAN|[Ss]tream|STREAM|[Rr]ain|RAIN|PREC|[Pp]rec"
flood <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

Next, we want all the wind related weather events which aren't already part of one of the previously generated groups:

```r
regex <- "WIND|[Ww]ind|[Gg]ust|GUST"
wind <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

For drought related events we match "dry", "drought" and "dust":

```r
regex <- "Drought|DROUGHT|Dry|DRY|DUST|Dus|dry"
dry <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

For the group of extreme heat events, we match "high temp", "warm", and "heat":

```r
regex <- "WARM|[Ww]arm|[Hh]eat|HEAT|[Hh]OT|HIGH TEMP|[Hh]igh [Tt]emp"
heat <- grep(regex, eventnames, value = TRUE)
eventnames <- eventnames[!grepl(regex, eventnames)]
```

Finally, we want to capture of the coastal and oceanic related weather events, including currents, tides, waves, and surges:

```r
regex <- "TIDE|[Tt]ide|SURGE|[Ss]urge|[Cc]urrent|CURRENT|TSUNAMI|[Tt]sunami|WAVE|[Ww]ave|SURF|[Ss]urf|[S
coastal <- grep(regex, eventnames, value = TRUE)
```

The remaining events are stored in the "other" group, these include events such as hail storms and volcanic events. We see that around 100 event types fall into the other group. Hence, 800 of the 900 different event types fall into one of our groups.

```r
other <- eventnames[!grepl(regex, eventnames)]
length(other)
```

```
## [1] 107
```

We use these groups to form a data frame containing total injuries/fatalites corresponding to each group, as well as total count of events corresponding to at least one fatality/injury. We extract out the subsets of our data corresponding to those event types which fall into the above groups, and sum up the total number of injuries/fatalities. We put these values into a data.frame, along with a variable indicating the total number of individual events involving a death or fatality.

```
eventnames = c("coastal","dry","flood","heat","other","tornado","tropical","tstorm","wildfire","wind","w
eventlist = list(coastal, dry, flood, heat, other, tornado, tropical, tstorm, wildfire, wind, winter)
injury <- sapply(eventlist, function(x){
    temp <- subset(injuryevents, EVTYPE %in% x)
    sum(temp$INJURIES)
})
injtotal <-sapply(eventlist, function(x){
    sum(injuryevents$EVTYPE %in% x)
})
fatal <- sapply(eventlist, function(x){
    temp <- subset(fatalevents, EVTYPE %in% x)
    sum(temp$FATALITIES)
})
deathtotal <-sapply(eventlist, function(x){
    sum(fatalevents$EVTYPE %in% x)
})
healtheffects <- data.frame(names = eventnames, injuries = injury, inj.events = injtotal, avg.inj = inju
                            fat.events = deathtotal, avg.fat = fatal/deathtotal)
```

To determine economic damage, we consider the damage totals for both the crops and property. We subset our data to include only the events where nonzero crop loss or property damage occured.

```
croploss <- subset(NOAA, CROPDMG > 0)
propdamage <- subset(NOAA, PROPDMG > 0)
```

Next, we sum the total losses for events falling into the groups which we determined previously. We also include a count of the total number of events which result in damages for each event group. We form a data frame from these totals, which we will use for our analysis. We add one more event type group, for our damages analysis, containing the hail events:

```
regex <- "[Hh]ail|HAIL"
hail <- grep(regex, other, value = TRUE)
other <- other[!grepl(regex, other)]
eventnames = c("coastal","dry","flood","hail", "heat", "other","tornado","tropical","tstorm","wildfire"
eventlist = list(coastal, dry, flood, hail, heat, other, tornado, tropical, tstorm, wildfire, wind, win
```

We use the groups to summary the economic damage data:

```
croptotal <- sapply(eventlist, function(x){
    temp <- subset(croploss, EVTYPE %in% x)
    sum(temp$CROPDMG)
})
cropevent <-sapply(eventlist, function(x){
    sum(croploss$EVTYPE %in% x)
})
proptotal  <- sapply(eventlist, function(x){
    temp <- subset(propdamage, EVTYPE %in% x)
    sum(temp$PROPDMG)
})
propevents <-sapply(eventlist, function(x){
    sum(propdamage$EVTYPE %in% x)
})
economicloss <- data.frame(Names = eventnames, crop.loss = croptotal, crp.evnts = cropevent, avg.crp = c
```

```
economicloss
```

