Severe Weather Events in the United States - 1950 through 2011

Synopsis

We analyze severe weather events in the United States to determine what event types lead to the greatest affect population health and economic consequences. As regards this analysis, population health is used to describe both fatalities and injuries. Similarly, economic consequences is used to describe both property damage and crop damage. Based on this analysis, tornadoes have the greatest impact on population health while floods generate the greatest economic consequences.

Data Processing

All necessary packages for this analysis are loaded first.

```
options(rpubs.upload.method = "internal")
library(R.utils)
```

```
## Loading required package: R.oo
## Loading required package: R.methodsS3
## R.methodsS3 v1.6.1 (2014-01-04) successfully loaded. See ?R.methodsS3 for help.
## R.oo v1.18.0 (2014-02-22) successfully loaded. See ?R.oo for help.
##
## Attaching package: 'R.oo'
##
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
##
## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save
##
## R.utils v1.33.0 (2014-08-24) successfully loaded. See ?R.utils for help.
##
## Attaching package: 'R.utils'
##
## The following object is masked from 'package:utils':
##
##     timestamp
##
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

```
library(ggplot2)
library(plyr)
```

Next the data set is downloaded and extracted. This set is then loaded into the 'data' data frame.

```
setwd(".")
if(!file.exists("repdata-data-StormData.csv.bz2")){
  file.create("repdata-data-StormData.csv.bz2")
  URL <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  destfile <- "repdata-data-StormData.csv.bz2"
  download.file(URL, destfile)
}

data<-read.table("repdata-data-StormData.csv.bz2", header=T, sep = ",",
              colClasses=c("NULL",NA,rep("NULL",4),NA,NA,rep("NULL",3)
                        ,NA,rep("NULL",10),NA,NA,NA,NA,NA,NA,rep("NULL",9)))

summary(data)
```

```
##             BGN_DATE          STATE                  EVTYPE
##   5/25/2011 0:00:00:  1202   TX     : 83728   HAIL            :288661
##   4/27/2011 0:00:00:  1193   KS     : 53440   TSTM WIND       :219940
##   6/9/2011 0:00:00 :  1030   OK     : 46802   THUNDERSTORM WIND: 82563
##   5/30/2004 0:00:00:  1016   MO     : 35648   TORNADO         : 60652
##   4/4/2011 0:00:00 :  1009   IA     : 31069   FLASH FLOOD     : 54277
##   4/2/2006 0:00:00 :   981   NE     : 30271   FLOOD           : 25326
##   (Other)          :895866   (Other):621339   (Other)         :170878
##             END_DATE        FATALITIES       INJURIES        PROPDMG
##                   :243411   Min.   :  0   Min.   :   0.0   Min.   :   0
##   4/27/2011 0:00:00:  1214   1st Qu.:  0   1st Qu.:   0.0   1st Qu.:   0
##   5/25/2011 0:00:00:  1196   Median :  0   Median :   0.0   Median :   0
##   6/9/2011 0:00:00 :  1021   Mean   :  0   Mean   :   0.2   Mean   :  12
##   4/4/2011 0:00:00 :  1007   3rd Qu.:  0   3rd Qu.:   0.0   3rd Qu.:   0
##   5/30/2004 0:00:00:   998   Max.   :583   Max.   :1700.0   Max.   :5000
##   (Other)          :653450
##    PROPDMGEXP         CROPDMG        CROPDMGEXP
##           :465934   Min.   :  0.0           :618413
##   K       :424665   1st Qu.:  0.0   K       :281832
##   M       : 11330   Median :  0.0   M       :  1994
##   0       :   216   Mean   :  1.5   k       :    21
##   B       :    40   3rd Qu.:  0.0   0       :    19
##   5       :    28   Max.   :990.0   B       :     9
##   (Other):    84                    (Other):     9
```

Next, we consider our two categories of analysis: human loss and economic loss. With regards to human loss, we create a subset using the appropriate columns (event type, fatalities, injuries). We then create a new column, casualities, that sums the fatalities and injuries columns. This column better reflects the true human loss created by the event. This information is then summarized by event type using the plyr package.

```
health <- data[,c("EVTYPE","FATALITIES","INJURIES")]


health$CASUALTIES <- health$FATALITIES + health$INJURIES


healthsummary <- ddply(health,
                       .(EVTYPE),
                       summarize,
                       TOTAL=sum(CASUALTIES))
```

Similar processes are conducted for the appropriate economic damage related columns (property damage and crop damage). One caveat with this data is the use of exponents for damage calculations (i.e. the PROPDMGEXP column contains the value 'k' when the PROPDMG value is in 1,000s). Additional columns are constructed to input the multiplier needed for the most common exponents (billions, millions, thousands, and hundreds). This information is then computed into 'TOTAL' column. The package, plyr, is again used to summarize the data across event types.

