US Severe Weather Data Analysis

Yinyan Guo

October 17, 2015

Synopsis

- Explore the most harmful (with respect to population health) event;
- Explore the temperal and spatial distribution of the most harmful event;
- Explore the types of events have the greatest economic consequences;

Data

U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database recorded from 1950 to Nevember 2011 downloaded at the course web site.

Processing Data

```
if(!file.exists("./data")){
    dir.create("data")
}
#download file
filename = "./data/repdata_data_StormData.csv.bz2"
if (!file.exists(filename)) {
    ## download url and destination file
    file.url <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
    file.dest <- './data/repdata_data_StormData.csv.bz2'
    ## download from the URL
    download.file(file.url, file.dest)
    unlink(file.url)
    require(R.utils)
    bunzip2("./data/repdata_data_StormData.csv.bz2",  "./data/repdata_data_StormData.csv", remove = FALSE, skip = TRUE)
}
```

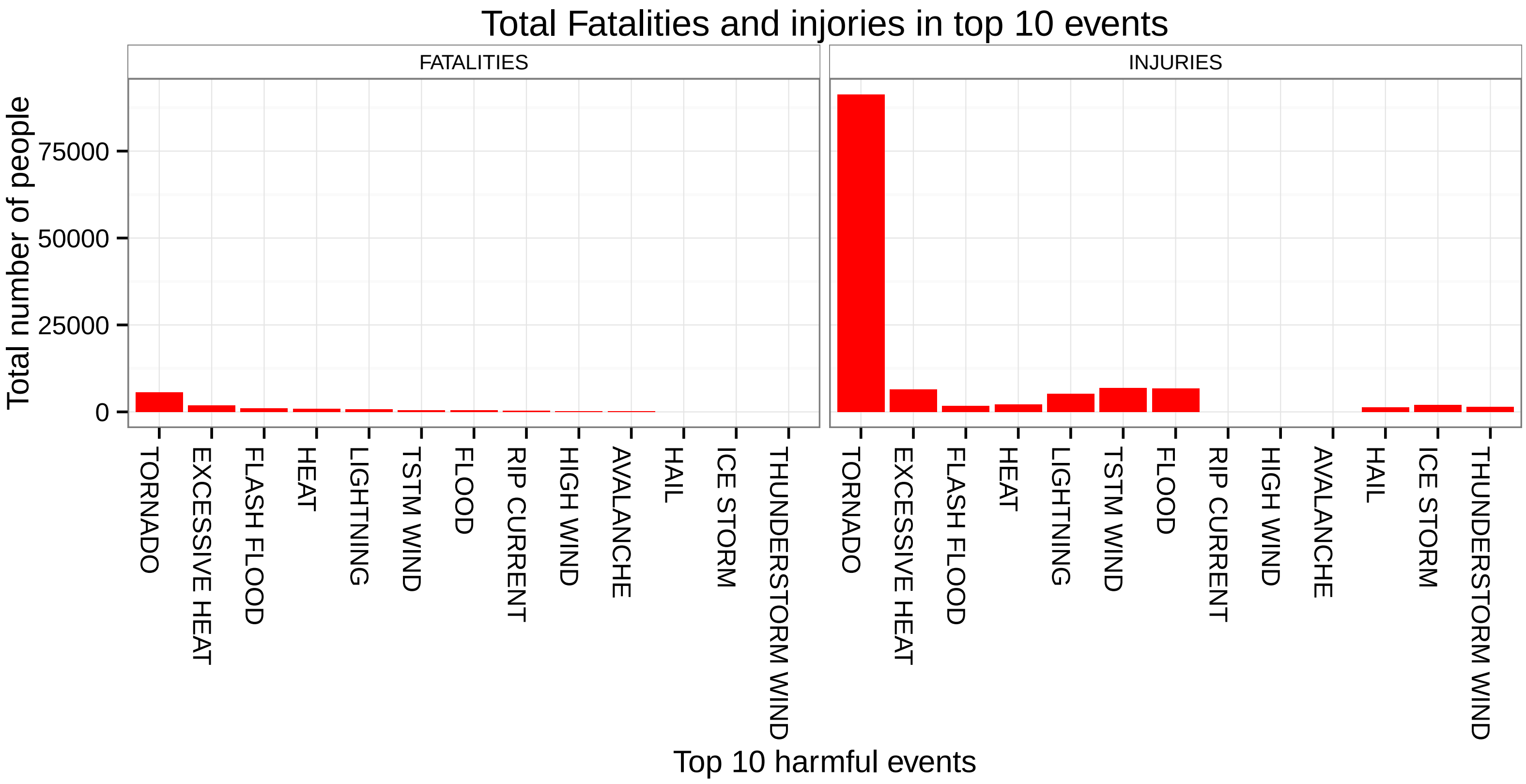
```
stormData <-  read.csv(filename, header=T,
                       sep=",", stringsAsFactors=F, na.strings=""
                       )
## subset with columns needed
mycol <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
mystormData <- stormData[mycol]
```

