NOAA Data Analysis

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

The goal of this report is to explore the NOAA Storm Database and answer some basic questions about severe weather events.

- Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

Dataset

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database 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.

Data Processing

1. obtain the data

```
url <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
dest<- 'StormData.csv.bz2'
#download.file(url, dest, method = "wget")
```

2. read in data

```
storm_data <- read.csv(bzfile(dest,open = 'r'), header=TRUE)
dat        <- storm_data[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "CROPDMG")]
dat$EVTYPE  <- as.factor(tolower(dat$EVTYPE))
```

3. get the top 10

```
#
fatality_n  <- tapply(dat$FATALITIES, dat$EVTYPE, sum, na.rm = TRUE)
injury_n    <- tapply(dat$INJURIES, dat$EVTYPE, sum, na.rm = TRUE)
propdamage  <- tapply(dat$PROPDMG, dat$EVTYPE, sum , na.rm = TRUE)
cropdamage  <- tapply(dat$CROPDMG, dat$EVTYPE, sum , na.rm = TRUE)
#
fatality_10 <- sort(fatality_n, decreasing =  TRUE)[1:10]
injury_10   <- sort(injury_n, decreasing =  TRUE)[1:10]
propdam_10  <- sort(propdamage, decreasing = TRUE)[1:10]
cropdam_10  <- sort(cropdamage, decreasing = TRUE)[1:10]
```

