Impact of Weather Events on Population Health and Economic Losses

An Analysis of NOAA's Storm Events Database

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. This report provides an analysis as to which weather events effect both aspects across the United States. It was found that "heat" related events have caused the most fatalities, followed by "tornado" and "flood" related events. "Flash flood" had the highest occurrences among the top 5 fatality events. In respect of economic losses, it was found that "flood" events have caused the most property damage, followed by "hurricane/typhoon". "Drought" events, on the other hand, have caused the most crop damage, followed by "flood". However, it was observed that the property damage caused by flood was significantly higher than hurricane/typhoon. Considering the recent events such as Hurricane Sandy and Hurricane Katrina, a further investigation would be necessary to see if any recording errors have occurred.

The NOAA's storm events database collected records from 1950 to 2011. Due to lack of diverse event types, records prior to 1993 were excluded in this analysis.

Data Source

Database

The dataset was obtained from the course web site (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2). It came in the form of a comma-separated-value file compressed in the bzip2 algorithm.

file name: "repdata-data-StormData.csv.bz2"

Storm Data Event Table

The standard 48 event types were obtained from NWS Directive 10-1605 (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) page6, section 2.1.1

This NOAA webpage (http://www.ncdc.noaa.gov/stormevents/details.jsp?type=eventtype) provided a timeline of how event type records have progressed over the years.

Consumer Price Index

The data ranged from 1950 to 2011, therefore inflation must be considered. In this analysis, all dollar amounts were adjusted with the U.S. Consumer Price Index(CPI).

CPI was download from BUREAU of Labor Statistics (http://www.bls.gov/data/) :

series ID CUUR0000SA0 (Not Seasonally Adjusted. Area: U.S. city average. Item: All items. Base Period: 1982-84=100).

The downloaded file was in MS Excel (.xlsx) format and then saved as a comma-separated-value file for R coding.

Because obtaining CPI data required website "point-and-click" by hand, the detailed process were documented in this github repository (https://github.com/blackszu/NOAA_Storm_DataAnalysis/blob/gh-pages/Getting%20CPI%20data%20from%20BUREAU%20of%20Labor%20Statistics.pdf)

file name: "cpi_1950_2011.csv", "Getting CPI data from BUREAU of Labor Statistics.pdf"

```r
# Check and load required R packages
pkg<-c("knitr", "ggplot2", "grid", "dplyr", "tidyr")
pkgCheck<-pkg %in% rownames(installed.packages())
for(i in 1:length(pkg)) {
    if(pkgCheck[i]==FALSE) {
        install.packages(pkg[i])
    }
    library(pkg[i],character.only = TRUE)
}
```

Data Processing

Loading Raw Data

```r
con<-bzfile("repdata-data-StormData.csv.bz2", "r")
storm<- read.csv(con)
close(con)
```

```r
#set table and figure counter
tnum<- 0L
fnum<- 0L
```

Variables

```r
colnames(storm)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Among 37 *variables* in the dataset. We were interested in only the following for the intended analysis:

1. event type: *EVTYPE*

2. population health : *FATALITIES / INJURIES*

3. economic consequences: *PROPDMG / CROPDMG* for "number", *PROPDMGEXP / CROPDMGEXP* for "magnitude". The damage dollar amount would be "number" x "magnitude"

4. Others that may be useful for analysis: *STATE / BGN_DATE*

```
str(storm[, c(2,7,8,23,24:28)])
```

```
## 'data.frame':    902297 obs. of  9 variables:
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224
2224 2260 383 3980 3980 ...
##  $ 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 83
4 834 834 834 834 ...
##  $ 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 ...
```

```
# convert string to Date format
storm$BGN_DATE<- strptime(storm$BGN_DATE, "%m/%d/%Y %H:%M:%S")
storm$YEAR<- storm$BGN_DATE$year+1900
```

Cleaning and Processing Data

1. Property and Crop Damage

1.1 Check data quality

As stated in NWS Directive 10-1605 (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) page12, section 2.7 Damage: "…Alphabetical characters used to signify magnitude include"K" for thousands, "M" for

millions, and "B" for billions…"

However, the dataset had many magnitudes that seemed invalid:

```
levels(storm$PROPDMGEXP) #property damage
```

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

```
levels(storm$CROPDMGEXP) # crop damage
```

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

1.2 Examine the invalid damage magnitudes

It was assumed that "H" and "h" signified "hundred" and thus included it as a "valid" magnitudes along with "K", "M", and "B". Below codes examined how many events were marked with invalid damage magnitudes. All letter magnitudes were unified to lower cases.

