

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

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.

Synopsis

TODO synopsis

Data Processing

Load necessary libraries.

```
library(ggplot2)
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)
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)
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])
```

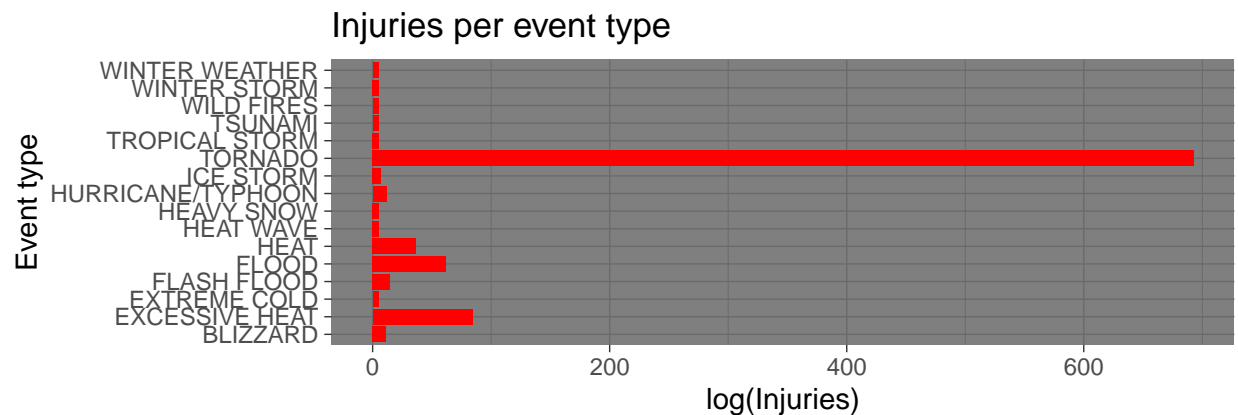
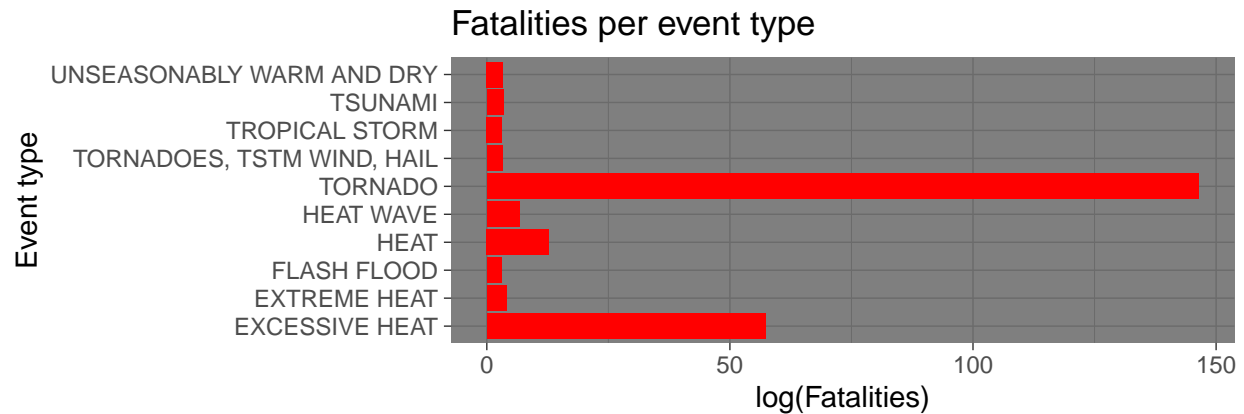
Results

Witch events are most harmful.

```
f <- ggplot(fatalities_99, aes(x=log(FATALITIES),y=EVTTYPE)) +
  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=EVTTYPE)) +
  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:

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

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

Injuries:

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

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

As we can see vast majority of fatalities and injuries 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 injuries but almost none of the fatalities.

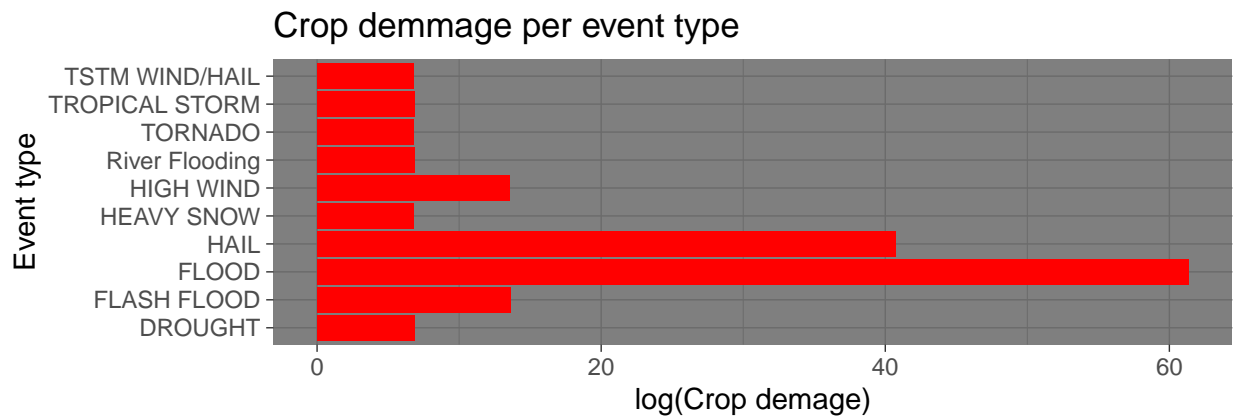
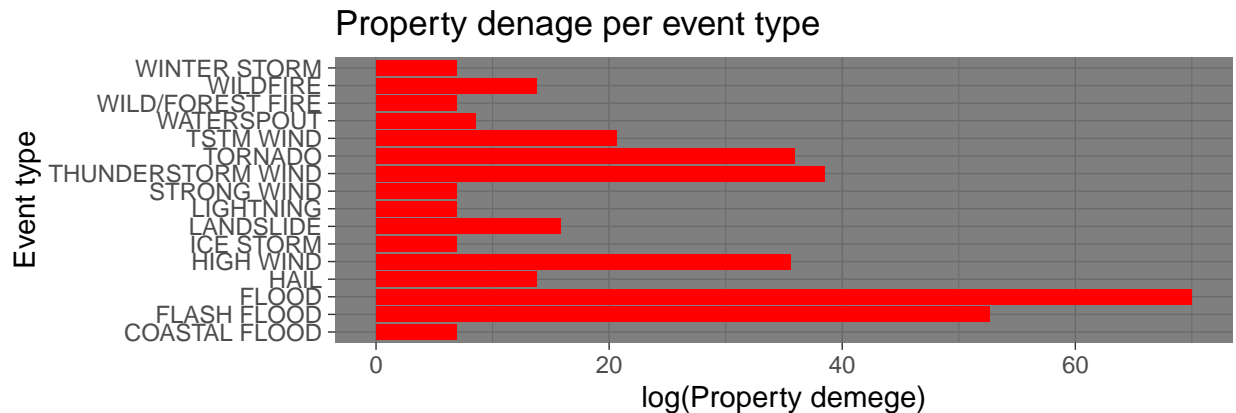
Which events have the greatest economic consequences.

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

i <- ggplot(crop_damage_99, aes(x=log(CROPDGM), y=EVTYPE)) +
```

```
geom_bar(fill="red",stat="identity" ) +
ggtitle("Crop demmage per event type") +
xlab("log(Crop damage)") +
ylab("Event type") +
theme_dark()

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



Property damage:

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

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

Crop damage:

```
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.