Reproducible Research: Assignment 2

Analysis of the NOAA storm database to estimate fatalities, injuries, and property damage due to severe weather events.

Introduction

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.

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.

In order to analyze the NOAA storm database we'll pose and answer the follwing questions:

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

Data Processing

The data used for this analysis come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. There is also some documentation of the database available. By clicking the following link you will find how some of the variables are constructed/defined: https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years can be considered more complete.

Loading packages

```
library(dplyr)
library(ggplot2)
```

Loading data

```
data <- read.csv("C:/Users/ldewit/Documents/coursera_local/Data/repdata-data-StormData.csv.bz2", header
```

Harmful effects with respect to population health

For this part, only features related to event type, population health, and economic consequences are relevant.

```
data_health <- select(data, "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CRO
```

Aggregate data per event

```
fatalities <- aggregate(FATALITIES ~ EVTYPE, data_health, sum, na.rm=TRUE)
fatalities <- fatalities[order(fatalities$FATALITIES, decreasing = TRUE),][1:10,]

injuries <- aggregate(INJURIES ~ EVTYPE, data_health, sum, na.rm=TRUE)
injuries <- injuries[order(injuries$INJURIES, decreasing = TRUE),][1:10,]
```

Greatest economic consequence

As can be found in the documentation, the characters in the columns PROPDMGEXP and CROPDMGEXP imply the magnitude of the numbers PROPDMG. Combining these columns gives an idea of the economic consequence.

Property

```
data$pd <- 0
data[data$PROPDMGEXP == "H", ]$pd <-
        data[data$PROPDMGEXP == "H", ]$PROPDMG * 10^2

data[data$PROPDMGEXP == "K", ]$pd <-
        data[data$PROPDMGEXP == "K", ]$PROPDMG * 10^3

data[data$PROPDMGEXP == "M", ]$pd <-
        data[data$PROPDMGEXP == "M", ]$PROPDMG * 10^6

data[data$PROPDMGEXP == "B", ]$pd <-
        data[data$PROPDMGEXP == "B", ]$PROPDMG * 10^9
```

Crops

```
data$crops <- 0
data[data$CROPDMGEXP == "H", ]$crops <-
        data[data$CROPDMGEXP == "H", ]$CROPDMG * 10^2

data[data$CROPDMGEXP == "K", ]$crops <-
        data[data$CROPDMGEXP == "K", ]$CROPDMG * 10^3

data[data$CROPDMGEXP == "M", ]$crops <-
        data[data$CROPDMGEXP == "M", ]$CROPDMG * 10^6

data[data$CROPDMGEXP == "B", ]$crops <-
        data[data$CROPDMGEXP == "B", ]$CROPDMG * 10^9
```

Next, we'll aggregate the sum of property and crop damages and then take the top 10 most damaging events again.

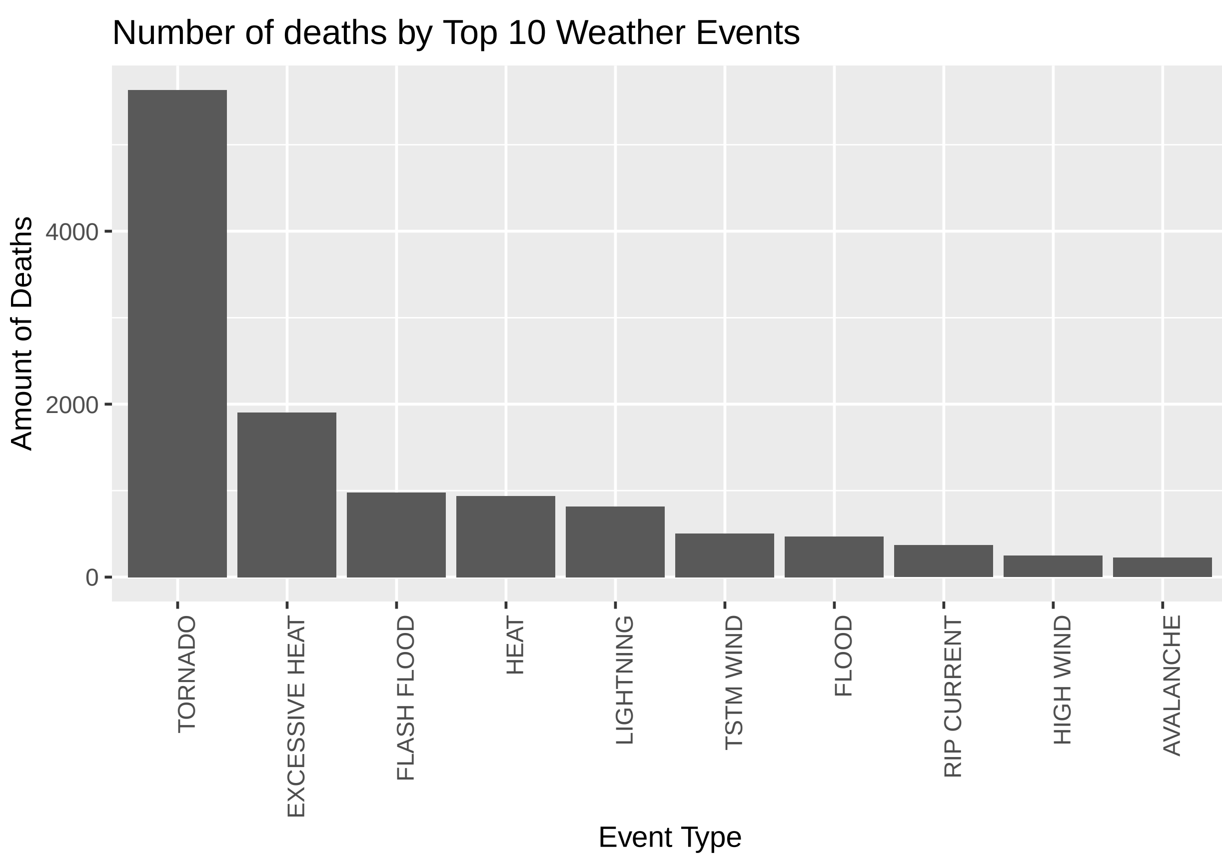
```
data_damage <- aggregate(pd + crops ~ EVTYPE, data = data, sum)
names(data_damage) <- c("EVTYPE", "ECON_DAM")
data_damage <- data_damage[order(data_damage$ECON_DAM, decreasing = TRUE), ][1:10, ]
data_damage$EVTYPE <- factor(data_damage$EVTYPE, levels = data_damage$EVTYPE)
```