```
# assume h="hundred". use lower cases
mag<-c("h", "k", "m", "b")

# (inline codes) check invalid magnitudes with numbers(ie. PROPDMG>0 or CROPDMG>0)
invalidP<- which(!tolower(storm$PROPDMGEXP) %in% mag & storm$PROPDMG>0)
invalidPROPDMG<- list(Count=length(invalidP),
                    Count.Perc=length(invalidP)/nrow(storm[storm$PROPDMG>0,]),
                    Number.Perc=sum(storm$PROPDMG[invalidP])/sum(storm$PROPDMG),
                    Count.byYear=with(storm[invalidP, ], table(tolower(PROPDMGEXP), YEA
R)))

invalidC<- which(!tolower(storm$CROPDMGEXP) %in% mag & storm$CROPDMG>0)
invalidCROPDMG<- list(Count=length(invalidC),
                    Count.Perc=length(invalidC)/nrow(storm[storm$CROPDMG>0,]),
                    Number.Perc=sum(storm$CROPDMG[invalidC])/sum(storm$CROPDMG),
                    Count.byYear=with(storm[invalidC, ], table(tolower(CROPDMGEXP), YEA
R)))
```

Number of events with invalid damage magnitude by year, Property and Crop:

```
invalidPROPDMG$Count.byYear
```

```
##      YEAR
##       1993 1994 1995
##          6    3   67
##    -     0    0    1
##    +     0    1    4
##    0     1   25  183
##    2     0    0    1
##    3     0    0    1
##    4     0    1    3
##    5     0    1   17
##    6     0    0    3
##    7     0    0    2
```

```
invalidCROPDMG$Count.byYear
```

```
##      YEAR
##       1994 1995
##          3    0
##    0     8    4
```

Among all events with a property damage number record (ie. PROPDMG > 0), ~0.1% (320 events) were marked with an invalid magnitude, and all of them were before year 1995.

Among all the events with a crop damage number record (ie. CROPDMG > 0), ~0.1% (15 events) were marked with an invalid magnitude, and all of them were before year 1995.

Rather than assuming (guessing) how to correct the invalid damage magnitudes events, we decided to ignore them since they had very minimum impact to the intended analysis.

1.3 Calculate damage dollar amount

Calculated property and crop damage dollar amount in **Millions** for each event:

- if number>0 and magnitude was valid, the dollar amount would be calculated by multiplying the two

- if the number=0, the dollar amount would be 0 as well

- if the magnitude was invalid, the dollar amount would be "NA"

```
# set multipliers
mag10<-10^c(2, 3, 6, 9)

storm$PROPDMG.M<- storm$PROPDMG * mag10[match(tolower(storm$PROPDMGEXP), mag)]/10^6
storm$CROPDMG.M<- storm$CROPDMG * mag10[match(tolower(storm$CROPDMGEXP), mag)]/10^6

#if damage number=0, its dollar amount should be also 0, not NA
storm$PROPDMG.M[storm$PROPDMG==0]<- 0
storm$CROPDMG.M[storm$CROPDMG==0]<- 0
```

1.4 Adjustment for inflation

```
# read cpi file, use the "Annual" column
cpi<- read.csv("cpi_1950_2011.csv", skip=10, colClasses=c(NA,rep("NULL",12),NA,"NULL","NU
LL"))

# index for converting to present value
cpi<- transform(cpi, pv=Annual[Year==2011]/Annual)

# calculate adjusted dollar amount
storm<- transform(storm, adjPROPDMG.M=round(PROPDMG.M * cpi[match(YEAR, cpi$Year),"pv"],
0),
                  adjCROPDMG.M=round(CROPDMG.M * cpi[match(YEAR, cpi$Year),"pv"],0))
```

2. Event Type Grouping

2.1 Check data quality

```
# EVTYPE factor levels
nlevels(storm$EVTYPE)
```

```
## [1] 985
```

```
# event types with equal or less than 5 & only 1 record in the database
list(five.or.less= sum(table(storm$EVTYPE)<6),
     only.one= sum(table(storm$EVTYPE)==1))
```

```
## $five.or.less
## [1] 760
##
## $only.one
## [1] 489
```

Among the dataset's 985 event types, many had only 5 or less records, and almost half had only one record. It would be beneficial to the analysis to group them into the standard 48 event types defined in NWS Directive 10-1605 (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf). Below 2.2 to 2.5 sections show how the event-matching was achieved.

2.2 Get standard event type table

In keeping with the principle of "Don't do things by hand", we import the event table from the NWS Directive 10-1605 (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) document.

