Reproducible Research - Peer assessment 2

Jure Bordon

Sunday, February 22, 2015

Damage of storms and other severe weather events

Synopsis

In this report we analyse the data from the storms and other sever weather events dataset. We are interested in which events cause the most injuries, the most fatalities and are the most devastating in regards to property and crop damage

Data processing

Unzip the file (download from here - copy it to the working directory) and read it:

```
data <- read.csv(bzfile("repdata-data-StormData.csv.bz2"), header = TRUE, stringsAsFactors = FALSE)
```

Required packages:

```
install.packages("ggplot2", repos='http://cran.us.r-project.org')
install.packages("gridExtra", repos='http://cran.us.r-project.org')
install.packages("dplyr", repos='http://cran.us.r-project.org')
require(ggplot2)
require(gridExtra)
require(dplyr)
```

Check the data:

```
head(data)
```

```
##   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
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
```

```
## 3         NA         0                        0.1   123 2   0          0
## 4         NA         0                        0.0   100 2   0          0
## 5         NA         0                        0.0   150 2   0          0
## 6         NA         0                        1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

We only require the following columns from the data set:

- EVTYPE - Type of event
- FATALITIES - # of fatalities
- INJURIES - # of injuries
- PROPDMG - Property damage in orders of magnitude
- PROPDMGEXP - Order of magnitude of property damage
- CROPDMG - Crop damage in orders of magnitude
- CROPDMGEXP - Order of magnitude of crop damage

```r
dmg_data <- select(data,EVTYPE,FATALITIES,INJURIES,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
```

We need to check how orders of magnitude are expressed:

```r
unique(dmg_data$PROPDMGEXP)
```

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

Express the property exponent data in numerically, so we can later compute property damage value, according to documentation:

```r
# Replace the exponent information, in order of appearance

dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "K"] <- 1000
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "M"] <- 1e+06
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == ""] <- 1
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "B"] <- 1e+09
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "m"] <- 1e+06
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "0"] <- 1
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "5"] <- 1e+05
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "6"] <- 1e+06
```

```r
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "4"] <- 10000
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "2"] <- 100
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "3"] <- 1000
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "h"] <- 100
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "7"] <- 1e+07
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "H"] <- 100
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "1"] <- 10
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "8"] <- 1e+08

# Invalid data is set to 0, so it will not be added in the computation

dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "+"] <- 0
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "-"] <- 0
dmg_data$PROPEXP[dmg_data$PROPDMGEXP == "?"] <- 0

# Compute the property damage

dmg_data$PROPDMG.VAL <- dmg_data$PROPDMG*dmg_data$PROPEXP
```

Same as with property exponent data, we need to transform the crop exponent data:

```r
unique(dmg_data$CROPDMGEXP)
```

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

```r
# Replace the exponent information, in order of appearance

dmg_data$CROPEXP[dmg_data$CROPDMGEXP == ""] <- 1
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "M"] <- 1e+06
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "K"] <- 1000
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "m"] <- 1e+06
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "B"] <- 1e+09
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "0"] <- 1
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "k"] <- 1000
dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "2"] <- 100

# Invalid data is set to 0, so it will not be added in the computation

dmg_data$CROPEXP[dmg_data$CROPDMGEXP == "?"] <- 0

# Compute the crop damage

dmg_data$CROPDMG.VAL <- dmg_data$CROPDMG * dmg_data$CROPEXP
```

We now compute total fatalities, injuries, property damage and crop damage by event type and arrange them in descending order:

```r
fatalities <- summarise(group_by(dmg_data,EVTYPE),FATALITIES.TOTAL = sum(FATALITIES))
fatalities <- arrange(fatalities,desc(FATALITIES.TOTAL))

injuries <- summarise(group_by(dmg_data,EVTYPE),INJURIES.TOTAL = sum(INJURIES))
injuries <- arrange(injuries,desc(INJURIES.TOTAL))
```

