NOAA Storm Data Analysis

Synopsis

This is an data analysis for the NOAA storm data from 1950-2011. It focuses on two points:

1. Population casualties
2. Economic damage

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.

Data

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site: Storm Data (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

National Weather Service Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)

National Climatic Data Center Storm Events FAQ (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.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 should be considered more complete.

Data Processing

Loading required packages

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(knitr)
```

Setting working directory and reading data and seeing dimension of data

```
setwd("D:/Data Science Certification/datasciencecoursera/5_Reproducable Research/Assignment
 2")
data <- read.csv("StormData.csv.bz2")
dim(data)
```

```
## [1] 902297     37
```

Selecting only required columns from the entire dataset

```
storm_damage <- select(data, BGN_DATE,STATE, EVTYPE,
                       FATALITIES, INJURIES, PROPDMG, PROPDMGEXP,CROPDMG,CROPDMGEXP)
```

In order to get correct values for PROPDMG and CROPDMG, we need to also consider values in PROPDMGEXP and CROPDMGEXP. Based on PROPDMGEXP to be H,K,M,B we multiply the corresponding value to make it that many billions. Here billion is used because, it will be easy to visualize in higher units in graphs.

Initialize to 0.

```
storm_damage$PROP_US <- 0
storm_damage$CROP_US <- 0
storm_damage$damage  <- 0
storm_damage$health  <- 0
```

Performing the required calculation

```
#1 billion = 1,000,000,000
storm_damage$PROP_US <- ifelse(storm_damage$PROPDMGEXP =="H" | storm_damage$PROPDMGEXP =="h",
                               storm_damage$PROPDMG*0.0000001, storm_damage$PROP_US)
storm_damage$CROP_US <- ifelse(storm_damage$CROPDMGEXP =="H"|  storm_damage$CROPDMGEXP =="h",
                               storm_damage$CROPDMG*0.0000001, storm_damage$CROP_US)
storm_damage$PROP_US <- ifelse(storm_damage$PROPDMGEXP =="K"|  storm_damage$PROPDMGEXP =="k",
                               storm_damage$PROPDMG*0.000001,  storm_damage$PROP_US)
storm_damage$CROP_US <- ifelse(storm_damage$CROPDMGEXP =="K"|  storm_damage$CROPDMGEXP =="k",
                               storm_damage$CROPDMG*0.000001,  storm_damage$CROP_US)
storm_damage$PROP_US <- ifelse(storm_damage$PROPDMGEXP =="M"|  storm_damage$PROPDMGEXP =="m",

                               storm_damage$PROPDMG*0.001,     storm_damage$PROP_US)
storm_damage$CROP_US <- ifelse(storm_damage$CROPDMGEXP =="M"|  storm_damage$CROPDMGEXP =="m",

                               storm_damage$CROPDMG*0.001,     storm_damage$CROP_US)
storm_damage$PROP_US <- ifelse(storm_damage$PROPDMGEXP =="B"|  storm_damage$PROPDMGEXP =="b",

                               storm_damage$PROPDMG*1,         storm_damage$PROP_US)
storm_damage$CROP_US <- ifelse(storm_damage$CROPDMGEXP =="B"|  storm_damage$CROPDMGEXP =="b",
                               storm_damage$CROPDMG*1,         storm_damage$CROP_US)
```

Adding up the values to get final numbers for health and properties (and) crop damage

```
storm_damage$health <- storm_damage$FATALITIES + storm_damage$INJURIES
storm_damage$damage <- storm_damage$PROP_US + storm_damage$CROP_US
```

Finding total health and properties value based on individual event type.

```
health <- aggregate(storm_damage$health, by=list(storm_damage$EVTYPE), FUN = sum)
health <- arrange(health, desc(x))
health <- head(health,10)
health <- select(health, Event_Type = Group.1, Number_of_Injuries= x)
```

Finding total crop damage value based on individual event type.

```
storm_PDMG <- aggregate(storm_damage$damage, by=list(storm_damage$EVTYPE), FUN = sum)
storm_PDMG <- arrange(storm_PDMG, desc(x))
storm_PDMG <- head(storm_PDMG,10)
storm_PDMG <- select(storm_PDMG, Event_Type = Group.1, Economic_Impact= x)
```

Results

Top 10 harmful type base on the sum of casualties.