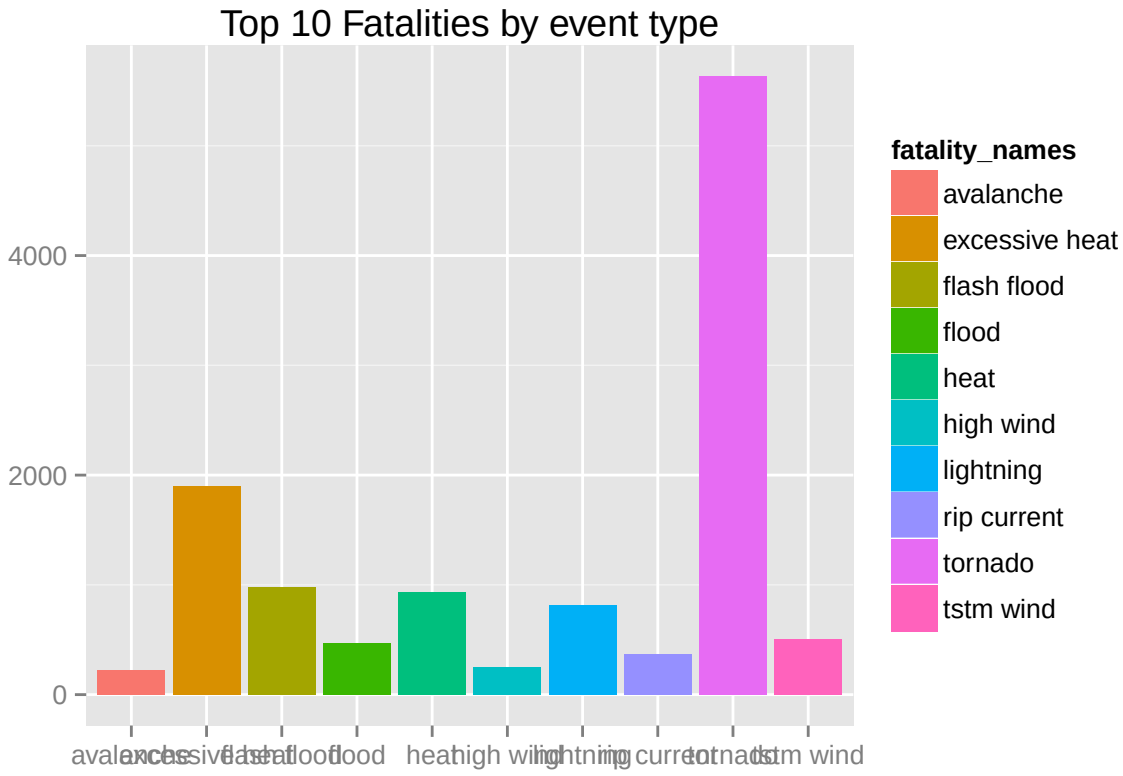
Results

- which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

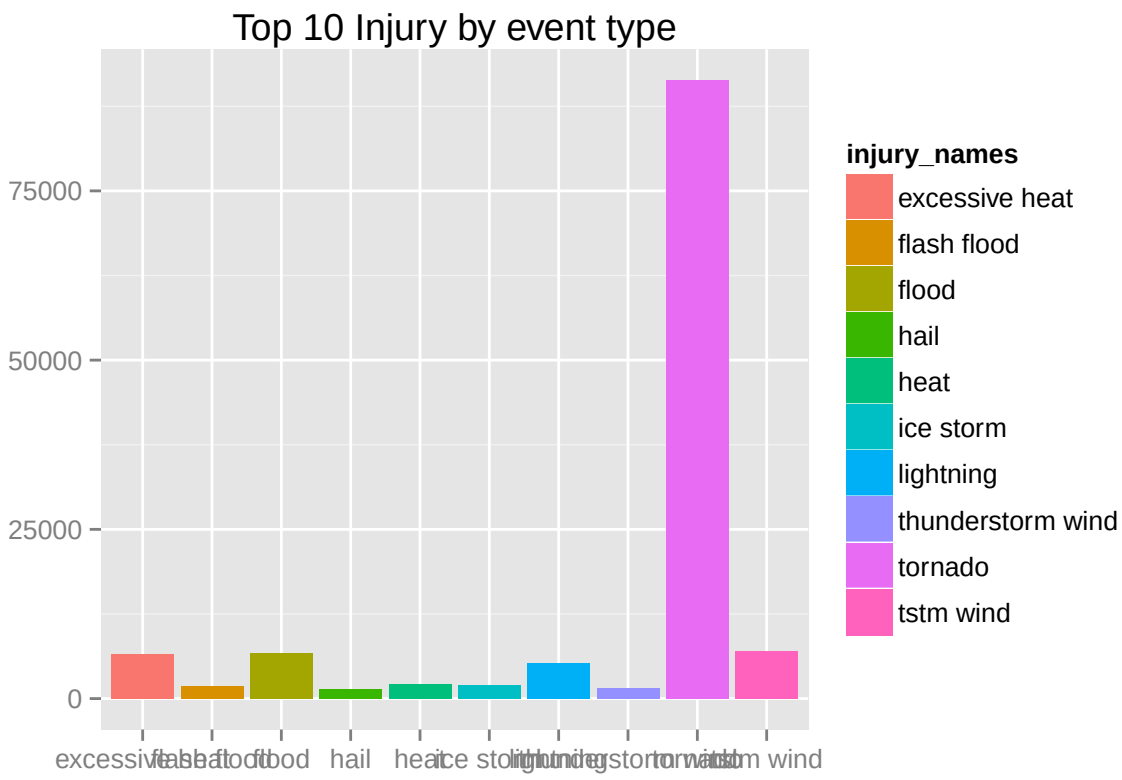
```
library(ggplot2)
library(gridExtra)
```

```
## Loading required package: grid
```

```
fatality_names <- as.factor(names(fatality_10))
fatality_plot  <- qplot(x = fatality_names, y = fatality_10, fill = fatality_names, geom="bar", stat="i
      ylab="", xlab="", main = "Top 10 Fatalities by event type")
injury_names   <- as.factor(names(injury_10))
injury_plot    <- qplot(x = injury_names, y = injury_10, fill = injury_names, geom="bar", stat="identity
      ylab="", xlab="", main = "Top 10 Injury by event type")
#grid.arrange(fatality_plot, injury_plot, ncol = 2)
fatality_plot
```