```
propdmg <- summarise(group_by(dmg_data,EVTYPE),PROPDMG.VAL.TOTAL = sum(PROPDMG.VAL))
propdmg <- arrange(propdmg,desc(PROPDMG.VAL.TOTAL))

cropdmg <- summarise(group_by(dmg_data,EVTYPE),CROPDMG.VAL.TOTAL = sum(CROPDMG.VAL))
cropdmg <- arrange(cropdmg,desc(CROPDMG.VAL.TOTAL))
```

Results

Questions

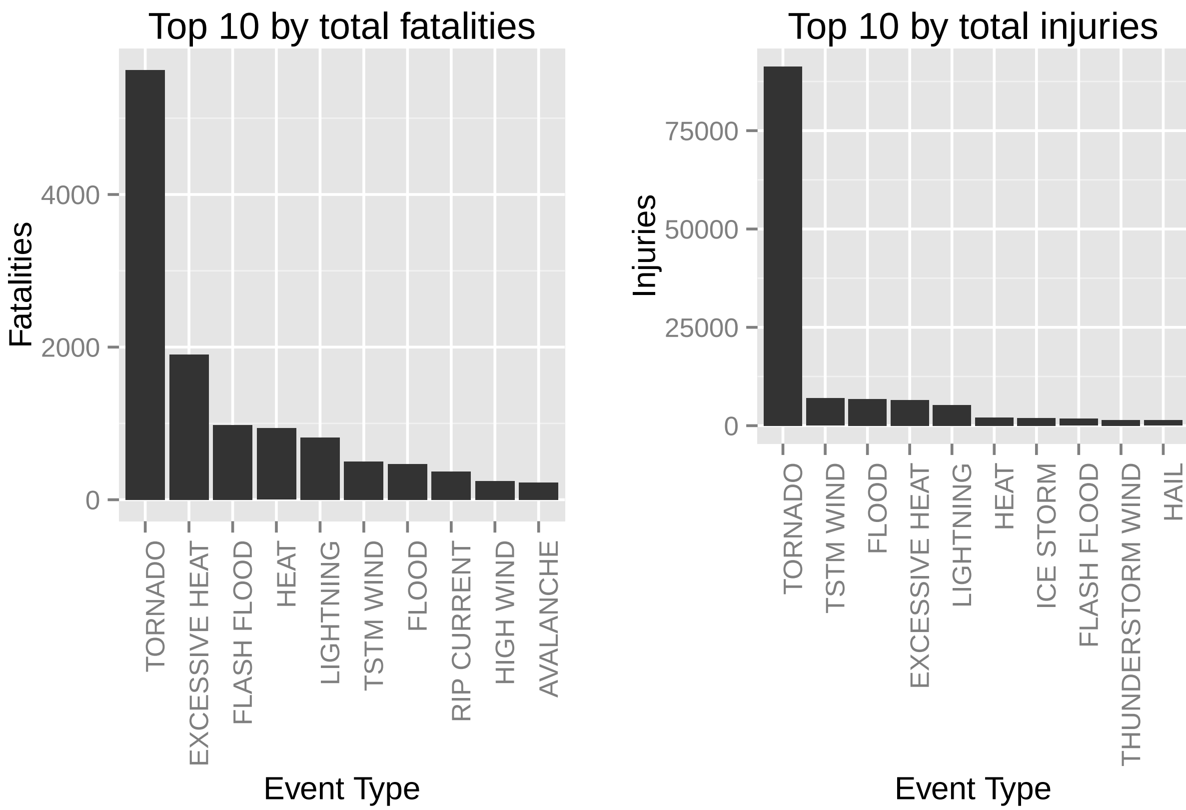
Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? The top 10 most damaging weather events in regards to fatalities and injuries:

```
fatalities <- head(fatalities,10)
injuries <- head(injuries,10)

p1 <- ggplot(fatalities,aes(x=reorder(EVTYPE,-FATALITIES.TOTAL),y=FATALITIES.TOTAL)) +
    geom_bar(stat="identity") +
    labs(x="Event Type", y="Fatalities", title="Top 10 by total fatalities") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))

p2 <- ggplot(injuries,aes(x=reorder(EVTYPE,-INJURIES.TOTAL),y=INJURIES.TOTAL)) +
    geom_bar(stat="identity") +
    labs(x="Event Type", y="Injuries", title="Top 10 by total injuries") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))

