Analysis of severe weather events on public health and economic consequences across United States

Maverix13

September 21, 2015

Synopsis

This report explores the NOAA Storm Database and answer some basic questions about severe weather events. The analysis covers:

- Across United States, events that are most harmful to public health.
- Across United States, events that have greatest economic consequences.

Results, covered later in the document, shows that 50% of the fatalities are caused by Extreme Temperature and Convection related weather events while combined with Flood and Wind accounts for about 80% of fatalities. Similarly, Convection and Extreme Temperatures cause about 60% of injuries, which combined with combined with Flood and Wind causes over 80% injuries. Economic impact of weather related activities is mostly caused by Flood and Cyclone resulting in over $250 billion. Property damage accounts for over 90% of the economic impact.

Data Processing

Data is downloaded, if required. Code checks if data already exists. Data is then loaded into R. Note that a zip file is directly read using read.csv. There is no need to unzip it as read.csv will handle it.

```
#install.packages("devtools")
#devtools::install_github("renkun-ken/formattable")
library(knitr)
library(dplyr)
library(ggplot2)
library(formattable)
library(reshape2)
require(gridExtra)
```

```
if(!file.exists("data")){
  dir.create("data")
}
if(!file.exists("data/stormData.csv.bz2")) {
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "data/storml

}

stormData <- read.csv("data/stormData.csv.bz2")
dim(stormData)
```

```
## [1] 902297     37
```

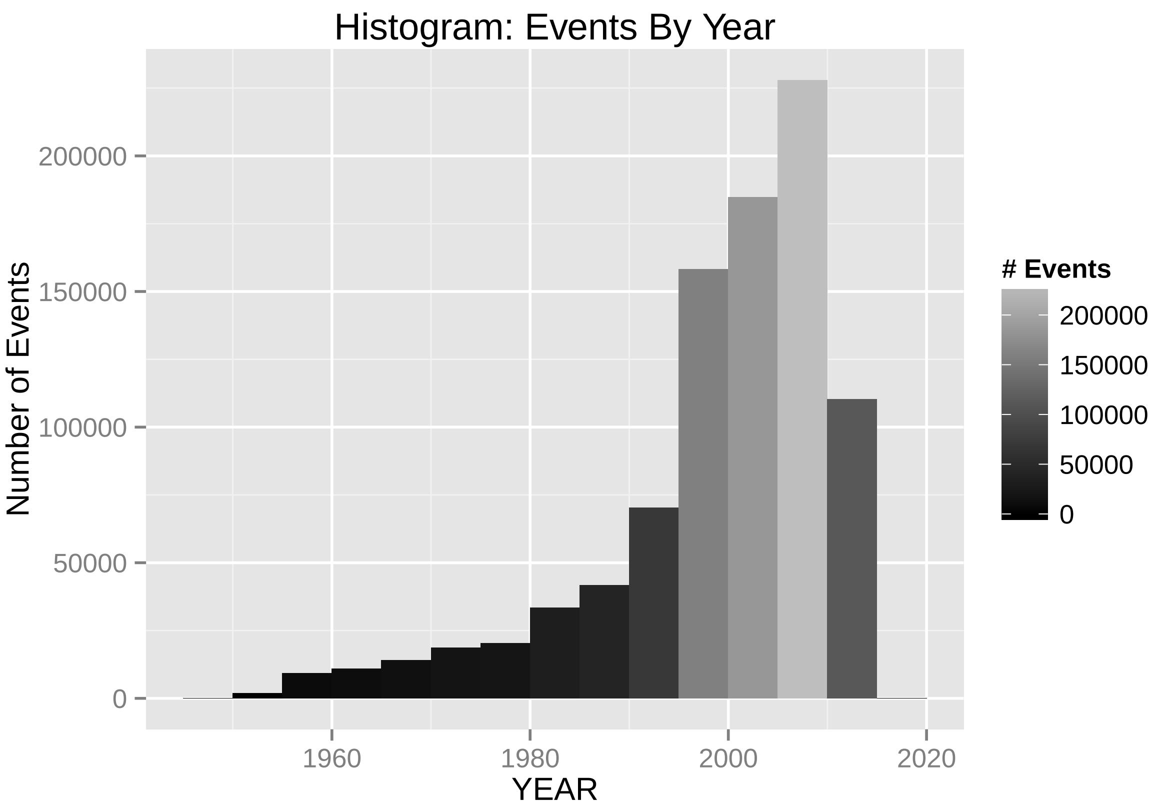
```r
head(stormData[,2:28])
```

```
##             BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES
## 1         0                      14.0   100 3   0          0       15
## 2         0                       2.0   150 2   0          0        0
## 3         0                       0.1   123 2   0          0        2
## 4         0                       0.0   100 2   0          0        2
## 5         0                       0.0   150 2   0          0        2
## 6         0                       1.5   177 2   0          0        6
##   PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1    25.0          K       0
## 2     2.5          K       0
## 3    25.0          K       0
## 4     2.5          K       0
## 5     2.5          K       0
## 6     2.5          K       0
```