```
url <- "http://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf"
dest <- tempfile(fileext = ".pdf")
download.file(url, dest, mode = "wb")

# set path to pdftotxt.exe and convert pdf to text
exe <- "C:\\Program Files\\Git\\mingw64\\binpdftotext.exe"
system(paste("\"", exe, "\" \"", dest, "\"", sep = ""), wait = F)

# get txt-file name and rename it
filetxt <- sub(".pdf", ".txt", dest)
fsave<- ".\\NWS Directive 10-1605.txt"
file.rename(filetxt, fsave)
evtb<- readLines(fsave)
```

```
t<-grep("Event Name", evtb)[1] # get the line where the event type table begins
evtb<-tolower(evtb[c((t+1):(t+24), (t+29):(t+52))])
evtb[25]<- "hurricane/typhoon"  #change from "hurrican (typhoon)"

#. move the order of the following event types up in the hierachy: "Flood", "Heat", "stro
ng wind", "thunderstorm wind"
evtb<-evtb[c(15,20,38,39,1:14,16:19,21:37, 40:48)]
print(evtb)
```

```
##  [1] "flood"                   "heat"
##  [3] "strong wind"             "thunderstorm wind"
##  [5] "astronomical low tide"   "avalanche"
##  [7] "blizzard"                "coastal flood"
##  [9] "cold/wind chill"         "debris flow"
## [11] "dense fog"               "dense smoke"
## [13] "drought"                 "dust devil"
## [15] "dust storm"              "excessive heat"
## [17] "extreme cold/wind chill" "flash flood"
## [19] "frost/freeze"            "funnel cloud"
## [21] "freezing fog"            "hail"
## [23] "heavy rain"              "heavy snow"
## [25] "high surf"               "high wind"
## [27] "hurricane/typhoon"       "ice storm"
## [29] "lake-effect snow"        "lakeshore flood"
## [31] "lightning"               "marine hail"
## [33] "marine high wind"        "marine strong wind"
## [35] "marine thunderstorm wind" "rip current"
## [37] "seiche"                  "sleet"
## [39] "storm surge/tide"        "tornado"
## [41] "tropical depression"     "tropical storm"
## [43] "tsunami"                 "volcanic ash"
## [45] "waterspout"              "wildfire"
## [47] "winter storm"            "winter weather"
```

2.3 Group (match) database events with the standard 48 event types

By referencing at the variable *EVTYPE*, some general discrepancies were noticed and modified before the main event-matching task as below :

```r
# transform database event type factors to lower cases
f<- tolower(levels(storm$EVTYPE))

# thunderstorm wind
f<- gsub("tstm","thunderstorm wind", f)
# NOAA glossary(http://forecast.weather.gov/glossary.php?word=tstm), tstm=thunderstorm
f<- gsub("thunderstorm","thunderstorm wind", f)
f<- gsub("burst","thunderstorm wind", f)

# from NWS Directive  10-1605 page 1 Summary of Revisions: "Landslide" was renamed to "De
bris Flow"
f<- gsub("landslide","debris flow", f)

# hurricane/typhoon
f<- gsub("^typhoon","hurricane/typhoon", f)
f<- gsub("hurricane","hurricane/typhoon", f)

# others/typos
f<- gsub("wintery","winter", f)
f<- gsub("wintry","winter", f)
f<- gsub("avalance","avalanche", f)
f<- gsub("fld","flood", f)
```

Main event-matching task using for-loop. Events that failed to match were grouped as "others" :

```r
fClean<- rep("others", times=length(f))
for (i in 1:length(evtb)) {
    fClean[grep(evtb[i], f)]<- evtb[i]
}
rm(i)
```

Further grouping for events that could be clearly matched:

```r
fClean[grepl("wild[ /]", f) & fClean=="others"]<- "wildfire"
fClean[grepl("extreme cold", f) & fClean=="others"]<- "extreme cold/wind chill"
fClean[grepl("surge", f) & fClean=="others"]<- "storm surge/tide"
fClean[grepl("surf", f) & fClean=="others"]<- "high surf"

#(inline codes) number of event typs in the "others" group
unmatchedEV<- sum(fClean=="others")
```

2.4 Exame unmatched event types

There were 451 event types not successfully matched. They were generally ambiguous, and were examed with following codes:

```
unmatched<- which(storm$EVTYPE %in% levels(storm$EVTYPE)[fClean=="others"])

# check unmatched event types that have less than 10/only one record throught out 62 year
s of data
list(five.or.less= sum(table(droplevels(storm$EVTYPE[unmatched]))<6),
     only.one= sum(table(droplevels(storm$EVTYPE[unmatched]))==1))
```

```
## $five.or.less
## [1] 361
##
## $only.one
## [1] 223
```

```
# (inline codes) exame the unmatched events
unmatchedCheck<- list(Count=length(unmatched),
                      Count.Perc=length(unmatched)/nrow(storm),
                      Harm.Dmg.Perc=colSums(storm[unmatched, c("FATALITIES","INJURIES","a
djPROPDMG.M","adjCROPDMG.M")], na.rm=T)/colSums(storm[,c("FATALITIES","INJURIES","adjPROP
DMG.M","adjCROPDMG.M")], na.rm=T))
```

```
tnum=tnum+1
kable(data.frame(Variables=names(unmatchedCheck[[3]]), Perc=paste0(format(unmatchedCheck
[[3]]*100, digit=1),"%")), col.names=c("Variables","Unmatched event %"), row.names=F)
```

Variables	Unmatched event %
FATALITIES	2.22%
INJURIES	1.15%
adjPROPDMG.M	0.04%
adjCROPDMG.M	2.68%

Table 1. Percentage of harm and damage variables from unmatched events.

The unmatched events represented 0.5% (4781 rows) of the dataset. Many had only 5 or less records, and almost half had only one record. They were accounted for little share of fatalities, injuries, property and crop damage. It was therefore decided that no further event-matching was needed. The unmatched events were grouped as "others".

2.5. Create "Clean" event type variable

```
storm$EV.Clean<- storm$EVTYPE
levels(storm$EV.Clean)<- fClean

#check against 48 standard event type table
length(intersect(levels(storm$EV.Clean), evtb)) ; setdiff(levels(storm$EV.Clean), evtb)
```

```
## [1] 48
```

```
## [1] "others"
```

The "Clean" event types were labeled with 49 categories, 48 from the standard event and 1 from "others".

3. Event Type over Years

According to the NOAA website (http://www.ncdc.noaa.gov/stormevents/details.jsp?type=eventtype), only 3 event types(tornado, thunderstorm wind, and hail) were reported prior to 1993. This was examined with the codes below:

```
main3EV<- c("tornado", "thunderstorm wind", "hail")
tnum=tnum+1

#(inline cods) number of "non-tornado/thunder"
tb<-table(storm$EV.Clean %in% main3EV, storm$YEAR>1992)
```

Event Type	Before 1993	1993 Onward
Main 3*	187559	486993
All Others	0	227745

* *Main 3 event types: tornado, thunderstorm, and hail*

Table 2 Number of events reported before and after 1993 by event types