Results

Explore the most harmful events

```
##calculate the total fatalities and total injuries for each event
stormFatal <- aggregate(FATALITIES ~ EVTYPE, stormData, FUN = sum)
## top 10 events
stormFatalTop10 <- stormFatal[with(stormFatal, order(-FATALITIES)),][1:10,]
stormInju  <- aggregate(INJURIES ~ EVTYPE, stormData, FUN = sum)
## top 10 events
stormInjuTop10 <- stormInju[with(stormInju, order(-INJURIES)),][1:10,]
## top 10 fatalities and injuries events
stormHarmTop <- merge(stormFatalTop10,stormInjuTop10, by="EVTYPE",all=TRUE)
##decended sort by FATALITIES
stormHarmTop <-stormHarmTop[with(stormHarmTop, order(-FATALITIES)),]
##record the event order
flevel <- stormHarmTop$EVTYPE

require(reshape2)
stormHarmMelt <-  melt(stormHarmTop, id=c("EVTYPE"))
stormHarmMelt$EVTYPE <- ordered(stormHarmMelt$EVTYPE, levels = flevel)

#if(!file.exists("./result")){
#  dir.create("result")
#}

## make bar graph of top 10 most harmful events
require(ggplot2)
#png('./result/plot1.png', width=600, height=480)
```

```r
g <- ggplot(stormHarmMelt, aes(EVTYPE, value))
g <- g + geom_bar(stat="identity", fill="red") +
  facet_grid(. ~ variable, scales="free_y") +
  theme_bw() +
  xlab("Top 10 harmful events") +
  theme(axis.text.x=element_text(angle = -90, hjust = 0),
        strip.text.x = element_text(size=8, angle=0),
        strip.background = element_rect(fill="white")
  ) +
  ylab('Total number of people') +
  ggtitle('Total Fatalities and injories in top 10 events')
g
```



```
#dev.off()
stormHarmTop
```

```
##              EVTYPE FATALITIES INJURIES
## 12          TORNADO       5633    91346
## 2    EXCESSIVE HEAT       1903     6525
## 3       FLASH FLOOD        978     1777
## 6              HEAT        937     2100
## 9         LIGHTNING        816     5230
## 13        TSTM WIND        504     6957
## 4              FLOOD        470     6789
## 10       RIP CURRENT        368       NA
## 7          HIGH WIND        248       NA
## 1          AVALANCHE        224       NA
## 5               HAIL         NA     1361
## 8          ICE STORM         NA     1975
## 11 THUNDERSTORM WIND         NA     1488
```

Conclusion: TORNADO is the most harmful (in respect to population health) event.