Preprocessing Data

As the histogram below shows, in the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records.

```r
stormData <- mutate(stormData, YEAR = as.numeric(format(as.Date(BGN_DATE, format = "%m/%d/%Y %H:%M:%S")
g <- ggplot(stormData, aes(YEAR, fill = ..count.., removePanelGrid=TRUE,removePanelBorder=TRUE)) +
  geom_histogram(binwidth=5) +
  ylab("Number of Events") +
  scale_fill_gradient("# Events", low = "black", high = "gray") +
  ggtitle("Histogram: Events By Year")
print(g)
```



Hence for this report, data after 1996 is considered for analysis.

```
stormData <- stormData[stormData$YEAR >= 1996,]
dim(stormData)
```

```
## [1] 653530     38
```

Following columns are of relevance for the current analysis

Column Name	Description
EVTYPE	Type of storm event. Take note that similar storm events can be listed using different wording e.g. "coa
FATALITIES	Number directly killed
INJURIES	Number directly injured
PROPDMG	Property damage in whole numbers and hundredths
PROPDMGEXP	A multiplier where Hundred (H), Thousand (K), Million (M), Billion (B)
CROPDMG	Crop damage in whole numbers and hundredths
CROPDMGEXP	A multiplier where Hundred (H), Thousand (K), Million (M), Billion (B)

This above table is constructed based on information from http://ire.org/nicar/database-library/databases/storm-events/

Further reduction of stormData based on columns and row values

```
## Just keeping relevant columns
relevantColumns <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXI
stormData <- stormData[,relevantColumns]
```

```r
## Keep rows having any data
stormData <- subset(stormData, FATALITIES > 0 | INJURIES > 0 | PROPDMG > 0 | CROPDMG > 0)
```

Mapping of Events to Group Many events can be grouped together based on similarity. The categorization for this report is based on NOAA 2009 Annual Summaries (http://www.ncdc.noaa.gov/oa/climate/sd/annsum2009.pdf).

```r
stormData$EVGROUP <- "Others"
stormData[grepl("tornado|thunderstorm|hail|funnel|tstm|lig(.*?)ing",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Convection"
stormData[grepl("temperature|cold|cool|heat|hot|fire|dry|warm|wet|freez",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Extreme Temperature"
stormData[grepl("flood|rain|mud(.*?)slid",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Flood"
stormData[grepl("marine|sea|tide|coast|tsunami|current|surf|wave",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Marine"
stormData[grepl("tropic|cyclone|hurricane",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Cyclone"
stormData[grepl("avalanche|snow|blizzard|winter|wintry|ice",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Winter"
stormData[grepl("wind|dust",
                stormData$EVTYPE, ignore.case = TRUE), "EVGROUP"] <- "Wind"

stormData$EVGROUP <- as.factor(stormData$EVGROUP)
```

Analysis on public health

Analysis of fatalities and injuries caused by a given storm event group.

```r
pubHealthData <- stormData %>%
  group_by(EVGROUP) %>%
  summarize(fatalities = sum(FATALITIES), injuries = sum(INJURIES))
kable(pubHealthData)
```

EVGROUP	fatalities	injuries
Convection	2170	25533
Cyclone	182	1661
Extreme Temperature	2267	9202
Flood	1404	8670
Marine	751	989
Others	160	1302
Wind	1034	7127
Winter	764	3491

Analysis on Economic Impact

Analysis of property and crop damage caused by a given event group.

```r
unique(stormData$PROPDMGEXP)
```

```
## [1] K   M B
## Levels:  - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

```r
unique(stormData$CROPDMGEXP)
```

```
## [1] K   M B
## Levels:  ? 0 2 B k K m M
```

Unique values for both property and crop damage exponents is [‘K’, ‘M’, ‘B’]. These values are mulitpliers as mentioned above in column descriptions.