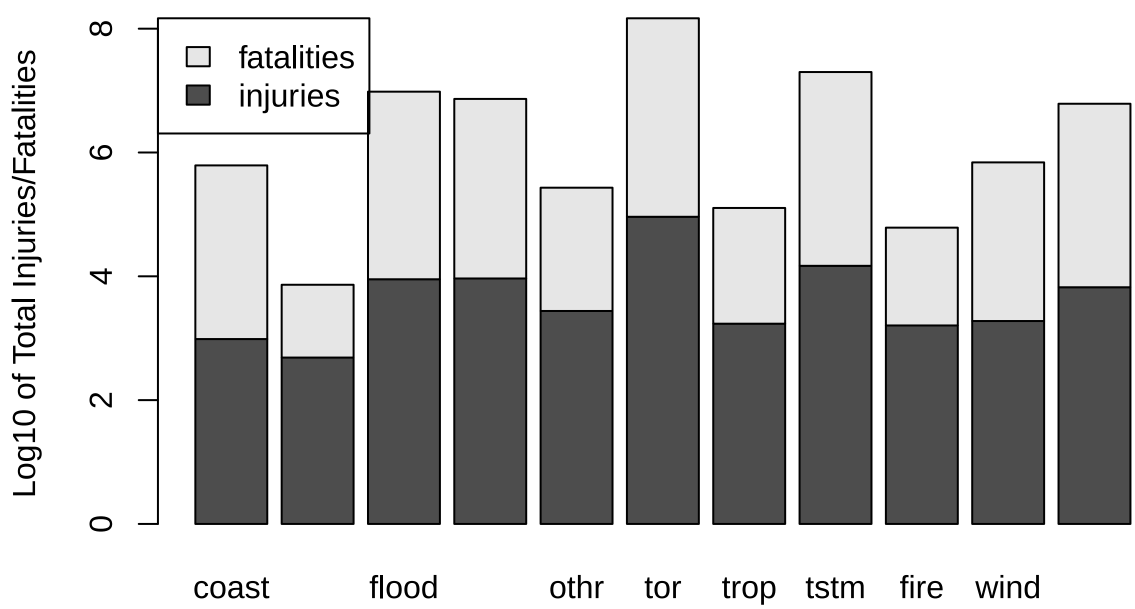
```
##       Names crop.loss crp.evnts   avg.crp   prop.dmg prp.evnts   avg.prp
## 1   coastal    875.00         4 218.75000   35781.96       358  99.94961
## 2       dry  35570.90       271 131.25793   11106.15       199  55.80980
## 3     flood 379849.83      4395  86.42772 2514905.24     33234  75.67266
## 4      hail 581418.36      9389  61.92548  689306.78     23062  29.88929
## 5      heat   1427.40        19  75.12632    3032.86        43  70.53163
## 6     other   1071.40        32  33.48125   41592.10       393 105.83232
## 7   tornado 100029.27      1502  66.59738 3226251.94     39217  82.26667
## 8  tropical  18102.91       163 111.06080   75857.33       654 115.98980
## 9    tstorm 202970.99      5556  36.53186 3281772.14    126700  25.90191
## 10 wildfire   9565.74       130  73.58262  125323.29      1057 118.56508
## 11     wind  21808.91       333  65.49222  453069.89      9391  48.24512
## 12   winter  25136.61       305  82.41511  426500.33      4866  87.64906
```

Results

Injuries and Fatalities by Event Type

Since the total fatalities/injuries vary by orders of magnitude, we take the log of the total injury/fatality counts:

```
barplot(t(cbind(log10(injury),log10(deathtotal))), names.arg = c("coast", "dry", "flood", "heat", "othr
```



We can see that the largest number of reported injuries and fatalities are due to tornado related weather events.

Flooding, winter events, heat-related events, and thunderstorms also have relatively large fatality/injury counts. Also, our the "other" group accounts for a relatively small portion of the remaining deaths/injuries. We can also look at the data frame directly:

```
healtheffects
```

