U.S. National Oceanic and Atmospheric Administration's (NOAA) data analize

Synopsis

Summary

Course Project 2 for coursera reproducible research course. Goal of analysis is to find which types of events are most harmful with respect to population health and which have the greatest economic consequences. Data come from U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database which tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The events in the database start in the year 1950 and end in November 2011. Data documentation can be found here data documentatnion

Analysis steps

1. Required libraries was installed and loaded (ggplot2 and cowplot).
2. Data was downloaded from Storm Data and loaded to "data] data frame
3. Data subsets was Create with 0.99 quantile, use to find most harmful events.
4. In case of property and corp damage the value need to be multiplied by 10 to power of exponent.
5. Result plots was Create using ggplot2 and cowplot. Log function is use to scale the data the dominance of the plots by outlier data.

Data Processing

Load necessary libraries.

```
if (!require('ggplot2'))
  install.packages('ggplot2');
library(ggplot2)

if (!require('cowplot'))
   install.packages('cowplot');
library(cowplot)
```

Download and read csv file.

```
url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
file = "Data/FStormData.csv.bz2"
if(!file.exists(file)){
  download.file(url,destfile = file)
}
data <- read.csv(file)
```

Create data set "fatalities_99", which is a subset of original data with events with number of fatalities greater than 99% of data.

```
fatalites <- subset(data, FATALITIES>0)
q_fatalites <-quantile(fatalites$FATALITIES, probs=c(.25,.50,0.95,0.99))
fatalites_99 <- subset(fatalites, FATALITIES > q_fatalites[4])
```

Create data set "injuries_99", which is a subset of original data with events with number of injuries greater than 99% of data.

```
injuries <- subset(data, INJURIES>0)
q_injuries <-quantile(injuries$INJURIES, probs=c(.25,.50,0.95,0.99))
injuries_99 <- subset(injuries, INJURIES > q_injuries[4])
```

Create data set "property_damage_99", which is a subset of original data with events with number of property greater than 99% of data.

```
property_damage <- subset(data, PROPDMG>0 & !(PROPDMGEXP %in% c("+","-","?")))
property_damage$PROPEXP[property_damage$PROPDMGEXP == "K"] <- 3
property_damage$PROPEXP[property_damage$PROPDMGEXP == "M"] <- 6
property_damage$PROPEXP[property_damage$PROPDMGEXP == ""] <- 0
property_damage$PROPEXP[property_damage$PROPDMGEXP == "B"] <- 9
property_damage$PROPEXP[property_damage$PROPDMGEXP == "m"] <- 6
property_damage$PROPEXP[property_damage$PROPDMGEXP == "0"] <- 0
property_damage$PROPEXP[property_damage$PROPDMGEXP == "5"] <- 5
property_damage$PROPEXP[property_damage$PROPDMGEXP == "6"] <- 6
property_damage$PROPEXP[property_damage$PROPDMGEXP == "4"] <- 4
property_damage$PROPEXP[property_damage$PROPDMGEXP == "2"] <- 2
property_damage$PROPEXP[property_damage$PROPDMGEXP == "3"] <- 3
property_damage$PROPEXP[property_damage$PROPDMGEXP == "h"] <- 2
property_damage$PROPEXP[property_damage$PROPDMGEXP == "7"] <- 7
property_damage$PROPEXP[property_damage$PROPDMGEXP == "H"] <- 2
property_damage$PROPEXP[property_damage$PROPDMGEXP == "1"] <- 1
property_damage$PROPEXP[property_damage$PROPDMGEXP == "8"] <- 8

property_damage$PROPDMG <- property_damage$PROPDMG * (10 ** property_damage$PROPEXP)

q_property_damage <-quantile(property_damage$PROPDMG, probs=c(.25,.50,0.95,0.9998))
property_damage_99 <- subset(property_damage, PROPDMG > q_property_damage[4])
```

Create data set "crop_damage_99", which is a subset of original data with events with number of crop damages greater than 99% of data.

```
crop_damage <- subset(data, CROPDMG>0  & !(CROPDMGEXP %in% c("+","-","?")))
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "M"] <- 6
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "K"] <- 3
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "m"] <- 6
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "B"] <- 9
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "0"] <- 0
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "k"] <- 3
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == "2"] <- 2
crop_damage$CROPEXP[crop_damage$CROPDMGEXP == ""] <- 0

crop_damage$CROPDMG <- crop_damage$CROPDMG * (10 ** crop_damage$CROPEXP)

q_crop_damage <-quantile(crop_damage$CROPDMG, probs=c(.25,.50,0.95,0.998))
crop_damage_99 <- subset(crop_damage, CROPDMG > q_crop_damage[4])
```