TORNADO temporal (year, month) distribution

- TORNADO most damaged years and month

```
TORNADO <- stormData[stormData$EVTYPE=="TORNADO", c("BGN_DATE","STATE","FATALITIES","INJURIES")]
TORNADOYearMonth <- TORNADO
TORNADOYearMonth$BGN_DATE <-  as.Date(TORNADOYearMonth$BGN_DATE, "%m/%d/%Y")
TORNADOYearMonth$year <- format(TORNADOYearMonth$BGN_DATE,"%Y")
TORNADOYearMonth$month <- format(TORNADOYearMonth$BGN_DATE,"%b")

TORNADOYearFAT <- aggregate(FATALITIES ~ year, TORNADOYearMonth, FUN = sum)
TORNADOYearFATTop10 <- TORNADOYearFAT[with(TORNADOYearFAT, order(-FATALITIES)),][1:10,]
TORNADOYearInju <- aggregate(INJURIES ~ year, TORNADOYearMonth, FUN = sum)
TORNADOYearInjuTop10 <- TORNADOYearInju[with(TORNADOYearInju, order(-INJURIES)),][1:10,]
TORNADOHarmTopYear <- merge(TORNADOYearFATTop10,TORNADOYearInjuTop10, by="year",all=TRUE)
TORNADOHarmTopYear  <-TORNADOHarmTopYear[with(TORNADOHarmTopYear, order(-FATALITIES)),]
TORNADOHarmTopYear
```

```
##    year FATALITIES INJURIES
## 14 2011        587     6163
## 2  1953        519     5131
## 10 1974        366     6824
## 5  1965        301     5197
## 1  1952        230       NA
## 4  1957        193       NA
## 8  1971        159     2723
## 7  1968        131     2522
## 13 1998        130       NA
## 3  1955        129       NA
## 6  1967         NA     2144
## 9  1973         NA     2406
## 11 1979         NA     3014
## 12 1984         NA     2499
```

```
TORNADOMonthFAT <- aggregate(FATALITIES ~ month, TORNADOYearMonth, FUN = sum)
TORNADOMonthInju <- aggregate(INJURIES ~ month, TORNADOYearMonth, FUN = sum)
TORNADOHarmMonth <- merge(TORNADOMonthFAT,TORNADOMonthInju, by="month",all=TRUE)
TORNADOHarmMonth <- TORNADOHarmMonth[with(TORNADOHarmMonth, order(-FATALITIES)),]
TORNADOHarmMonth
```

```
##    month FATALITIES INJURIES
## 1    Apr       1793    29439
## 9    May       1253    17003
## 8    Mar        662     9559
## 7    Jun        565     9868
## 4    Feb        436     6027
## 10   Nov        251     4946
## 3    Dec        154     2928
## 5    Jan        137     2479
## 2    Aug        121     2804
## 11   Oct         99     2382
## 12   Sep         95     1799
## 6    Jul         67     2112
```

Conclusion: TORNADO in Year 2011, 1953, 1974, 1965 were most harmful.
TORNADO in April and May were most harmful.

TORNADO damage spatial (states) distribution in US

- TORNADO most harmful 10 tops states

```
## sum FATALITIES for each state
TORNADOStatFat <- aggregate(FATALITIES ~ STATE, TORNADO, FUN = sum)
TORNADOStatFatTop10 <- TORNADOStatFat[with(TORNADOStatFat, order(-FATALITIES)),][1:10,]
```

