Public health and economic effects of extreme weather events in the U.S. - 1950-2011

Synopsis

This is an exploratory analysis of the data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. It seeks to identify and present graphically the types of extreme weather events that have the most impact on the population's health (in terms of number of fatalities and injuries), and economic costs (in terms of the dollar value of property damage and crop damage). We find that tornadoes are the leading cause of deaths and injuries, whereas floods and droughts respectively cause the most property damage and crop damage.

Data Processing

The file 'repdata-data-StormData.csv.bz2'' is downloaded from the course website, as directed in the assignment instructions. We unzip it in R:

```
if (!file.exists("stormdata.csv"))
{
  library(R.utils)
  bunzip2("repdata-data-StormData.csv.bz2", "stormdata.csv", remove = FALS
E)
  }
```

We read the unzipped file to a R (as a dataframe).

```
 df_stormdata <- read.csv("stormdata.csv")
```

Explore the 37 fields contained in the raw data

```
head(df_stormdata, 2)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1 4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14   100 3   0          0
## 2         NA         0                       2   150 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAM
ES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
```

We create a more compact data set, retaining only the fields relevant to our analysis (as per the codebook) EVTYPE: Event type FATALITIES INJURIES PROPDMG: Property damage PROPDMGEXP: Propery damage exponential power key CROPDMG: Crop damage CROPDMGEXP: Crop damage exponential power key

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
df_stormdata <- select(df_stormdata, EVTYPE, FATALITIES, INJURIES, PROPDMG
, PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

We use the alphabetical keys for the exponential power to multiply the property and crop damage figures with the respective powers of 10

```
unique(df_stormdata$PROPDMGEXP)
```

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

We use this conversion for the symbolic exponent: H, h : 100 K: 1000 M, m : 1000000 B : 1000000000 Any numerical value: respective power of 10 Blank: 1 Any other value: 0

```
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "H"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "H"] * 100
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "h"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "h"] * 100
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "K"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "K"] * 1000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "M"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "M"] * 1000000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "m"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "m"] * 1000000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "B"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "B"] * 1000000000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "0"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "0"]
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "1"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "1"] * 10
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "2"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "2"] * 100
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "3"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "3"] * 1000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "4"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "4"] * 10000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "5"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "5"] * 100000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "6"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "6"] * 1000000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "7"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "7"] * 10000000
```

```
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "8"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "8"] * 100000000
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "+"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "+"] * 0
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "-"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "-"] * 0
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == "?"] <- df_stormdata$PROPD
MG[df_stormdata$PROPDMGEXP == "?"] * 0
df_stormdata$PROPDMG[df_stormdata$PROPDMGEXP == ""] <- df_stormdata$PROPDM
G[df_stormdata$PROPDMGEXP == ""]
```

```
unique(df_stormdata$CROPDMGEXP)
```

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

We use this conversion for the symbolic exponent: K, k: 1000 M, m : 1000000 B : 1000000000 Any numerical value: respective power of 10 Blank: 1 Any other value: 0

```
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "k"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "k"] * 100
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "K"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "K"] * 1000
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "M"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "M"] * 1000000
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "m"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "m"] * 1000000
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "B"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "B"] * 1000000000
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "0"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "0"]
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "2"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "2"] * 100
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == "?"] <- df_stormdata$CROPD
MG[df_stormdata$CROPDMGEXP == "?"] * 0
df_stormdata$CROPDMG[df_stormdata$CROPDMGEXP == ""] <- df_stormdata$CROPDM
G[df_stormdata$CROPDMGEXP == ""]
```

We group the entire data by the field EVTYPE

```
df_stormdata <- group_by(df_stormdata, EVTYPE)
```

We aggregate the fields FATALITIES, INJURIES, PROPDMG, CROPDMG by summing up in groups (EVTYPE)

```
df_stormdata <- summarise(df_stormdata, FATALITIES = sum(FATALITIES), INJU
RIES = sum(INJURIES), PROPDMG = sum(PROPDMG), CROPDMG = sum(CROPDMG))
```

We create separate dataframes, each containing the top ten event types by FATALITIES, INJURIES, PROPDMG, and CROPDMG respectively.