Results

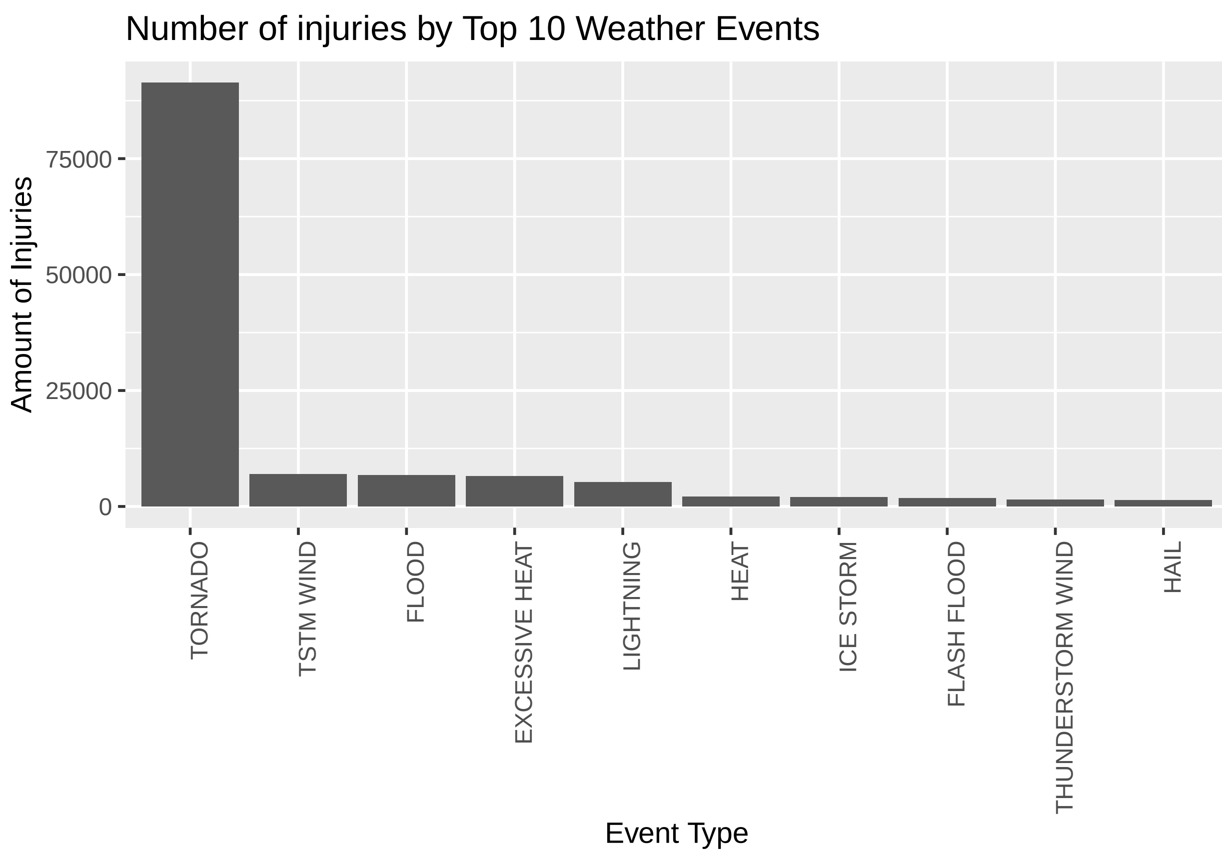
To demonstrate the effects of weather events on health and economic factors we'll plot our aggregated data.