```
damage <- data[,c("EVTYPE","PROPDMG","PROPDMGEXP","CROPDMG","CROPDMGEXP")]


damage$PROPMULT <- ifelse(damage$PROPDMGEXP == "B", 1000000000,
                   ifelse(damage$PROPDMGEXP == "M", 1000000,
                   ifelse(damage$PROPDMGEXP == "K", 1000,
                   ifelse(damage$PROPDMGEXP == "H", 100, 0))))


damage$CROPMULT <- ifelse(damage$CROPDMGEXP == "B", 1000000000,
                   ifelse(damage$CROPDMGEXP == "M", 1000000,
                   ifelse(damage$CROPDMGEXP == "K", 1000,
                   ifelse(damage$CROPDMGEXP == "H", 100, 0))))


damage$TOTAL <- damage$PROPDMG * damage$PROPMULT + damage$CROPDMG * damage$CROPMULT


damagesummary <- ddply(damage,
                       .(EVTYPE),
                       summarize,
                       TOTAL=sum(TOTAL))
```

Results

The below plot details the total number of casualties (fatalities and injuries) by event type. This indicates that tornados lead to the greatest human loss of any individual event type.

```
ggplot(healthsummary[order(healthsummary$TOTAL, decreasing = TRUE)[1:10],], aes(x = reorder(EVTYPE
, TOTAL), y = TOTAL)) + geom_bar(stat="identity") + coord_flip() + xlab("Event") + ylab("Casualtie
s") + ggtitle("Casualties by Event")
```

Casualties by Event
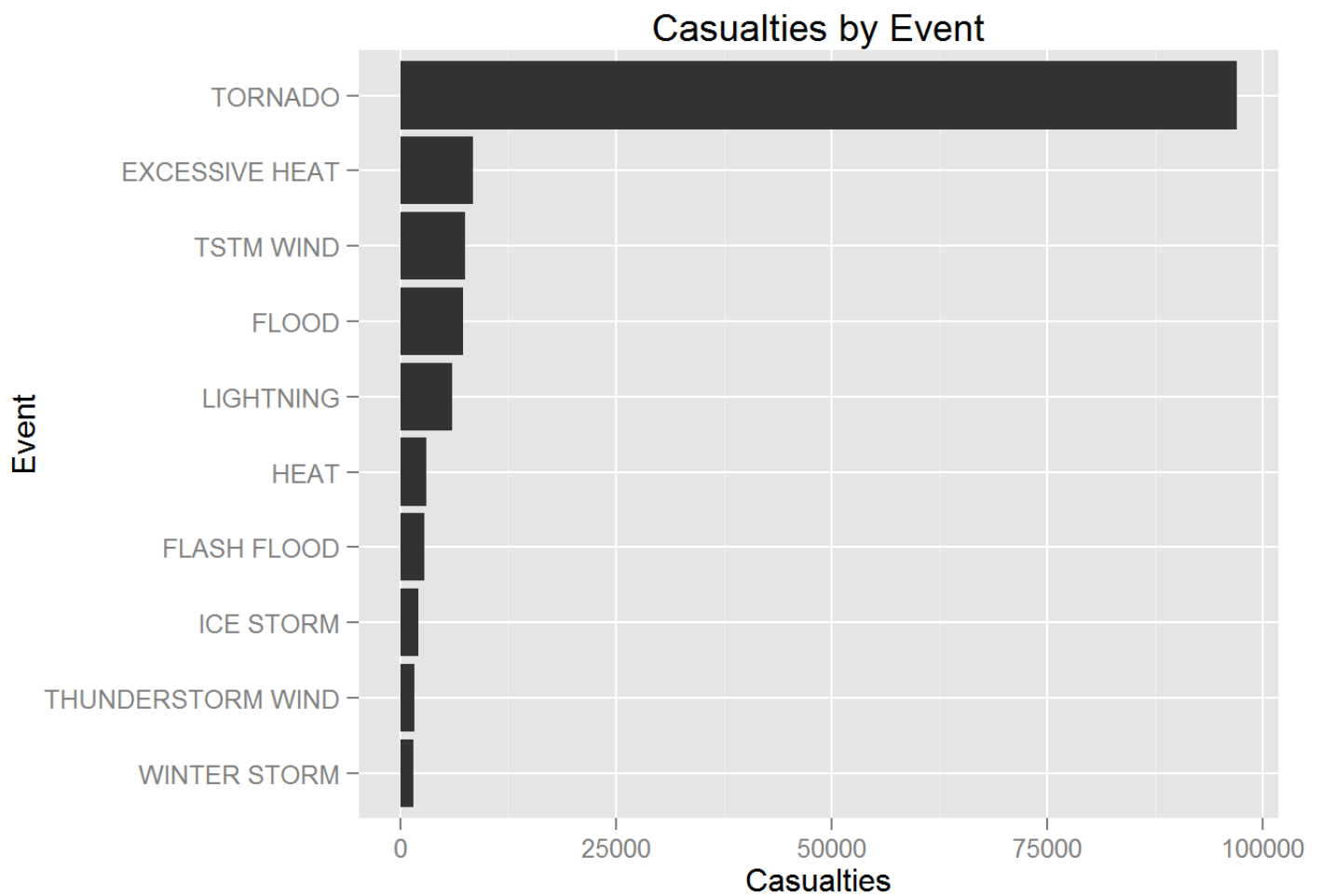
The below plot details the total economic loss (property and crop damage) by event type. This indicates that floods lead to the greatest economic loss of any individual event type.

```
ggplot(damagesummary[order(damagesummary$TOTAL, decreasing = TRUE)[1:10],], aes(x = reorder(EVTYPE
, TOTAL), y = TOTAL)) + geom_bar(stat="identity") + coord_flip() + xlab("Event") + ylab("Economic
Losses") + ggtitle("Economic Losses by Event")
```

Economic Losses by Event