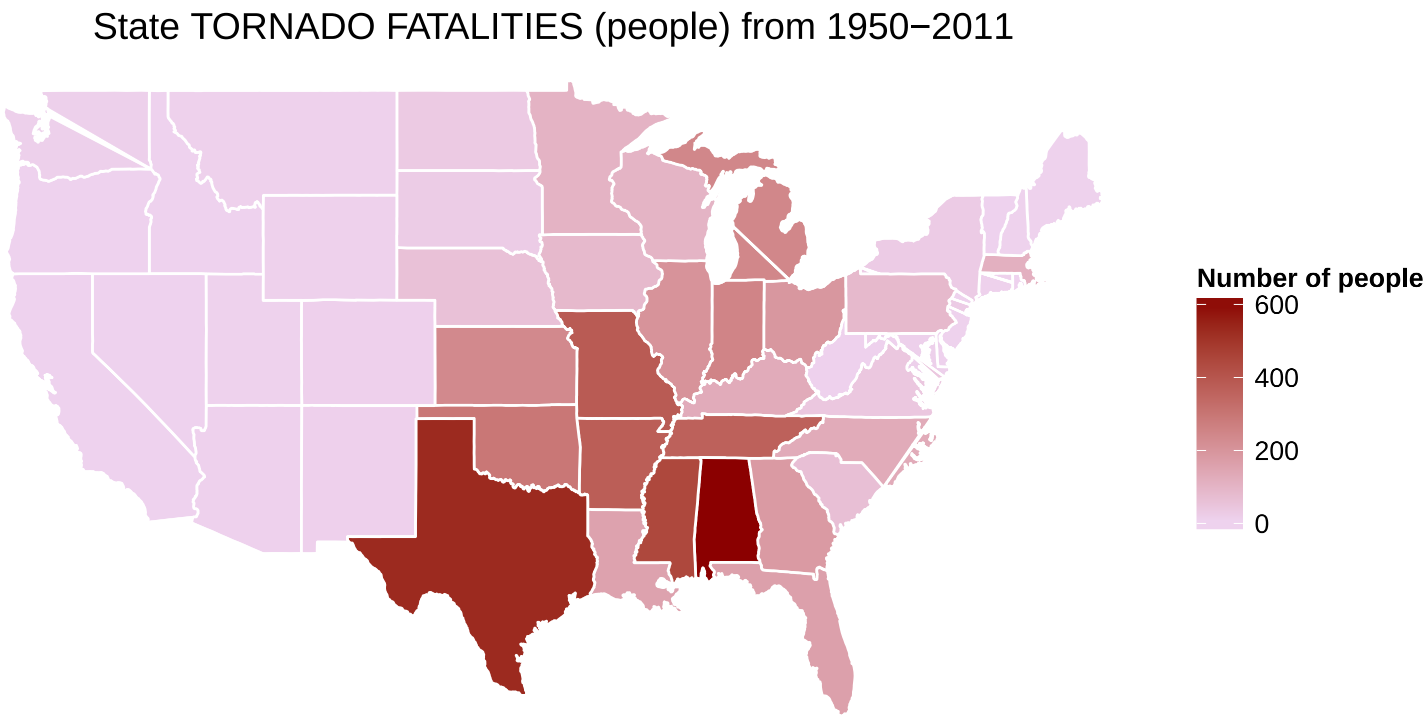
```
## sum INJURIES for each state
TORNADOStatInju <- aggregate(INJURIES ~ STATE, TORNADO, FUN = sum)
TORNADOStatInjuTop10 <- TORNADOStatInju[with(TORNADOStatInju, order(-INJURIES)),][1:10,]
TORNADOHarmStateTop10 <- merge(TORNADOStatFatTop10,TORNADOStatInjuTop10, by="STATE",all=TRUE)
TORNADOHarmStateTop10  <-TORNADOHarmStateTop10[with(TORNADOHarmStateTop10,order(-FATALITIES)),]
## TORNADO mosted harmful 10 states
TORNADOHarmStateTop10
```

```
##    STATE FATALITIES INJURIES
## 1     AL        617     7929
## 12    TX        538     8207
## 8     MS        450     6244
## 7     MO        388     4330
## 2     AR        379     5116
## 11    TN        368     4748
## 10    OK        296     4829
## 4     IN        252     4224
## 6     MI        243       NA
## 5     KS        236       NA
## 3     IL         NA     4145
## 9     OH         NA     4438
```

- The damaged TORNADO among states

```
## merge FATALITIES and INJURIES for all 50 states
TORNADOHarmState <- merge(TORNADOStatFat,TORNADOStatInju, by="STATE",all=TRUE)
## read in file to get full state name
statename <- read.csv("./data/statesNameID.csv", header=T, sep=";")
TORNADOHarmState <- merge(TORNADOHarmState,statename, by="STATE",all=TRUE)

require(ggplot2)
require(maps)
## state map file
all_states <- map_data("state")
##TORNADOHarmState merged with map file
TORNADOHarmStateMap <- merge(all_states,TORNADOHarmState,  by.x="region",by.y="name")
## draw map of FATALITIES
#png('./result/plot2.png', width=900, height=480)
```

```
p <- ggplot()
p <- p + geom_polygon(data=TORNADOHarmStateMap, aes(x=long, y=lat, group =  region, fill=TORNADOHarmStateMap$FATALITIES),colour="white") +
         scale_fill_continuous(low = "thistle2", high = "darkred", guide="colorbar")
P1 <- p + theme_bw()  + labs(fill = "Number of people"
                            ,title = "State TORNADO FATALITIES (people) from 1950-2011", x="", y="")
P1 + scale_y_continuous(breaks=c()) + scale_x_continuous(breaks=c()) + theme(panel.border =  element_blank())
```



```
#dev.off()
```