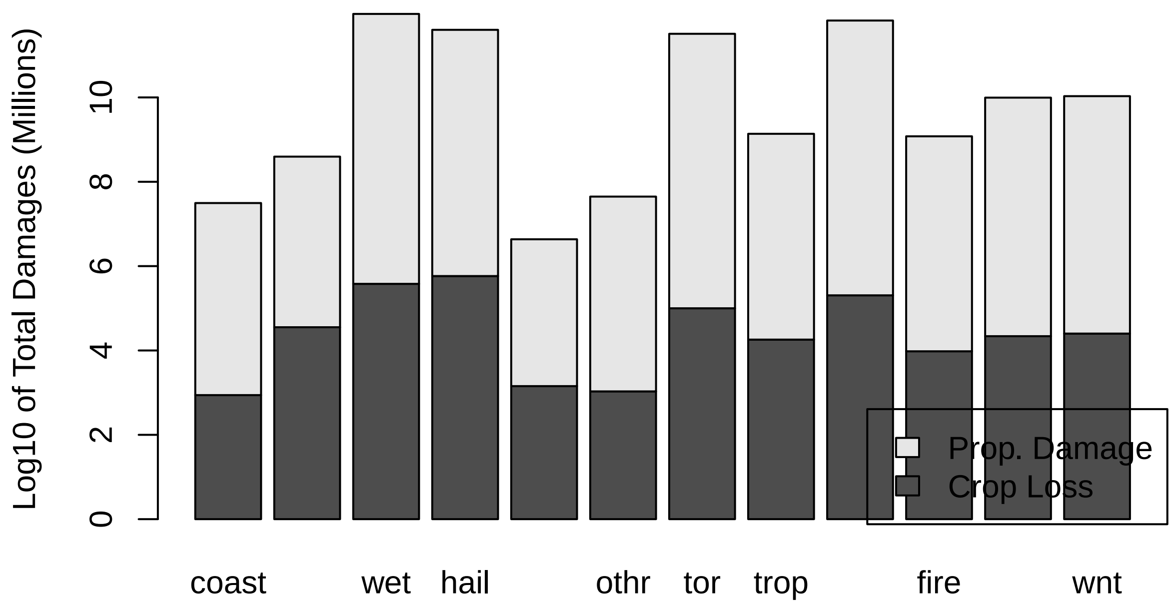
```
##       names injuries inj.events   avg.inj fatal fat.events  avg.fat
## 1   coastal      969        301  3.219269   818        638 1.282132
## 2       dry      487         57  8.543860    60         15 4.000000
## 3     flood     8961        663 13.515837  1654       1072 1.542910
## 4      heat     9228        232 39.775862  3143        795 3.953459
## 5     other     2756        441  6.249433   167         98 1.704082
## 6   tornado    91482       7728 11.837733  5667       1609 3.522063
## 7  tropical     1716         58 29.586207   201         74 2.716216
## 8    tstorm    14775       6468  2.284323  1547       1349 1.146775
## 9  wildfire     1608        316  5.088608    90         38 2.368421
## 10     wind     1896        699  2.712446   466        365 1.276712
## 11   winter     6650        641 10.374415  1332        921 1.446254
```

By inspecting the numbers directly we see that tornado events are clearly ahead of the other groups in terms of injury/death counts and also in terms of total number of events. Thunder storm events also account for a large number of deaths/injurys and total events. We see that tropical storms, tornadoes and heat events all have high average injury/death totals per event.

Economic Losses by Event Type

Since the total damages can vary by orders of magnitude across the event groups, we use a barplot with a logarithmic scale similar to one we made for fatalities and injuries to view the total damages:

```
barplot(t(cbind(log10(croptotal),log10(proptotal))), names.arg = c("coast", "dry", "wet", "hail", "heat
```



From this we can see that the hail, flooding, tornadoes, thunder storms corresponding to largest groups in terms of total damages. We also inspect the data frame directly.

```
economicloss
```

```
##       Names crop.loss crp.evnts   avg.crp   prop.dmg prp.evnts   avg.prp
## 1   coastal    875.00         4 218.75000   35781.96       358  99.94961
## 2       dry  35570.90       271 131.25793   11106.15       199  55.80980
## 3     flood 379849.83      4395  86.42772 2514905.24     33234  75.67266
## 4      hail 581418.36      9389  61.92548  689306.78     23062  29.88929
## 5      heat   1427.40        19  75.12632    3032.86        43  70.53163
## 6     other   1071.40        32  33.48125   41592.10       393 105.83232
## 7   tornado 100029.27      1502  66.59738 3226251.94     39217  82.26667
## 8  tropical  18102.91       163 111.06080   75857.33       654 115.98980
## 9    tstorm 202970.99      5556  36.53186 3281772.14    126700  25.90191
## 10 wildfire   9565.74       130  73.58262  125323.29      1057 118.56508
## 11     wind  21808.91       333  65.49222  453069.89      9391  48.24512
## 12   winter  25136.61       305  82.41511  426500.33      4866  87.64906
```

Our hail group is perhaps most notable in the number of events and total damages, in the previous analysis we considered hail events among the "other" group which accounts for a relatively small portion of the remaining totals. However, hail accounts for the largest portion of crop loss, and a considerable portion of overall damages. Unsurprisingly, flooding and thunderstorms account for relatively large portions oftotal damages. Wildfire and tropical storms, although relatively rare have high average losses per event.

Closing Remarks

From this simple analysis, we see that floording, tornadoes and thunder storms account for large numbers of deaths and injuries as well as economic damage. Extreme heat events are particularly dangerous to public health, while less significant for damages to property and crop loss. Hail events cause significant damages to crops and property, however do not seem to contribute much to the total injury/fatalities even if groups in with other miscellaneous events.