```r
economicData <- stormData %>%
  mutate(PROPDMGVAL = PROPDMG * (
    ifelse(PROPDMGEXP == "K", 1000,
           ifelse(PROPDMGEXP == "M", 1000000,
                  ifelse(PROPDMGEXP == "B", 1000000000,
                         1  )))))  %>%
  mutate(CROPDMGVAL = CROPDMG * (
    ifelse(CROPDMGEXP == "K", 1000,
           ifelse(CROPDMGEXP == "M", 1000000,
                  ifelse(CROPDMGEXP == "B", 1000000000,
                         1  ))))) %>%
  mutate(TOTALDMG = PROPDMGVAL + CROPDMGVAL) %>%
  group_by(EVGROUP) %>%
  summarize(propertyDamage = sum(PROPDMGVAL), cropDamage = sum(CROPDMGVAL), totalDamage = sum(TOTALDMG))
  arrange(desc(totalDamage))
economicData$EVGROUP <- factor(economicData$EVGROUP, levels=economicData$EVGROUP)
formattable(economicData, list(
  propertyDamage=currency,
  cropDamage=currency,
  totalDamage=currency
))
```

EVGROUP	propertyDamage	cropDamage	totalDamage
Flood	$159,904,859,000.00	$7,067,994,900.00	$166,972,853,900.00
Cyclone	$88,762,871,560.00	$6,026,993,800.00	$94,789,865,360.00
Others	$45,253,732,050.00	$13,439,935,500.00	$58,693,667,550.00
Convection	$39,955,370,310.00	$2,787,145,900.00	$42,742,516,210.00
Wind	$13,369,105,350.00	$1,736,767,400.00	$15,105,872,750.00
Extreme Temperature	$7,812,842,200.00	$3,530,582,630.00	$11,343,424,830.00
Winter	$6,413,677,850.00	$120,786,100.00	$6,534,463,950.00
Marine	$5,295,157,060.00	$42,522,500.00	$5,337,679,560.00

Results

Impact on public health

```
fData <- pubHealthData %>%
  arrange(desc(fatalities)) %>%
  select(EVGROUP, fatalities) %>%
  mutate(
    cumulativeSum = cumsum(fatalities),
    cumPercent = cumulativeSum/sum(fatalities)
  )
formattable(fData, list(
  cumPercent = percent
))
```

EVGROUP	fatalities	cumulativeSum	cumPercent
Extreme Temperature	2267	2267	25.96%
Convection	2170	4437	50.81%
Flood	1404	5841	66.89%
Wind	1034	6875	78.73%
Winter	764	7639	87.48%
Marine	751	8390	96.08%
Cyclone	182	8572	98.17%
Others	160	8732	100.00%

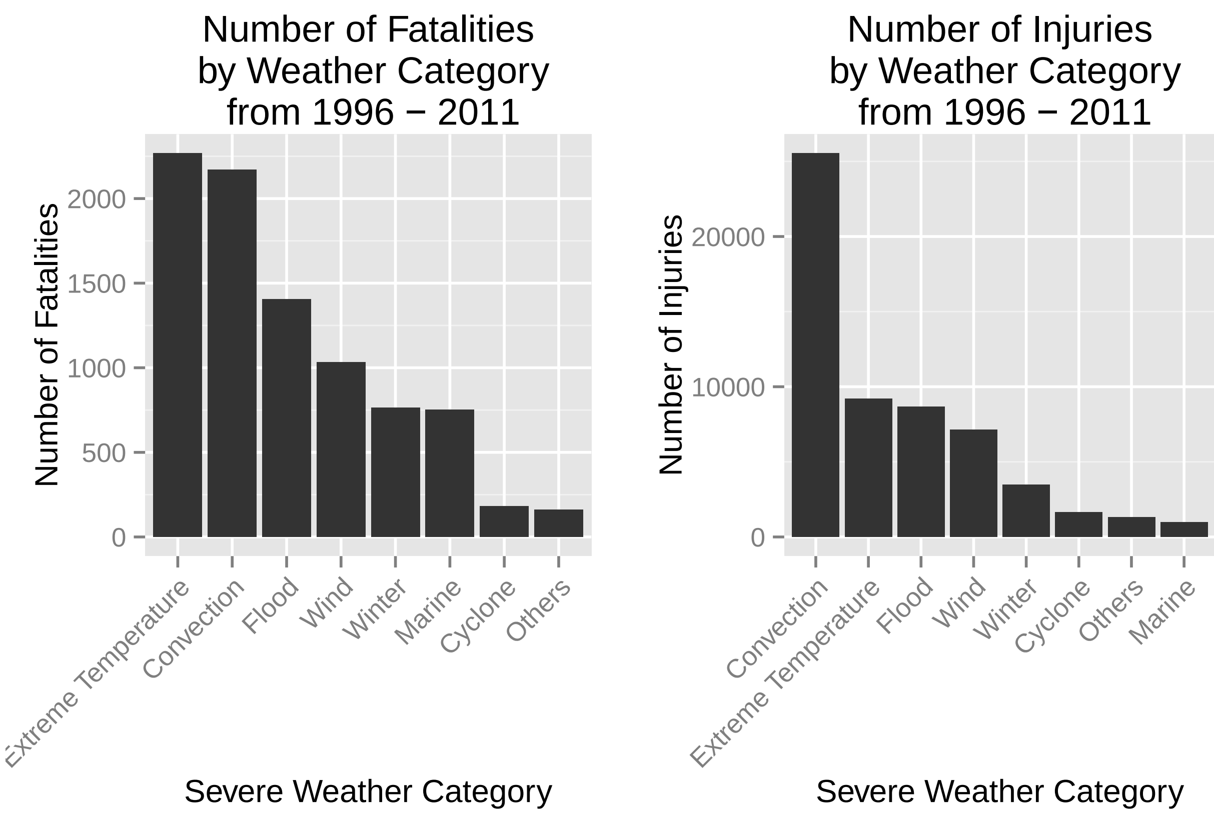
Table above shows fatalities along with a cumulative percent. Extreme Temperature, Convection, Flood and Wind together cause over 7500 fatalities which is close to 80% of total fatalities.