```
fnum=fnum+1
ggplot(storm, aes(x=YEAR)) +
    geom_histogram(aes(fill = EV.Clean %in% main3EV)) +
    labs(title = "Number of Events Reported\nAll States, 1993-2011", x = "Year", y = "Num
ber of Events Reported") +
    scale_fill_discrete(name="Event Type", breaks=c("TRUE", "FALSE"), labels=c("Main 3
*","All Others")) +
    theme_classic() +
    scale_x_continuous(limits=c(1950,2011), breaks=c(seq(1951,2011,6))) +
    theme(title = element_text(size=14, face="bold"), axis.title=element_text(size=12, fa
ce="bold"),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
    geom_vline(xintercept=1993, linetype="dotted", colour="red", size=1)
```
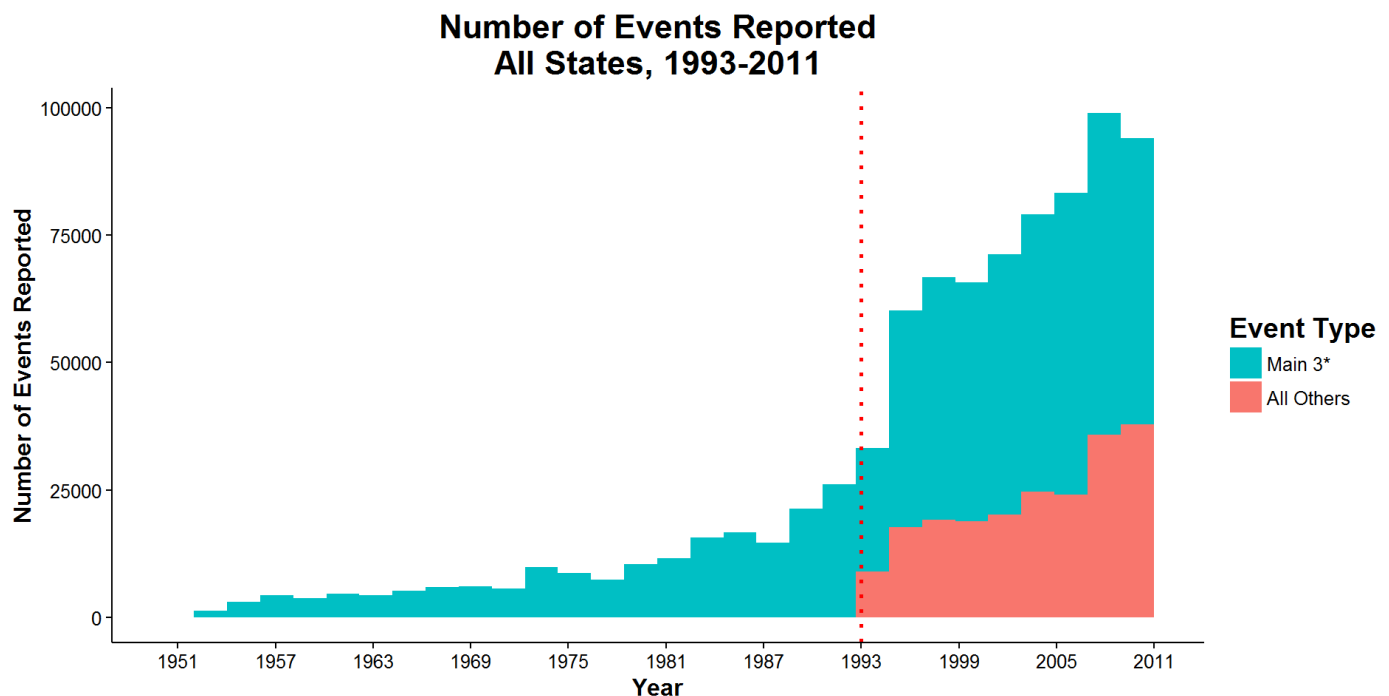

* Main 3 event types: tornado, thunderstorm, and hail

Figure 1 Number of events reported over time

Indeed there were only the main 3 event types reported before 1993. Therefore, data from 1950 to 1992 were excluded from this analysis.

4. Data Selection

Selected only data of year 1993-2011 and interested variables

```
storm1993on<- subset(storm, YEAR>1992, select=c("STATE","YEAR","EV.Clean","FATALITIES","I
NJURIES","adjPROPDMG.M","adjCROPDMG.M"))
```

Data Processing Summary

Property and Crop Damage

- Recognized only valid magnitudes "H, K, M, B"(hundred, thousand, million, and billion respectively). Ignored records with invalid magnitudes.

- Calculated damage dollar amount in million and adjusted by CPI

Event Type

- Regrouped event types, initially 985, into the standard 48 categories.

- Grouped ambiguous events as "others"

Years

- Used only data from 1993 to 2011 for this analysis due to the drastic change of event types reported.

Results

With Respect to Population Health

The Question: Across the United States, which types of events are most harmful with respect to population health?

```
harm.byEV <- storm1993on %>%
    select(EV.Clean, FATALITIES, INJURIES) %>%
    group_by(EV.Clean) %>%
    summarize(count=n(), fat=sum(FATALITIES), inj=sum(INJURIES)) %>%
    mutate(fat.perc=prop.table(fat), inj.perc=prop.table(inj)) %>%
    arrange(desc(fat))

harmTop5<- cbind(harm.byEV[c(1:5),c(1,3,5)], arrange(harm.byEV, desc(inj))[c(1:5),c(1,4,
6)])
countTop5<- arrange(harm.byEV, desc(count))[c(1:5),c(1:3)]

tnum<- tnum+1
kable(transform(harmTop5, fat.perc=paste0(format(fat.perc*100,digit=0),"%") ,inj.perc=pas
te0(format(inj.perc*100,digit=0),"%")),
        col.names=c("Top 5 Fatality Events","Fatalities","%", "Top 5 Injury Events","I
njuries","%"),
        row.names=F)
```

Top 5 Fatality Events	Fatalities	%	Top 5 Injury Events	Injuries	%
excessive heat	1922	18%	tornado	23328	34%
tornado	1646	15%	flood	6874	10%

heat	1212	11%	excessive heat	6525	9%
flash flood	1035	10%	thunderstorm wind	6116	9%
lightning	817	8%	lightning	5232	8%

Table 3 Top 5 events that caused most fatalities and injuries

"Excessive heat" was high on both fatality and injury top 5 lists. Another heat related event "Heat" also showed up on the top 5 fatality list. "Tornado" caused most injuries. Three weather events were on both top 5 lists: excessive heat, tornado, and lightning.