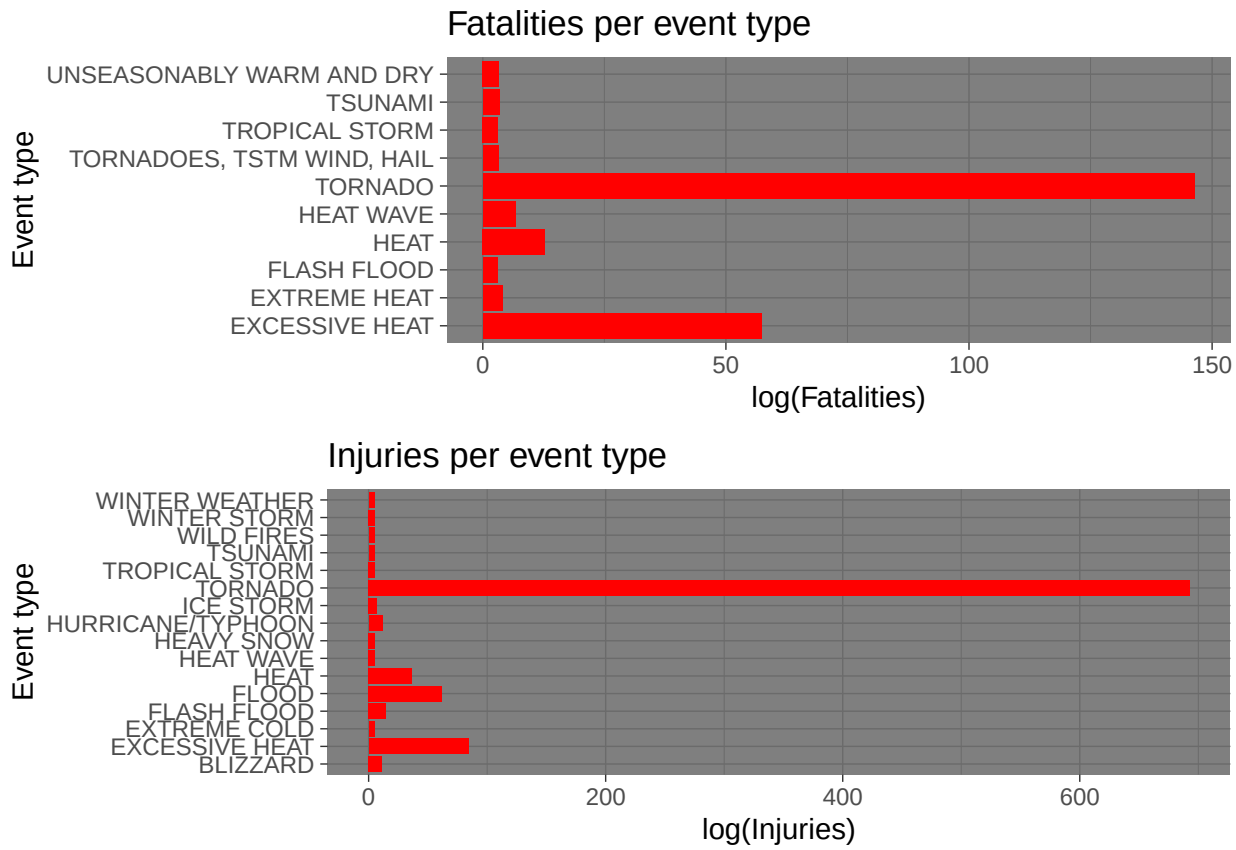
Results

Witch events are most harmful.

```
f <- ggplot(fatalites_99, aes(x=log(FATALITIES),y=EVTYPE)) +
      geom_bar(fill="red",stat="identity") +
      ggtitle("Fatalities per event type") +
      xlab("log(Fatalities)") +
      ylab("Event type") +
      theme_dark()

i <- ggplot(injuries_99, aes(x=log(INJURIES),y=EVTYPE)) +
      geom_bar(fill="red",stat="identity") +
      ggtitle("Injuries per event type") +
      xlab("log(Injuries)") +
      ylab("Event type") +
      theme_dark()

plot_grid(f,i,nrow = 2, ncol = 1)
```



Fatalities quantiles:

```
print(quantile(data$FATALITIES))
```

```
##   0%  25%  50%  75% 100%
##    0    0    0    0  583
```

Injuries quantiles:

```
print(quantile(data$INJURIES))
```

```
##   0%  25%  50%  75% 100%
##    0    0    0    0 1700
```