Health

```r
ggplot(fatalities, aes(x=reorder(EVTYPE, -FATALITIES), y=FATALITIES)) +
        geom_bar(stat = "identity") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        xlab("Event Type") + ylab("Amount of Deaths") +
        ggtitle("Number of deaths by Top 10 Weather Events")
```

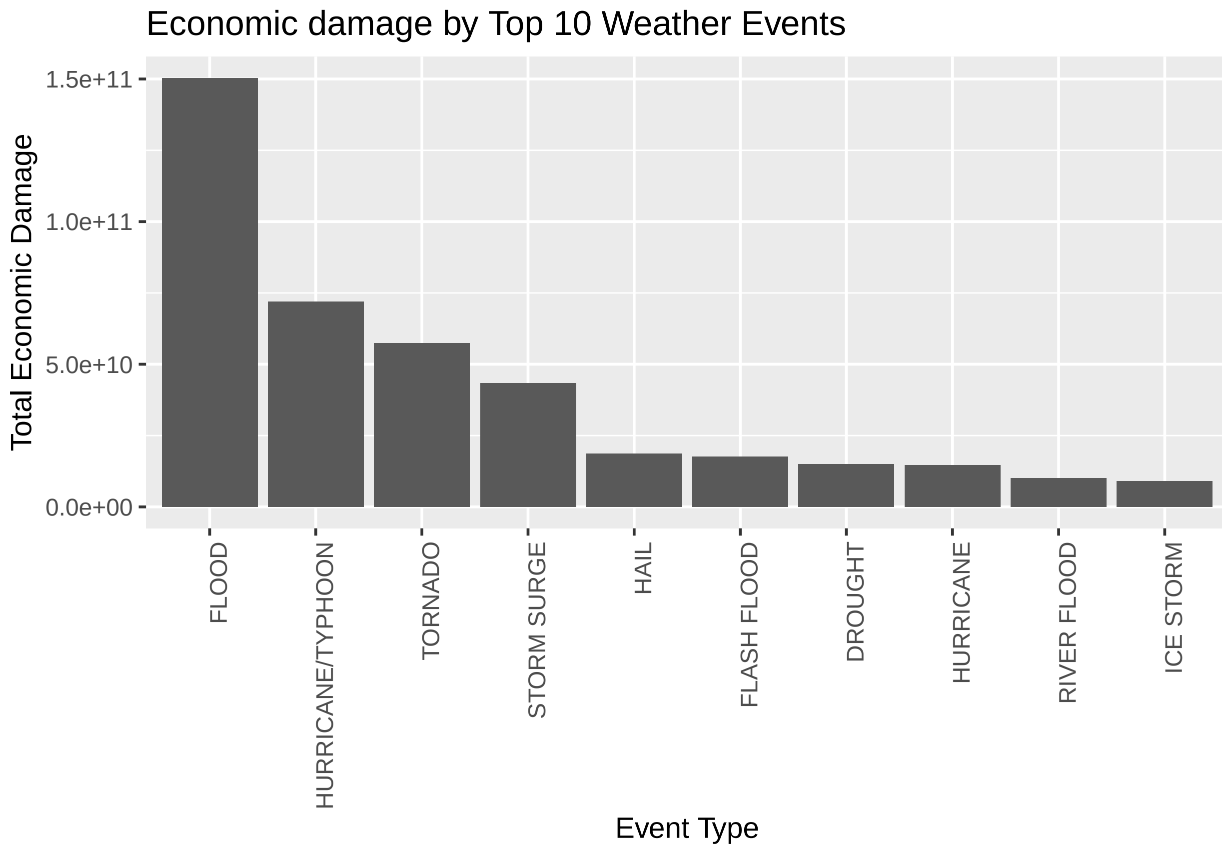



```r
ggplot(injuries, aes(x=reorder(EVTYPE, -INJURIES), y=INJURIES)) +
        geom_bar(stat = "identity") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        xlab("Event Type") + ylab("Amount of Injuries") +
        ggtitle("Number of injuries by Top 10 Weather Events")
```



Economic Damage

```
ggplot(data_damage, aes(x=reorder(EVTYPE, -ECON_DAM), y=ECON_DAM)) +
        geom_bar(stat = "identity") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        xlab("Event Type") + ylab("Total Economic Damage") +
        ggtitle("Economic damage by Top 10 Weather Events")
```



Conclusion

Based on the figured above the conclusion can be drawn that tornadoes, by far, lead to the most deaths and injuries whereas floods cause most economic damage.