Project 2

Sean

September 30, 2018

Health and economic impacts caused by severe weather events in the US (1950 - 2011)

Synopsis

This report compares servere weather events, in respect to economic costs and populations health in the United States (US) between 1950 and 2011.

Hypothesis

The hypothesis is that a prevention for unregularly weather events like floods is hard to establish. Tornados on the other hand are highly destructive and destroy large areas in a matter of hours. Therefore I assume that those events cause the most economic damages, while extreme heat is assumed to be the most harmful to people's health. This assumptions are based on thefact that measures taken to prevent people being harmed by tornados should be in place by now, since they occur on a regular basis. This does not apply to extreme heat, which occurs occasionally.

Results

We obtained storm data from the National Oceanic and Atmosphere Administration from 1950 - 2011 to verify the hypothesis. Factis that floods (on property) and droughts (on crop) caused the most economic damages, while the most harmful events to thepopulation's health was caused by tornados (injuries as well as fatalities). The main cost driver in the area of property damage was caused by a series ofstorms end of 2005, which produced significant runoff over much of northern California (NAPA). Floods can spread across large areas and are somewhat uncontrollable. They occur on a unregularly basis. Tacking measures to prevention floods as such costs billions. On the contrary, the harm on the population is relatively small. Tornadoes, on the other hand, are also able to destroy large areas, but occur relatively frequently in the US. Therefore the prevention on people's health should be manageable, but the figures show a different picture. The measures taken to prevent people being harmed seems inadequate, or storm warnings are being activated far too late. Another possible cause might be that not enough shelters in homes and public places might be available.

Data Processing

Load Data

The data and further information were obtained from: * **Dataset** Storm Data[47Mb] * **National Weather Service** Storm Data Documentation * **National Climate Data Center Storm Events** FAQ

Load packages:

```
library(knitr)
library(formatR)
library(plyr)
```

```r
library(dplyr)
library(ggplot2)
library(gridExtra)
```

Download data and upzip

```r
setwd(paste0("E:/Courses/Coursera/John_Hopkins_University_Data_Science",
    "/5_Reproducible_Research/Week 4"))
zipfile <- "repdata-data-StormData.csv.bz2"

# if(!file.exists(paste0('F:/Courses/Coursera/John_Hopkins_University_Data_Science',
# '/5_Reproducible_Research/Week 4',
# '/repdata-data-StormData.csv.bz2'))){ dataURL <-
# 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
# download.file(dataURL, zipfile, mode = 'wb')

# }

# Load storm data
storm_data <- read.csv(bzfile(zipfile), header = TRUE)
```

Only a subset of columns is needed to be extracted from the dataset, including: **'EVTYPE'**, **'INJURIES'**, **'FATALITIES'**, **'PROPDMG'**, **'PROPDMGEXP'**, **'CROPDMG'**, **'CROPDMGEXP'**.

```r
#Define the columns
Colm <- c('EVTYPE',
          'INJURIES',
          'FATALITIES',
          'PROPDMG',
          'PROPDMGEXP',
          'CROPDMG',
          'CROPDMGEXP')

#Remove unnecessary columns
data <- storm_data[, Colm]
```