```
iData <- pubHealthData %>%
  arrange(desc(injuries)) %>%
  select(EVGROUP, injuries) %>%
  mutate(
    cumulativeSum = cumsum(injuries),
    cumPercent = cumulativeSum/sum(injuries)
  )
formattable(iData, list(
  cumPercent = percent
))
```

EVGROUP	injuries	cumulativeSum	cumPercent
Convection	25533	25533	44.04%
Extreme Temperature	9202	34735	59.91%
Flood	8670	43405	74.87%
Wind	7127	50532	87.16%
Winter	3491	54023	93.18%
Cyclone	1661	55684	96.05%
Others	1302	56986	98.29%
Marine	989	57975	100.00%

As with fatalities, injuries are also mostly caused by Convection, Extreme Temperature, Flood and Wind resulting in over 50,000 injuries.

```
fPlot <- ggplot(fData, aes(x=reorder(EVGROUP, desc(fatalities)))) +
  geom_bar(aes(y=fatalities), stat = "identity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  xlab("Severe Weather Category") +
  scale_y_continuous("Number of Fatalities") +
  ggtitle("Number of Fatalities\n by Weather Category\n from 1996 - 2011")
iPlot <- ggplot(iData, aes(x=reorder(EVGROUP, desc(injuries)))) +
  geom_bar(aes(y=injuries), stat = "identity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  xlab("Severe Weather Category") +
  scale_y_continuous("Number of Injuries") +
  ggtitle("Number of Injuries\n by Weather Category\n from 1996 - 2011")
grid.arrange(fPlot, iPlot, ncol = 2)
```



Based on the graphs above, we see that there are lot more injuries compared to fatalities but the top 2 weather event are Extreme Tmperature and Convection.

Impact on Economy

```
pData <- economicData %>%
  arrange(desc(propertyDamage)) %>%
  select(EVGROUP, propertyDamage) %>%
  mutate(
```