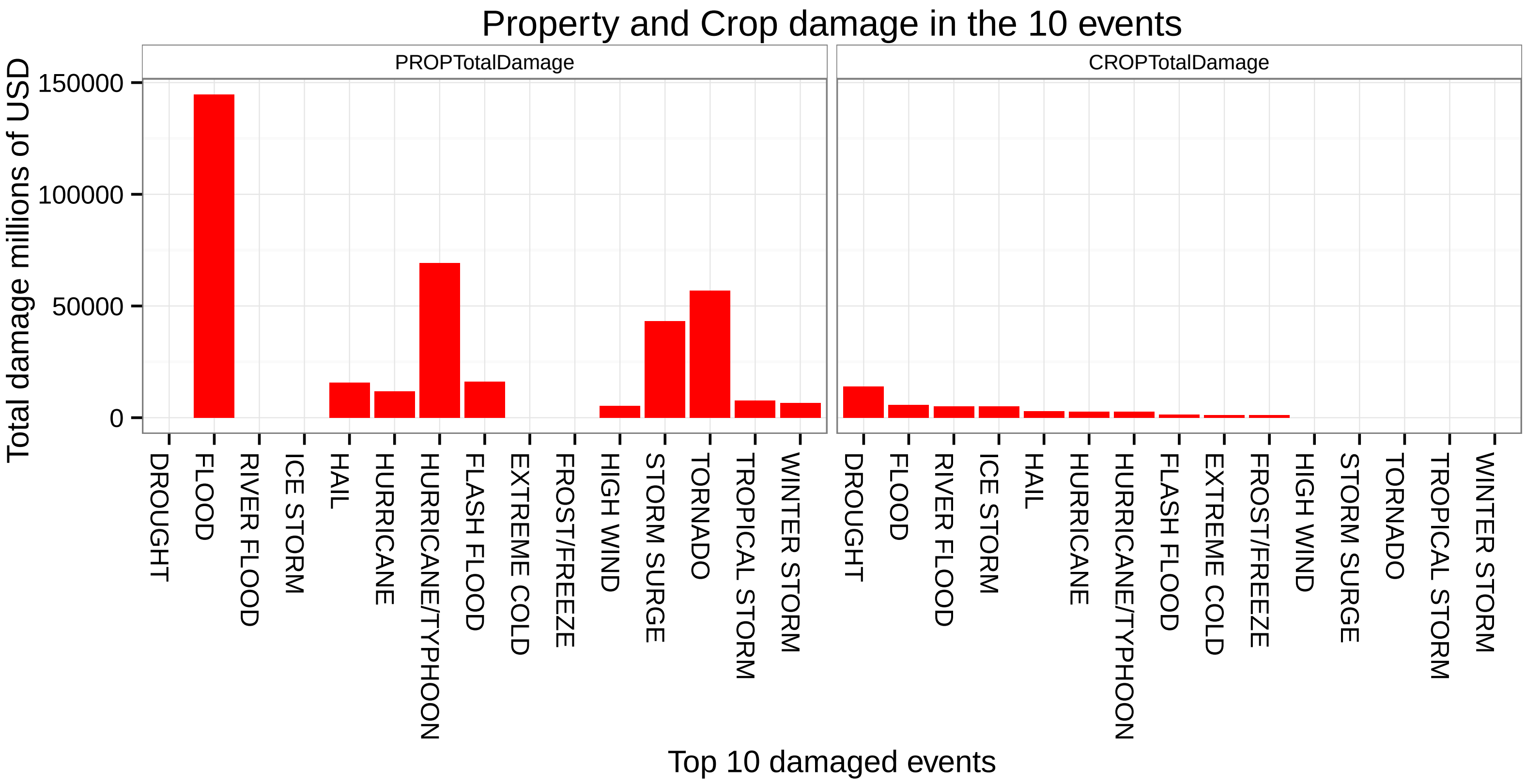
Conclusion: TORNADO in middle parts of US were most harmful.