```
health
```

```
##           Event_Type Number_of_Injuries
## 1            TORNADO              96979
## 2     EXCESSIVE HEAT               8428
## 3          TSTM WIND               7461
## 4              FLOOD               7259
## 5          LIGHTNING               6046
## 6               HEAT               3037
## 7        FLASH FLOOD               2755
## 8          ICE STORM               2064
## 9  THUNDERSTORM WIND               1621
## 10      WINTER STORM               1527
```

Top 10 harmful type base on the sum of damage.

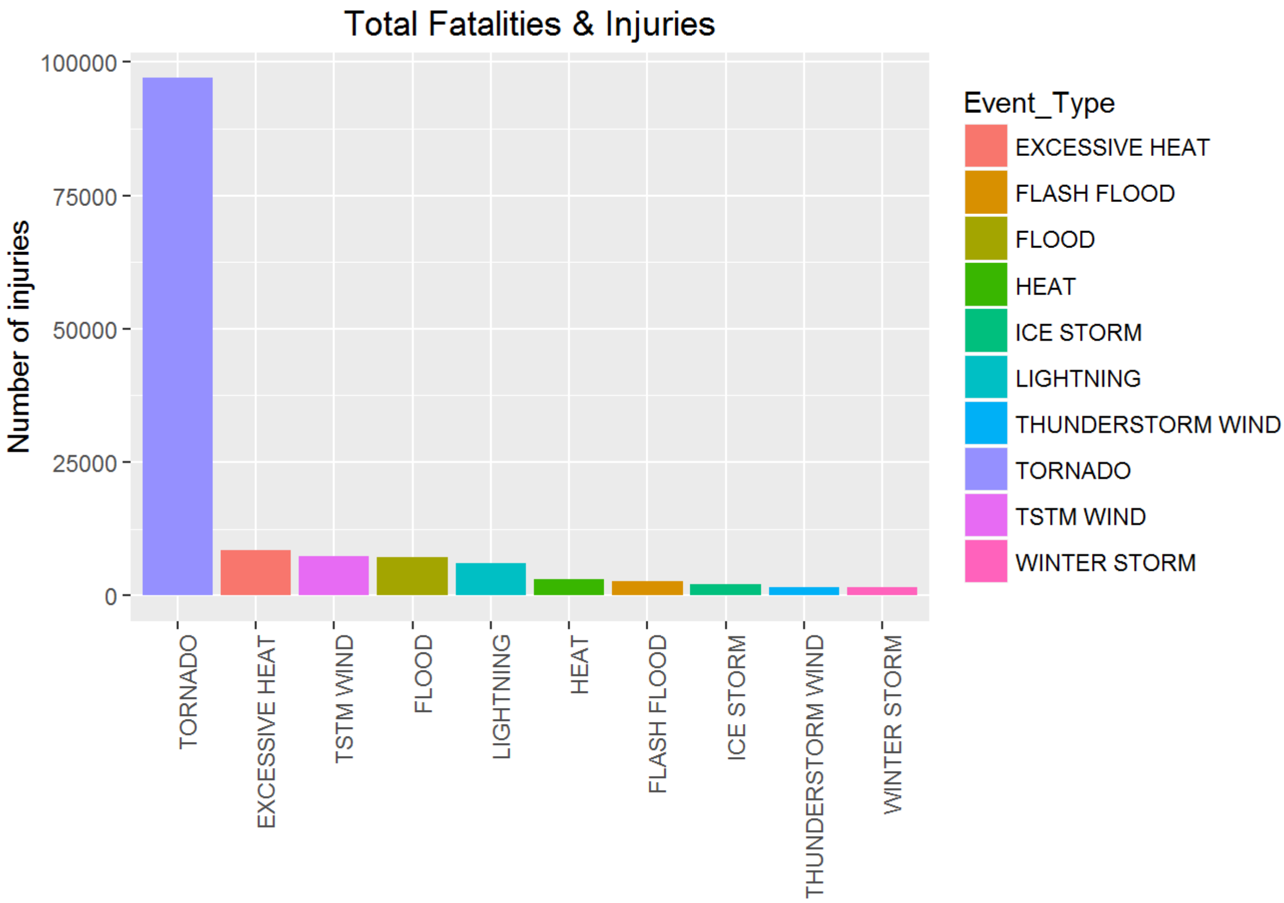
```
storm_PDMG
```