```
    cumulativeSum = cumsum(propertyDamage),
    cumPercent = cumulativeSum/sum(propertyDamage)
  )
formattable(pData, list(
  propertyDamage=currency,
  cumPercent = percent
))
```

EVGROUP	propertyDamage	cumulativeSum	cumPercent
Flood	$159,904,859,000.00	159904859000	43.60%
Cyclone	$88,762,871,560.00	248667730560	67.80%
Others	$45,253,732,050.00	293921462610	80.14%
Convection	$39,955,370,310.00	333876832920	91.03%
Wind	$13,369,105,350.00	347245938270	94.68%
Extreme Temperature	$7,812,842,200.00	355058780470	96.81%
Winter	$6,413,677,850.00	361472458320	98.56%
Marine	$5,295,157,060.00	366767615380	100.00%

Table above shows that over $250 billion in prpoerty damage is caused by Flood and Cyclone.

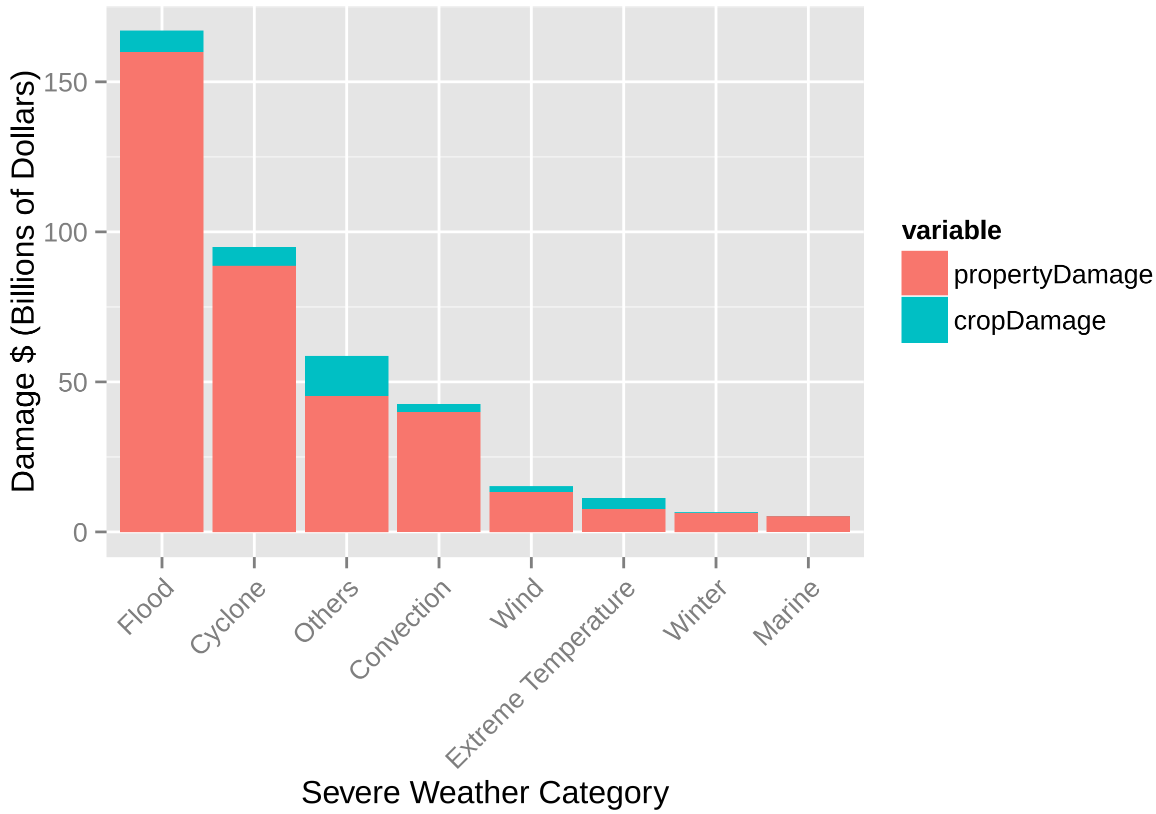
```
cData <- economicData %>%
  arrange(desc(cropDamage)) %>%
  select(EVGROUP, cropDamage) %>%
  mutate(
    cumulativeSum = cumsum(cropDamage),
    cumPercent = cumulativeSum/sum(cropDamage)
  )
formattable(cData, list(
  cropDamage=currency,
  cumPercent = percent
))
```

EVGROUP	cropDamage	cumulativeSum	cumPercent
Others	$13,439,935,500.00	13439935500	38.67%
Flood	$7,067,994,900.00	20507930400	59.01%
Cyclone	$6,026,993,800.00	26534924200	76.35%
Extreme Temperature	$3,530,582,630.00	30065506830	86.51%
Convection	$2,787,145,900.00	32852652730	94.53%
Wind	$1,736,767,400.00	34589420130	99.53%
Winter	$120,786,100.00	34710206230	99.88%
Marine	$42,522,500.00	34752728730	100.00%

Comparing crop damage to property damage shows that crop damage does not have as significant economic impact as property damage. Crop damage totaled close to $34 billion across all weather events. Hence the chart below takes total property damage into account.

```
meltData <- melt(economicData[,c('EVGROUP','propertyDamage','cropDamage')], id.var="EVGROUP")
ePlot <- ggplot(meltData, aes(x = EVGROUP, y = value/1000000000, fill = variable)) +
  geom_bar(stat = "identity") +
```

```
  xlab("Severe Weather Category") +
  scale_y_continuous("Damage $ (Billions of Dollars)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(ePlot)
```



Majority of economic impact is attributed to property damage caused by Flood and Cyclone.