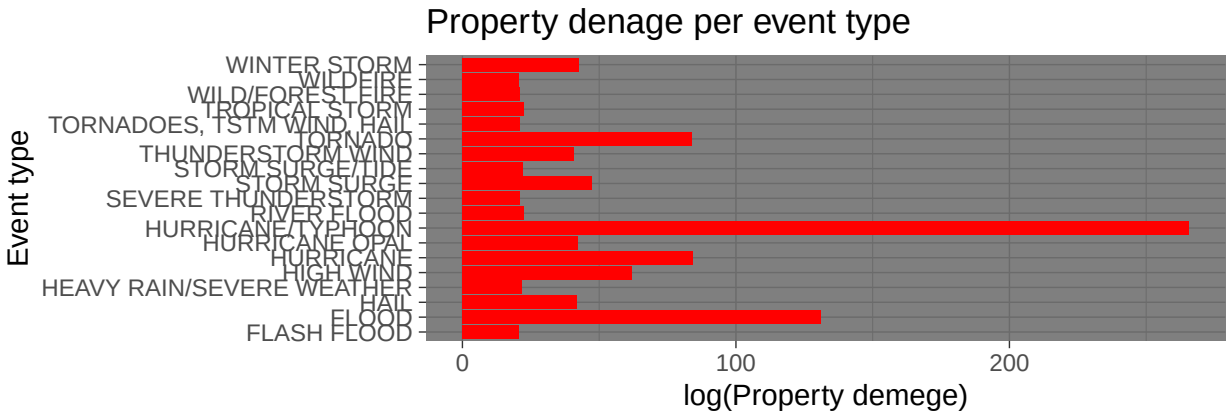
As we can see vast majority of fatalities and injures is caused by less than 1% of events.The most harmful ones are: tornadoes, floods, excessive and extreme heat and head in general. One interesting thing worth noting is that floods cause lots of the injures but almost non of the fatalities.

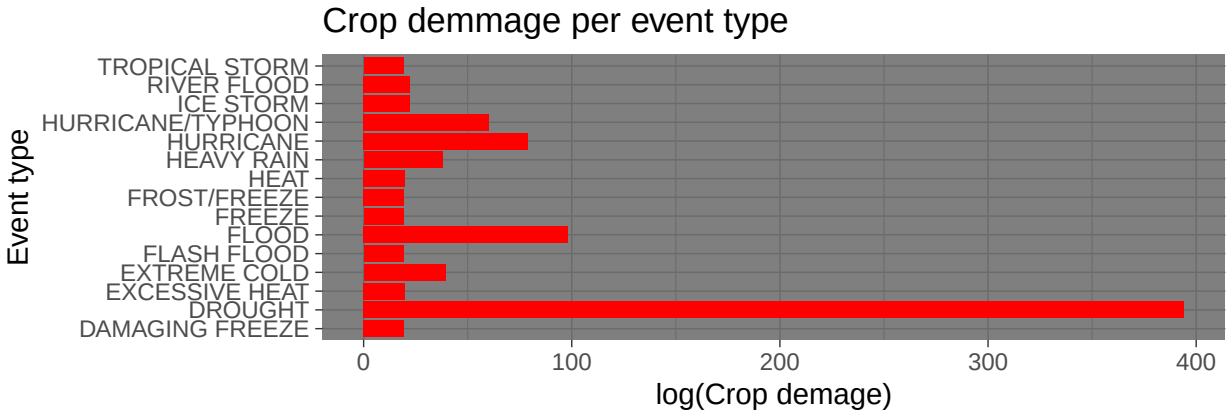
Witch events have the greatest economic consequences.

```
f <- ggplot(property_damage_99, aes(x=log(PROPDMG),y=EVTYPE)) +
      geom_bar(fill="red",stat="identity") +
      ggtitle("Property denage per event type") +
      xlab("log(Property demege)") +
      ylab("Event type") +
      theme_dark()

i <- ggplot(crop_damage_99, aes(x=log(CROPDMG),y=EVTYPE)) +
      geom_bar(fill="red",stat="identity") +
      ggtitle("Crop demmage per event type") +
      xlab("log(Crop demage)") +
      ylab("Event type") +
      theme_dark()

plot_grid(f,i,nrow = 2, ncol = 1)
```





Property damage quantiles:

```
print(quantile(data$PROPDMG))
```

```
##    0%   25%   50%   75%  100%
## 0e+00 0e+00 0e+00 5e-01 5e+03
```

Crop damage quantiles:

```
print(quantile(data$CROPDMG))
```

```
##   0%  25%  50%  75% 100%
##    0    0    0    0  990
```

As we can see vast majority of economic consequences is caused by less than 1% of events.The most harmful ones are: floods, tornadoes and event associated with wind. In case of crop damage another imported factor is heavy snow.