```
##           Event_Type Economic_Impact
## 1              FLOOD      150.319678
## 2  HURRICANE/TYPHOON       71.913713
## 3            TORNADO       57.352114
## 4        STORM SURGE       43.323541
## 5               HAIL       18.758222
## 6        FLASH FLOOD       17.562129
## 7            DROUGHT       15.018672
## 8          HURRICANE       14.610229
## 9        RIVER FLOOD       10.148404
## 10         ICE STORM        8.967041
```

Bar plot of the 10 most important events in United States

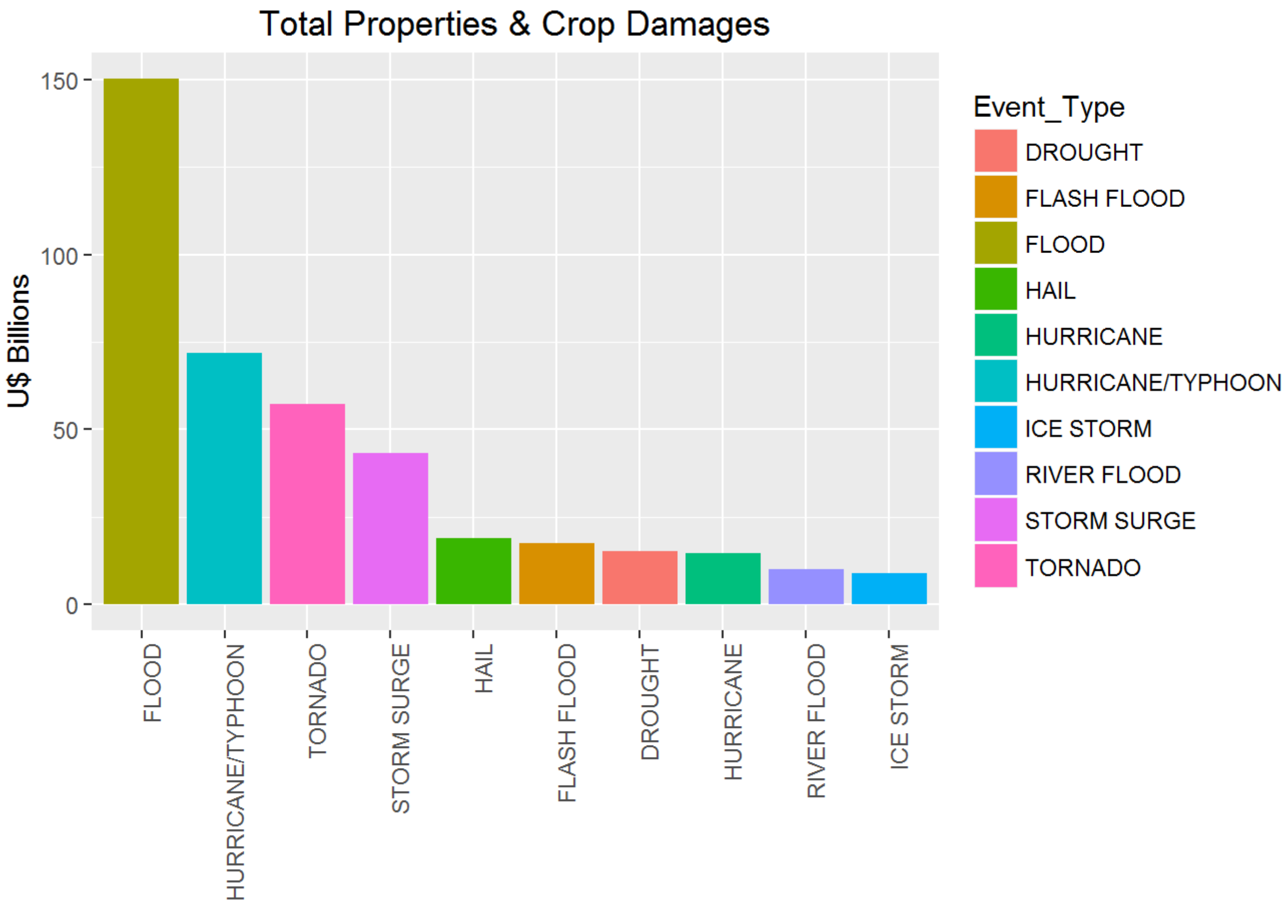
1. Population casualities

```
g<- ggplot(health, aes(reorder(Event_Type,-Number_of_Injuries), Number_of_Injuries)) +
    labs(title="Total Fatalities & Injuries") +
    xlab("") + ylab("Number of injuries")
plot1<- g + geom_bar(aes(fill=Event_Type),stat="identity") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(plot1)
```



2. Economic damage

```
g<- ggplot(storm_PDMG, aes(reorder(Event_Type,-Economic_Impact), Economic_Impact)) +
    labs(title="Total Properties & Crop Damages") +
    xlab("") + ylab("U$ Billions")
plot2<- g + geom_bar(aes(fill=Event_Type), stat="identity") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(plot2)
```



Summary

- Tornado is the highest for Economic damage.
- Flood is the highest for Properties and crop damage.