Data aggregation

Take into account the order of magnitude of property and crop damage (H = hundreds, K = thousands, M = millions, B= billions), as well as the injury and fatality data, and clean the dataset for plotting graphs.

```r
#################################
# Property and crop damage data#
#################################

# Convert to lowercase
data$PROPDMGEXP <- lapply(data$PROPDMGEXP, function(x) tolower(as.character(x)))
data$CROPDMGEXP <- lapply(data$CROPDMGEXP, function(x) tolower(as.character(x)))

# Prepare magnitudes for property damage
data$PROPDMGEXP[data$PROPDMGEXP %in% c('', '+', '-', '?')] <- "0"
data$CROPDMGEXP[data$CROPDMGEXP %in% c('', '?')] <- '0'
```

```
data$PROPDMGEXP[data$PROPDMGEXP %in% c('h')] <- '2'
data$PROPDMGEXP[data$PROPDMGEXP %in% c('k')] <- '3'
data$PROPDMGEXP[data$PROPDMGEXP %in% c('m')] <- '6'
data$PROPDMGEXP[data$PROPDMGEXP %in% c('b')] <- '9'

data$CROPDMGEXP[data$CROPDMGEXP %in% c('k')] <- '3'
data$CROPDMGEXP[data$CROPDMGEXP %in% c('m')] <- '6'
data$CROPDMGEXP[data$CROPDMGEXP %in% c('b')] <- '9'

# Convert to vector
data$PROPDMGEXP <- unlist(data$PROPDMGEXP, use.names = TRUE)
data$CROPDMGEXP <- unlist(data$CROPDMGEXP, use.names = TRUE)

# Calculate damage
data$PROPDMGEXP <- 10 ^ (as.numeric(data$PROPDMGEXP))
data$PROPDMG = as.numeric(data$PROPDMG) * data$PROPDMGEXP

data$CROPDMGEXP <- 10 ^ (as.numeric(data$CROPDMGEXP))
data$CROPDMG = as.numeric(data$CROPDMG) * data$CROPDMGEXP

# Remove unused columns
data$PROPDMGEXP <- NULL
data$CROPDMGEXP <- NULL

# convert EVTYPE fields to upper case
data$EVTYPE <- lapply(data$EVTYPE, function(x) toupper(as.character(x)))

# Convert to vector and as a factor
data$EVTYPE <- as.factor(unlist(data$EVTYPE, use.names = TRUE))

# Subsetting data and deviding by billion
Property_DMG <- ddply(data, .(EVTYPE), summarise, DMG = round((sum(PROPDMG) / 1e+9), 1))
Property_DMG <- subset(Property_DMG, DMG > 0)

#Subsetting data and deviding by million
Crop_DMG <- ddply(data, .(EVTYPE), summarise, DMG = round((sum(CROPDMG) / 1e+6), 1))
Crop_DMG <- subset(Crop_DMG, DMG > 0)

# Sort descendingly and filter top 10 (Property and Crop)
Property_DMG <- Property_DMG[order( -Property_DMG$DMG),]
data_Property_DMG <- filter(Property_DMG, DMG > 1) %>% top_n(10)
```

```
## Selecting by DMG
```

```
Crop_DMG <- Crop_DMG[order( - Crop_DMG$DMG),]
data_Crop_DMG <- filter(Crop_DMG, DMG > 1) %>% top_n(10)
```

```
## Selecting by DMG
```

Same applies to the injury and fatality data

```
#####################################
# Property injury and fatality data #
#####################################

# Subsetting data
inj_DMG <- ddply(data, .(EVTYPE), summarise, Injuries = round((sum(INJURIES)/1), 2))
fatal_DMG <- ddply(data, .(EVTYPE), summarise, Fatalities = round(sum(FATALITIES), 2))

# Sorting descendingly and filter top 10
inj_DMG <- inj_DMG[order( - inj_DMG$Injuries),]
data_inj_DMG <- filter(inj_DMG, Injuries > 1) %>% top_n(10)
```

```
## Selecting by Injuries
```

```
fatal_DMG <- fatal_DMG[order( - fatal_DMG$Fatalities),]
data_fatal_DMG <- filter(fatal_DMG, Fatalities > 1) %>% top_n(10)
```

```
## Selecting by Fatalities
```

```
data_inj <- filter(inj_DMG, Injuries > 1000)
data_fatal <- filter(fatal_DMG, Fatalities > 100)

#Clean unnecessary variables
rm(fatal_DMG)
rm(inj_DMG)
rm(Crop_DMG)
rm(Property_DMG)
```

Now, we are ready to plot.