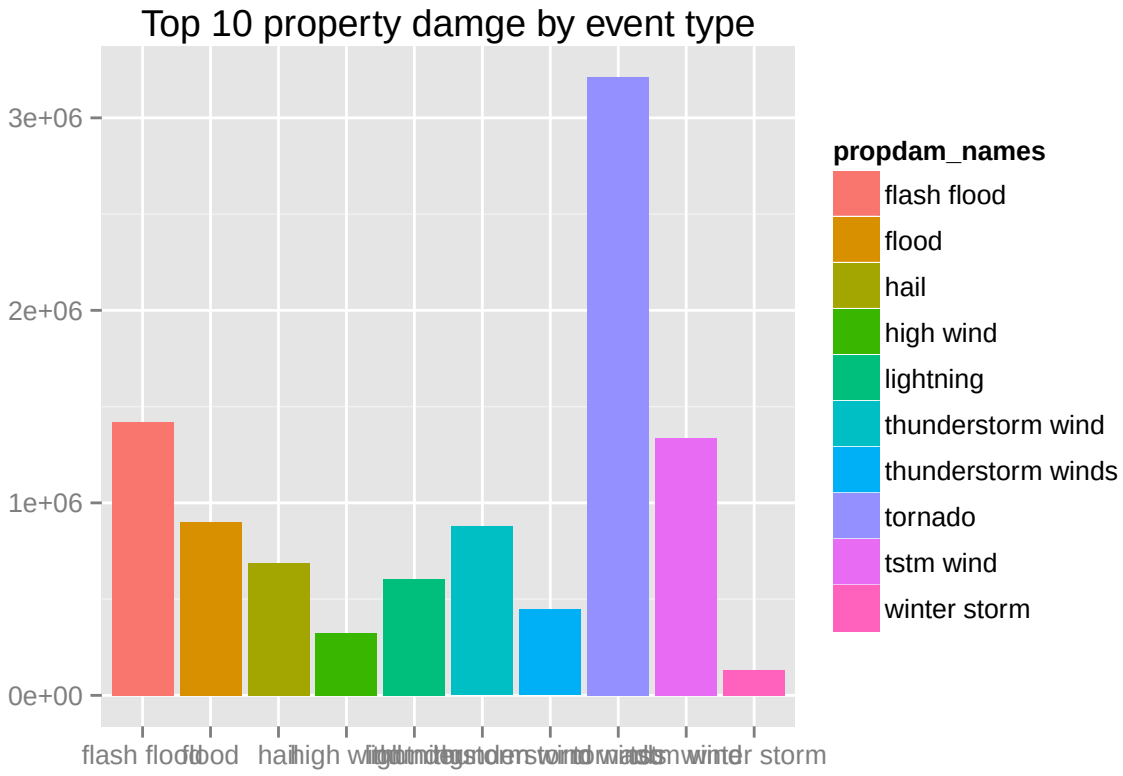
```
injury_plot
```



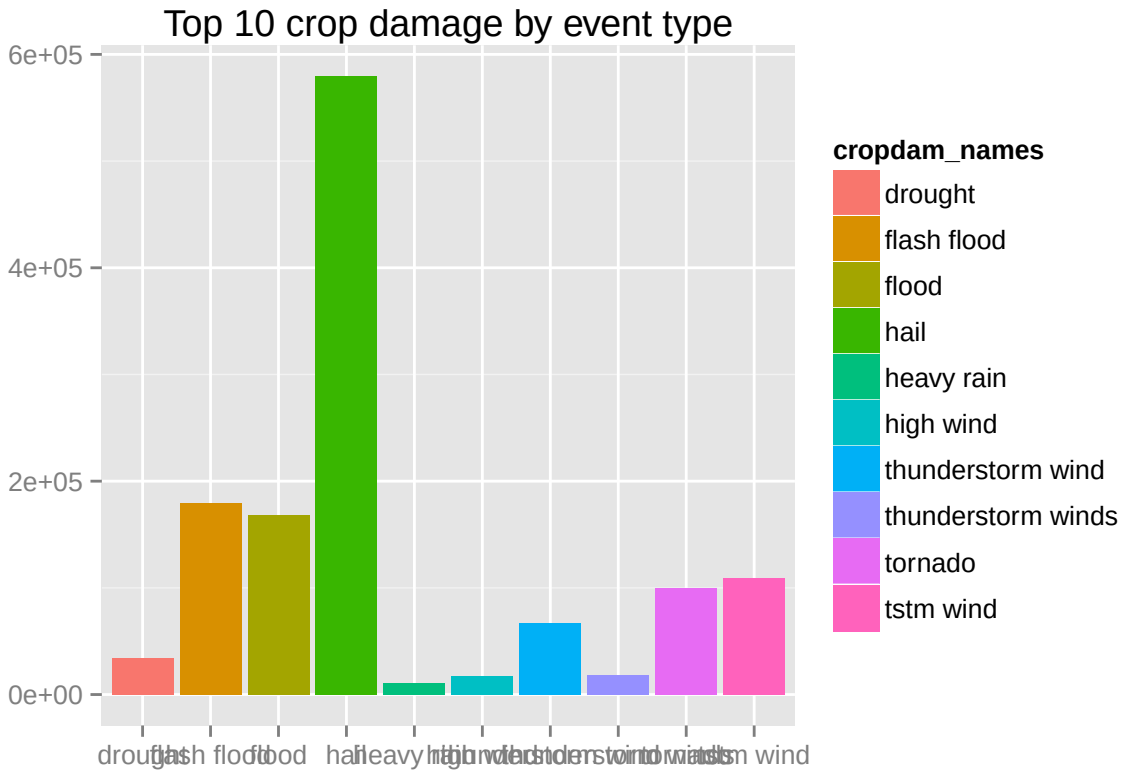
As the figure shows, tornado is most harmful.

- which types of events have the greatest economic consequences?

```
propdam_names <- as.factor(names(propdam_10))
propdam_plot  <- qplot(x = propdam_names, y = propdam_10, fill = propdam_names, geom="bar", stat="identi
                       ylab="", xlab="", main = "Top 10 property damge by event type")
cropdam_names   <- as.factor(names(cropdam_10))
cropdam_plot    <- qplot(x = cropdam_names, y = cropdam_10, fill = cropdam_names, geom="bar", stat="ide
                       ylab="", xlab="", main = "Top 10 crop damage by event type")
#grid.arrange(propdam_plot, cropdam_plot, ncol = 2)
propdam_plot
```




```
cropdam_plot
```



As the figure shows, tornado has most damge to human property, while hail has most damge to crops.