```
harmTop5.F<- as.character(harmTop5[,1])
harm.byEVplot<- transform(harm.byEV, Event.Type=EV.Clean)[,c(2,3,7)]
levels(harm.byEVplot$Event.Type)[which(!levels(harm.byEVplot$Event.Type) %in% harmTop5.
F)]<- "all other events"

fnum<- fnum+1
#the legend of this chart needs some work
ggplot(harm.byEVplot, aes(x=count, y=fat, group= Event.Type)) +
    geom_point(aes(color= Event.Type, size= Event.Type, alpha=Event.Type)) +
    scale_color_manual(values=c("black","red","deepskyblue","green4","purple","orange"))
+
    scale_alpha_manual(values=c(0.5,1,1,1,1,1)) +
    scale_size_manual(values=c(5,8,8,8,8,8)) +
    #scale_fill_manual("", name="Event Type", breaks=c(harmTop5.F, "all other events"), l
abels=c(harmTop5.F, "all other events"))  +
    labs(title = "Weather Events Fatalities by Number of Occurences\nAll States, 1993-201
1", x = "Number of Events Reported", y = "Fatalities") +
    #guides(fill=guide_legend(title="Event Type")) +
    scale_fill_discrete(name="Event Type", breaks=c(harmTop5.F, "all other events"), labe
ls=c(harmTop5.F, "all other events"))  +
    theme_bw() +
    theme(title = element_text(size=14, face="bold"), axis.title=element_text(size=12, fa
ce="bold"),
        panel.grid.major.x = element_line(colour="gray90", size=0.3), panel.grid.major.
y = element_line(colour="gray90", size=0.3), legend.position="right")
```

Weather Events Fatalities by Number of Occurences
All States, 1993-2011
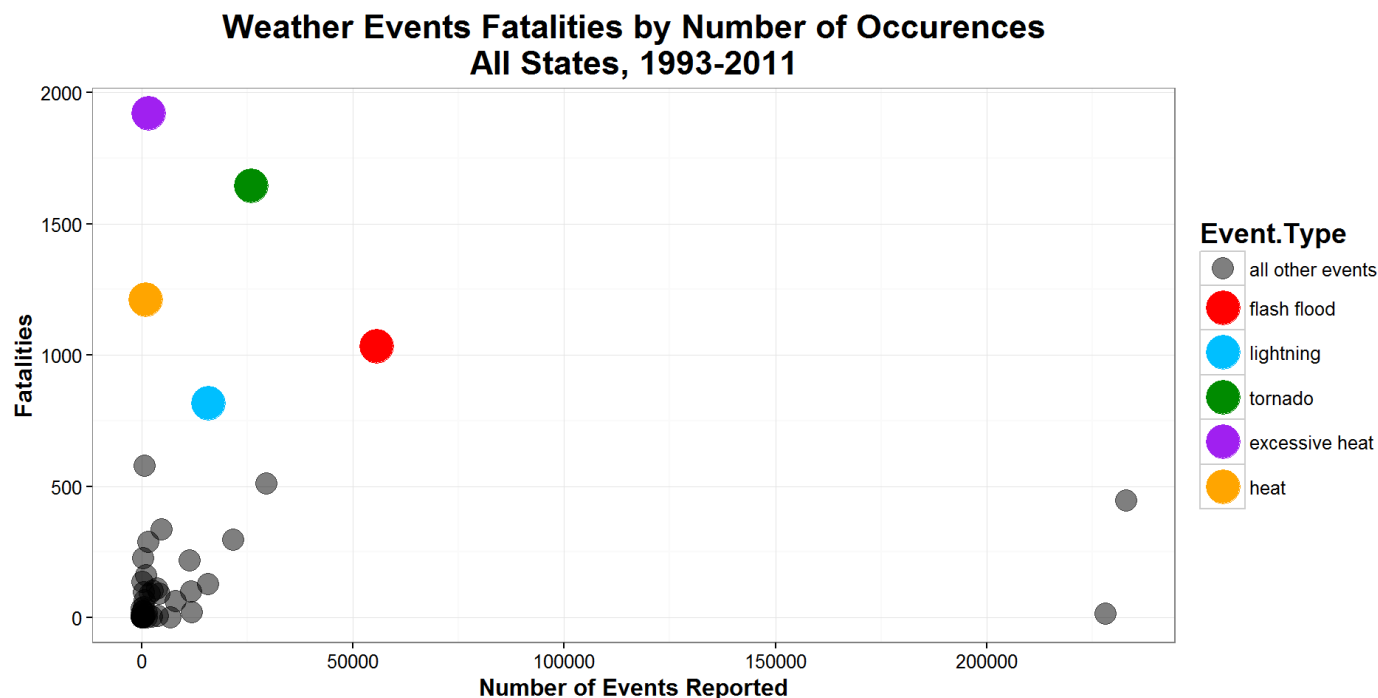
Figure2 Weather Events Fatalities by Number of Occurences

Among the top 5 fatality events, "flash flood" events have occured the most time. It was also observed that "thunderstorm wind", which ranked no.8 in fatality and no.4 in injury, has had the highest occurences in all events.

With Respect to Economic Consequences

The Question: Across the United States, which types of events have the greatest economic consequences?