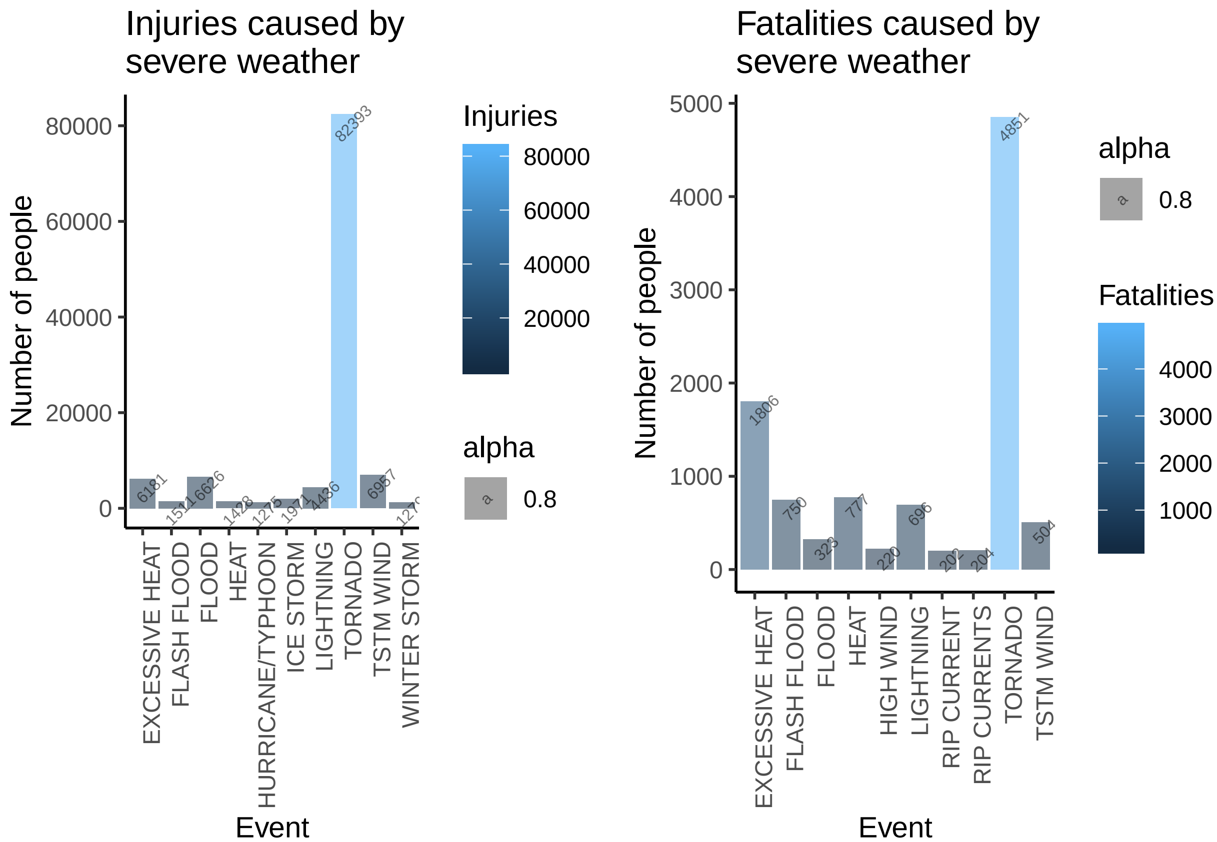
Results and Conclusions

1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

This will plot which of severe weather events causes the most harm to the population's health. Data is based on fatality and injurt, and the plot demonstrates the top 10 severe weather events with the highest fatality and injury rates.

```
#####################################
#   Plot both injury and fatality   #
#####################################

inj_plot <- ggplot(data = data_inj_DMG,
                  aes(x = data_inj_DMG$EVTYPE,
                      y = data_inj_DMG$Injuries,
                      fill = Injuries,
                      alpha = 0.8)) +
        geom_bar(stat = "identity") +
        labs(title = "Injuries caused by \nsevere weather",
             x = "Event",
```

```r
             y = "Number of people") +
        theme_bw() +
        theme(plot.background = element_blank(),
              panel.border = element_blank(),
              panel.grid.major = element_blank(),
              panel.grid.minor = element_blank(),
              axis.line = element_line(colour = "black")) +
        theme(axis.text.x = element_text(angle = 90,
                                         hjust = 1)) +
        geom_text(aes(label = Injuries),
                  angle = 45,
                  vjust = 1.6,
                  color = "black",
                  position = position_dodge(0.9),
                  size = 2)

fatal_plot <- ggplot(data = data_fatal_DMG,
                     aes(x = data_fatal_DMG$EVTYPE,
                         y = data_fatal_DMG$Fatalities,
                         fill = Fatalities,
                         alpha = 0.8)) +
        geom_bar(stat = "identity") +
        labs(title = "Fatalities caused by \nsevere weather",
             x = "Event",
             y = "Number of people") +
        theme_bw() +
        theme(plot.background = element_blank(),
              panel.border = element_blank(),
              panel.grid.major = element_blank(),
              panel.grid.minor = element_blank(),
              axis.line = element_line(colour = "black")) +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        geom_text(aes(label = Fatalities),
                  angle = 45,
                  vjust = 1.6,
                  color = "black",
                  position = position_dodge(0.9),
                  size = 2)

grid.arrange(inj_plot, fatal_plot, ncol = 2)
```



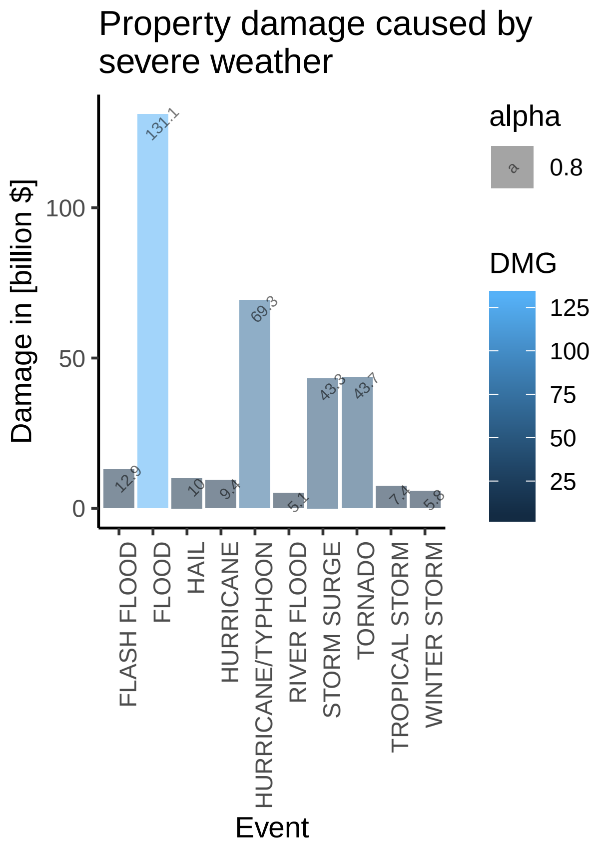
The above plot clearly illustrated that ***Tornado*** and ***Heat*** had caused most fatalities, and ***Tornado*** had caused the major injuries in the United States from 1995 to 2011.