grid.arrange(p1, p2, ncol=2)
```



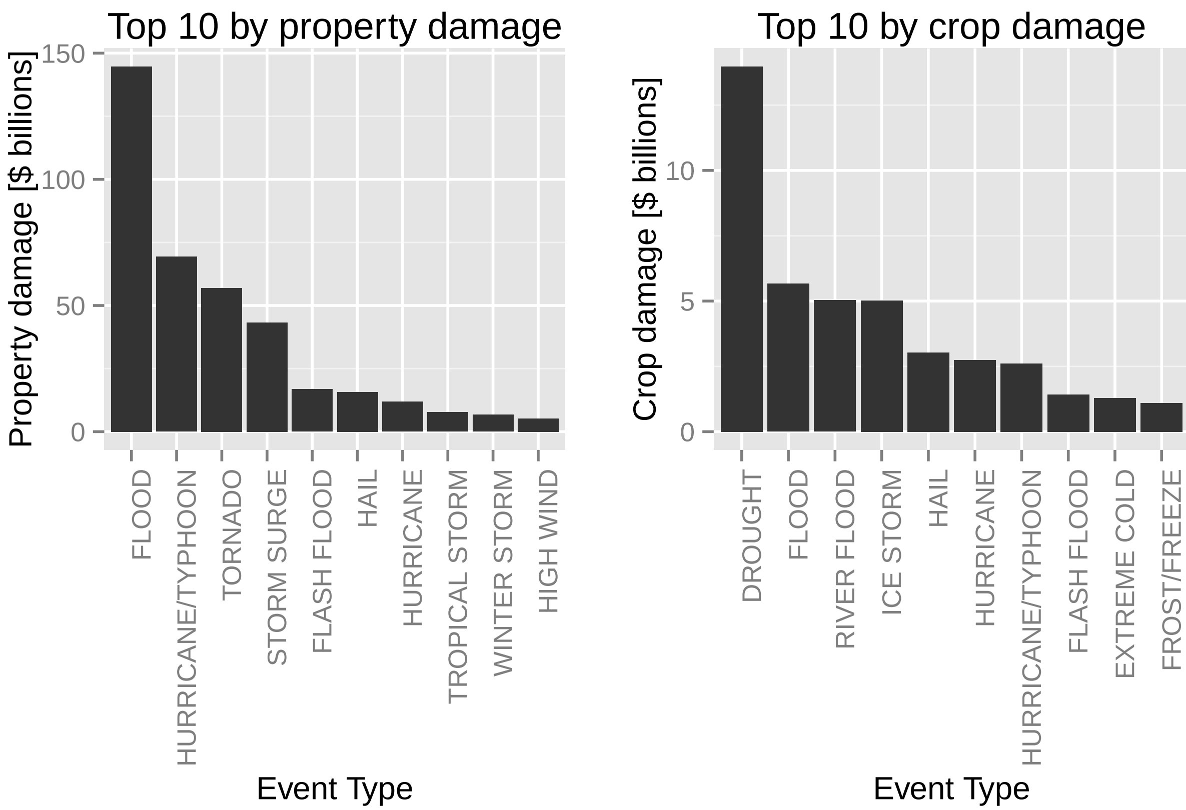
Across the United States, which types of events have the greatest economic consequences? The top 10 most damaging weather events in regards to property and crop damage in $ billion:

```
propdmg <- head(propdmg,10)
cropdmg <- head(cropdmg,10)

p1 <- ggplot(propdmg,aes(x=reorder(EVTYPE,-PROPDMG.VAL.TOTAL),y=PROPDMG.VAL.TOTAL/10^9)) +
    geom_bar(stat="identity") +
    labs(x="Event Type", y="Property damage [$ billions]", title="Top 10 by property damage") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))

p2 <- ggplot(cropdmg,aes(x=reorder(EVTYPE,-CROPDMG.VAL.TOTAL),y=CROPDMG.VAL.TOTAL/10^9)) +
    geom_bar(stat="identity") +
    labs(x="Event Type", y="Crop damage [$ billions]", title="Top 10 by crop damage") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))

grid.arrange(p1, p2, ncol=2)
```



Conclusions

According to the data plotted, the most damaging weather event in regards to both injuries and fatalities are tornadoes. For property damage the most devastating event are floods, while for crop damage the most impactful event is drought.