```
dmg.byEV <- storm1993on %>%
    select(EV.Clean, adjPROPDMG.M, adjCROPDMG.M) %>%
    group_by(EV.Clean) %>%
    summarize(count=n(), prop=sum(adjPROPDMG.M, na.rm=T)/10^3, crop=sum(adjCROPDMG.M, na.
rm=T)/10^3) %>%   # damage dollors changed to "Billions"
    mutate(prop.perc=prop.table(prop), crop.perc=prop.table(crop),
          total=prop+crop, total.perc=prop.table(total)) %>%
    arrange(desc(prop))

dmgTop5<- cbind(dmg.byEV[c(1:5),c(1,3,5)], arrange(dmg.byEV, desc(crop))[c(1:5),c(1,4,
6)],
                arrange(dmg.byEV, desc(total))[c(1:5),c(1,7,8)])
dmgCountTop5<- arrange(dmg.byEV, desc(count))[c(1:5),c(1:4,7)]

tnum<- tnum+1
kable(transform(dmgTop5, prop.perc=paste0(format(prop.perc*100,digit=0),"%"), crop.perc=p
aste0(format(crop.perc*100,digit=0),"%"), total.perc=paste0(format(total.perc*100,digit=
0),"%")),
          col.names=c("Top 5 Property Damage","$B","%", "Top 5 Crop Damage","$B","%", "T
op 5 Property+Crop Damage","$B","%"), row.names=F)
```

Top 5 Property Damage	$B	%	Top 5 Crop Damage	$B	%	Top 5 Property+Crop Damage	$B	%
flood	171.709	37%	drought	17.778	28%	flood	186.489	35%
hurricane/typhoon	102.190	22%	flood	14.780	23%	hurricane/typhoon	109.124	21%
storm surge/tide	54.801	12%	ice storm	7.619	12%	storm surge/tide	54.802	10%
tornado	32.420	7%	hurricane/typhoon	6.934	11%	tornado	32.902	6%
flash flood	19.937	4%	hail	3.581	6%	hail	22.420	4%

Table 4 Top 5 events that caused most proterty and crop damages (in $Billions)

```
dmgTop10EV<- as.character(arrange(dmg.byEV, desc(total))[c(1:10),]$EV.Clean)
dmgRegroup<- dmg.byEV %>%
    arrange(desc(total)) %>%
    select(c(1,3,4)) %>%
    gather(dmg.type, dmg.M, -EV.Clean)

levels(dmgRegroup$EV.Clean)[which(!levels(dmgRegroup$EV.Clean) %in% dmgTop10EV)]<- "all o
ther events"

dmg.EV.order<- droplevels(rbind(arrange(dmg.byEV, desc(total))[c(1:10),c(1,7)], data.fram
e(EV.Clean="all other events", total=sum(dmgRegroup$dmg.M[dmgRegroup$EV.Clean=="all other
events"]))))
dmg.EV.order<- as.character(arrange(dmg.EV.order, desc(total))$EV.Clean)