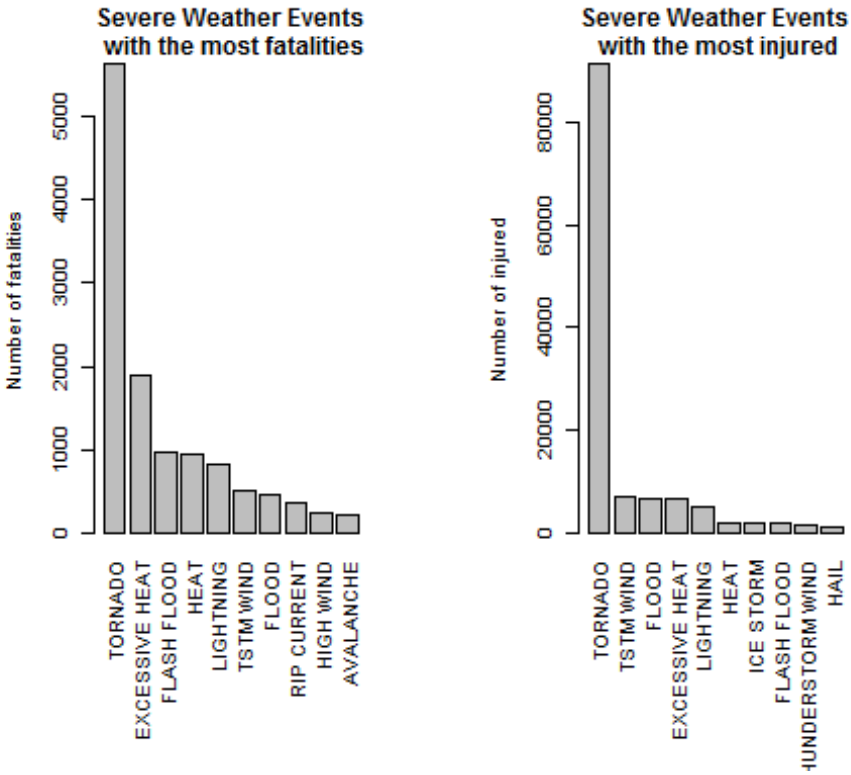
```
df_stormdata_popftl <- head(arrange(df_stormdata, desc(FATALITIES)), 10)
df_stormdata_popinjr <- head(arrange(df_stormdata, desc(INJURIES)), 10)
df_stormdata_propdmg <- head(arrange(df_stormdata, desc(PROPDMG)), 10)
df_stormdata_cropdmg <- head(arrange(df_stormdata, desc(CROPDMG)), 10)
```

Results

We create two plots - each with two panels: 1. Population health effects of severe weather events

```
par(mfrow = c(1, 2), cex = 0.6, mar=c(10,6,3,3))
barplot(df_stormdata_popftl$FATALITIES,
        names.arg = df_stormdata_popftl$EVTYPE,
        main = "Severe Weather Events\n with the most fatalities",
        ylab = "Number of fatalities",
        las = 3)

barplot(df_stormdata_popinjr$INJURIES,
        names.arg = df_stormdata_popinjr$EVTYPE,
         main = "Severe Weather Events\n with the most injured",
         ylab = "Number of injured",
         las = 3)
```



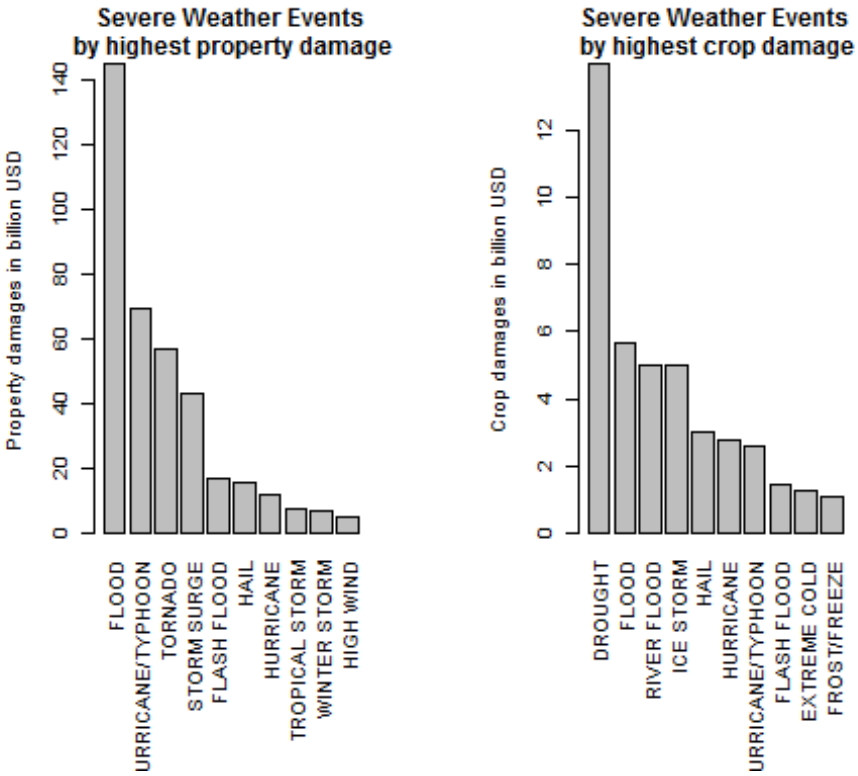
Across the U.S. tornadoes cause the most fatalities as well as injuries

2. Property and crop damage caused by severe weather events

```
par(mfrow = c(1, 2), cex = 0.6, mar=c(10,6,3,3))
barplot(df_stormdata_propdmg$PROPDMG / 1000000000,
        names.arg = df_stormdata_propdmg$EVTYPE,
        main = "Severe Weather Events\n by highest property damage",
        ylab = "Property damages in billion USD",
        las = 3)

barplot(df_stormdata_cropdmg$CROPDMG / 1000000000,
        names.arg = df_stormdata_cropdmg$EVTYPE,
```

```
        main = "Severe Weather Events\n by highest crop damage",
        ylab = "Crop damages in billion USD",
        las = 3)
```



Across the U.S., floods cause the greatest property damage, and droughts the greatest crop damage.