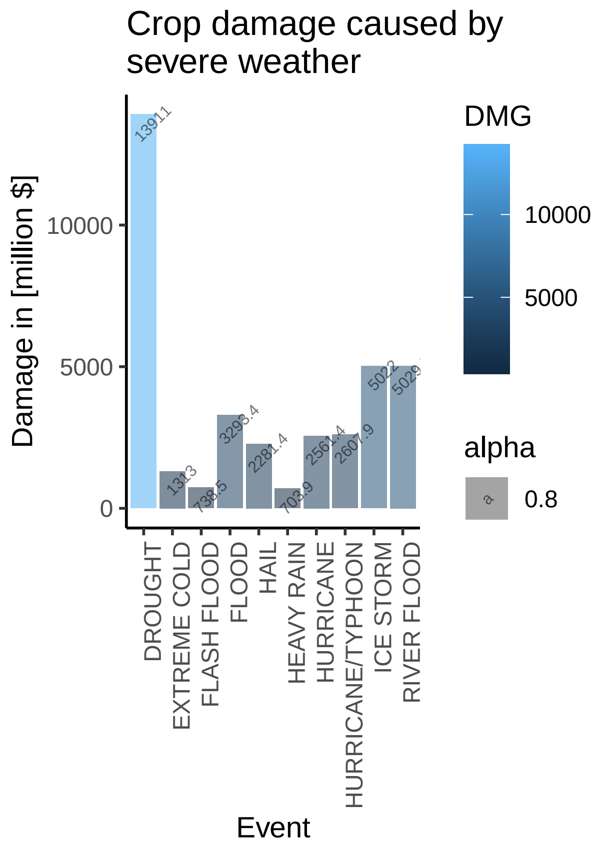
2. Across the United States, which types of events have the greatest economic consequences?

This will plot, which of severe weather events cause the most economic impact on property and crops, based on the top 10 economic damages.

```
############################################
#   Plot both property and crop damage   #
############################################

prop_plot <- ggplot(data = data_Property_DMG,
                    aes(x = data_Property_DMG$EVTYPE,
                        y = data_Property_DMG$DMG,
                        fill = DMG,
                        alpha = 0.8)) +
        geom_bar(stat = "identity") +
        labs(title = "Property damage caused by \nsevere weather",
             x = "Event",
             y = "Damage in [billion $]") +
        theme_bw() +
        theme(plot.background = element_blank(),
              panel.border = element_blank(),
              panel.grid.major = element_blank(),
              panel.grid.minor = element_blank(),
```

```
              axis.line = element_line(colour = "black")) +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        geom_text(aes(label = DMG),
                  angle = 45,
                  vjust = 1.6,
                  color = "black",
                  position = position_dodge(0.9),
                  size = 2)

crop_plot <- ggplot(data = data_Crop_DMG,
                    aes(x = data_Crop_DMG$EVTYPE,
                        y = data_Crop_DMG$DMG,
                        fill = DMG,
                        alpha = 0.8)) +
        geom_bar(stat = "identity") +
        labs(title = "Crop damage caused by \nsevere weather",
             x = "Event",
             y = "Damage in [million $]") +
        theme_bw() +
        theme(plot.background = element_blank(),
              panel.border = element_blank(),
              panel.grid.major = element_blank(),
              panel.grid.minor = element_blank(),
              axis.line = element_line(colour = "black")) +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        geom_text(aes(label = DMG),
                  angle = 45,
                  vjust = 1.6,
                  color = "black",
                  position = position_dodge(0.9),
                  size = 2)

grid.arrange(prop_plot, crop_plot, ncol = 2)
```





The plots indicate that ***Flood***, ***Hurricane/Typhoon***, ***Storm Surge***, and ***Tornado*** caused the greatest total economic damage on the property loss. And ***Drought***, ***Flood/River Flood***, and ***Ice Strom*** caused the greatest total economic damage on the crop loss.

Conclusion

From the analysis, ***Heat*** and ***Tornado*** are the two most harmful weather event that pose a major threat to uS population. The ***Flood*** and ***Drought*** have the greatest economic impacts on crops and properties.