fnum=fnum+1
ggplot(dmgRegroup, aes(x=EV.Clean, y=dmg.M, fill=dmg.type)) +
    geom_bar(stat = "identity", width=.7) + coord_flip() +
    labs(title = "Weather Events Property and Crop Damage\nAll States, 1993-2011", x = "E
vent Type (damage:  less  -->  more)", y = "Damage ($Billions)") +
    scale_fill_discrete(name="Damage Type", labels=c("Properties","Crops")) +
    scale_x_discrete(limits=rev(dmg.EV.order)) +
    scale_y_continuous(limits=c(0,200), breaks=seq(0,200,25)) +
    theme_bw() +
    theme(title = element_text(size=14, face="bold"), axis.title=element_text(size=12, fa
ce="bold"),
        panel.grid.major.x = element_line(colour="gray50", size=0.3, linetype="dotte
d"), panel.grid.major.y = element_blank(),
        legend.position=c(0.9,0.4))
```
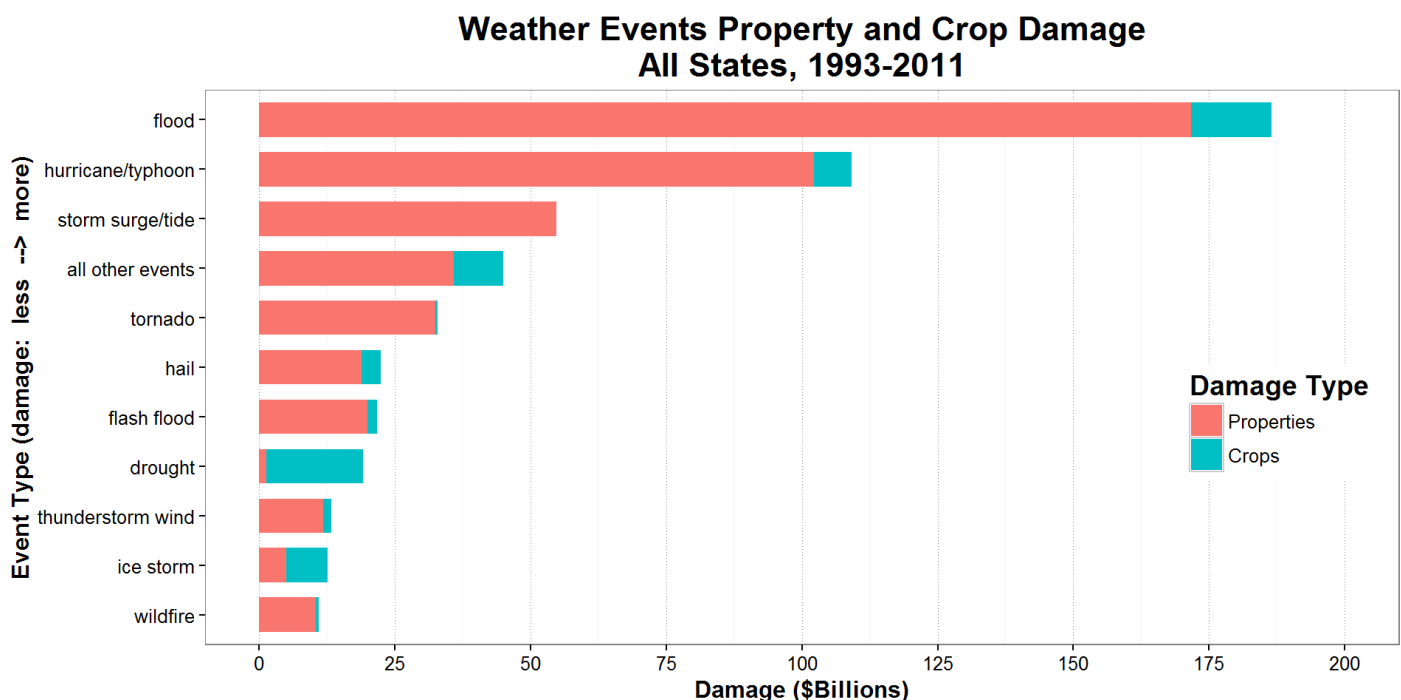

Figure 3 Weather Events Property and Crop Damages

"Flood" has caused the most property damage, followed by "hurricane/typhoon". "Drought" has caused the most crop damage, followed by "flood".

It was surprising how flood has caused significantly more (about 70%) of property damage than hurricane/typhoon, considering the recent events such as Hurricane Sandy and Hurricane Katrina. A further investigation would be necessary to see if any recording errors have occurred, for example, incorrect event type or incorrect damage magnitude.