The types of events have the greatest economic consequences

- Property damage and crop damage were recorded in NOAA database
- PROPDM (Proportional damage) and PROPDMGEXP (Proportional damage exponential) variables. The first variable has a proportion damage value, and the second variable has a measurement unit; for example, k category means thousands, m category means millions and b category billions.
- PROPTotalDamage = PROPDM * PROPDMGEXP while CROPTotalDamage =CROPDMG * CROPDMGEXP

```r
mystormData1<- mystormData
## change PROPDMGEXP to number based on k, m, b etc in PROPDMGEXP
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "h") | (mystormData1$PROPDMGEXP == "H")] <- 100
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "k") | (mystormData1$PROPDMGEXP == "K")] <- 1000
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "m") | (mystormData1$PROPDMGEXP == "M")] <- 1000000
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "B")] <- 1000000000
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "0") | (mystormData1$PROPDMGEXP == "")] <- 1
mystormData1$PROPDMGEXP[(mystormData1$PROPDMGEXP == "+") | (mystormData1$PROPDMGEXP == "-")  | (mystormData1$PROPDMGEXP == "?")] <- 0

## change CROPDMGEXP to number based on k, m, b etc in CROPDMGEXP
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "h") | (mystormData1$CROPDMGEXP == "H")] <- 100
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "k") | (mystormData1$CROPDMGEXP == "K")] <- 1000
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "m") | (mystormData1$CROPDMGEXP == "M")] <- 1000000
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "B")] <- 1000000000
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "0" | (mystormData1$CROPDMGEXP == ""))] <- 1
mystormData1$CROPDMGEXP[(mystormData1$CROPDMGEXP == "+") | (mystormData1$CROPDMGEXP == "-") | (mystormData1$CROPDMGEXP == "?")] <- 0

mystormData1$PROPDMGEXP <- as.numeric(mystormData1$PROPDMGEXP)
mystormData1$CROPDMGEXP <- as.numeric(mystormData1$CROPDMGEXP)

## calculate total USD damage amount
PROPTotalDamage  <- mystormData1$PROPDMG * mystormData1$PROPDMGEXP
CROPTotalDamage <- mystormData1$CROPDMG * mystormData1$CROPDMGEXP
mystormData2 <- cbind(mystormData1,PROPTotalDamage, CROPTotalDamage)
## calculate total USD damage on each event
PROPDMG <- aggregate(PROPTotalDamage ~ EVTYPE, mystormData2, FUN = sum)
PROPDMGTop10 <- PROPDMG[with(PROPDMG, order(-PROPTotalDamage)),][1:10,]

CROPDMG <- aggregate(CROPTotalDamage ~ EVTYPE, mystormData2, FUN = sum)
CROPDMGTop10 <- CROPDMG[with(CROPDMG, order(-CROPTotalDamage)),][1:10,]
## top most damaged events on property or crop
```

```r
DMGTop10 <- merge(PROPDMGTop10,CROPDMGTop10, by="EVTYPE",all=TRUE)
DMGTop10 <- DMGTop10[with(DMGTop10, order(-CROPTotalDamage)),]
flevels  <- DMGTop10$EVTYPE

##melt DMGTop10 on EVTYPE to be used in facet
DMGTop10Melt <-  melt(DMGTop10, id=c("EVTYPE"))
DMGTop10Melt$EVTYPE <- ordered(DMGTop10Melt$EVTYPE, levels=flevels)

##make graph
require(ggplot2)
#png('./result/plot3.png', width=600, height=480)
g <- ggplot(DMGTop10Melt, aes(EVTYPE, value/1000000))
g <- g + geom_bar(stat="identity", fill="red") +
  facet_grid(. ~ variable, scales="free_y") +
  theme_bw() +
  xlab("Top 10 damaged events") +
  theme(axis.text.x=element_text(angle = -90, hjust = 0),
        strip.text.x = element_text(size=8, angle=0),
        strip.background = element_rect(fill="white")
  ) +
  ylab('Total damage millions of USD ') +
  ggtitle('Property and Crop damage in the 10 events')
g
```



```
#dev.off()
DMGTop10
```

```
##               EVTYPE PROPTotalDamage CROPTotalDamage
## 1            DROUGHT              NA     13972566000
## 4              FLOOD    144657709800      5661968450
## 11       RIVER FLOOD              NA      5029459000
## 10         ICE STORM              NA      5022113500
## 6               HAIL     15732267606      3025954470
## 8          HURRICANE     11868319010      2741910000
## 9  HURRICANE/TYPHOON     69305840000      2607872800
## 3        FLASH FLOOD     16140812348      1421317100
## 2       EXTREME COLD              NA      1292973000
## 5       FROST/FREEZE              NA      1094086000
## 7          HIGH WIND      5270046260              NA
## 12       STORM SURGE     43323536000              NA
## 13           TORNADO     56937161125              NA
## 14    TROPICAL STORM      7703890550              NA
## 15      WINTER STORM      6688497251              NA
```

Conclusion: The most damaged event to crop is drought while the most damaged event to property is Flood.