Reproducible Research Week 4 – Assignment 2

**Disclaimer: because I was unable to install the R Markdown on my machine in order to create a .rmd file, I created a PDF document to simulate what the document might look like.

Introduction

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

This project involves exploring the US. 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, estimates of fatalities and injuries, and property damage.

Data

The data for this assignment comes in the form of a csv file compressed via the bzip2 algorithm to reduce file size. Download the file here: Storm Data

NWS Data Documentation and FAQs

*For this assignment, I looked at data from 1996 to 2011, earlier years of data were incomplete.

Data

Opening the data set, filtering out the data to use (1996 – 2011), and selecting out the data columns to use for analysis

```
stormdata <- read.csv("./ReproducibleResearch/StormData.csv.bz2")

data <- transform(stormdata, date=as.Date(BGN_DATE, format="%m/%d/%Y"))

data2 <- subset(data, data$date >= "1996-01-01")

datacols <- subset(data2, select=c(date, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP))
```

Health Related Analysis

Creating the dataset for health related items for analysis (deaths and injuries) and creating the dataset for damage related items (property and crop damage)

```
datahealth <- subset(datacols, !datacols$FATALITIES ==0 & !datacols$INJURIES ==0, select = c(EVTYPE, FATALITIES, INJURIES))

datacosts <- subset(datacols, !datacols$PROPDMG ==0 & !datacols$CROPDMG ==0, select = c(EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP))

datahealth_deaths <- aggregate(datahealth$FATALITIES, by = list(datahealth$EVTYPE), FUN = sum)

colnames(datahealth_deaths) <- c("TYPE", "FATALITIES")

datahealth_injuries <- aggregate(datahealth$INJURIES, by = list(datahealth$EVTYPE), FUN = sum)
```

colnames(datahealth_injuries) <- c("TYPE", "INJURIES")

Filtering data to show the top 5 types of weather events for both deaths and injuries

datahealth_deaths <- datahealth_deaths[order(datahealth_deaths$FATALITIES, decreasing = TRUE),][1:5,]

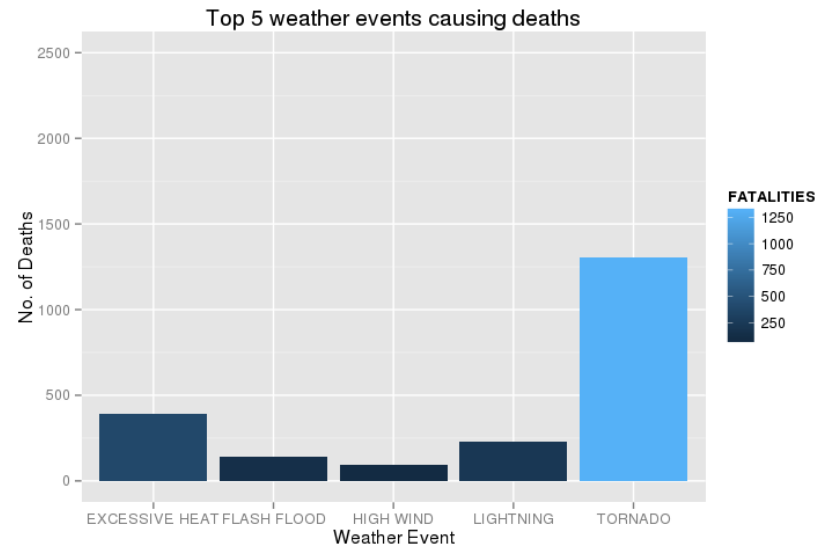
datahealth_injuries <- datahealth_injuries[order(datahealth_injuries$INJURIES, decreasing = TRUE),][1:5,]

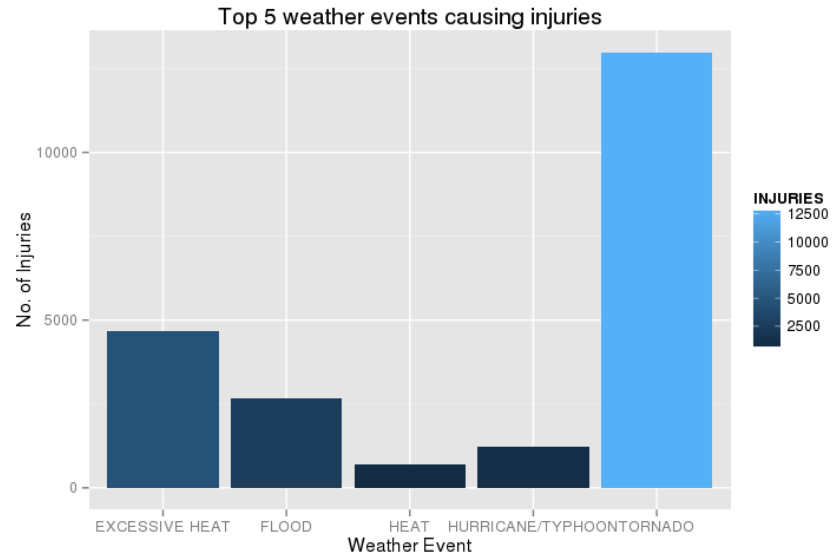
Analysis Question 1: Across the United States, which types for events are the most harmful with respect to the health of the population?

Create data plots for top 5 weather events that cause fatalities and injuries

ggplot(datahealth_deaths, aes(TYPE, FATALITIES, fill = FATALITIES)) + geom_bar(stat = "identity") + labs(x = "Weather Event", y = "No. of Deaths") + ggtitle("Top 5 weather events causing deaths") + theme_update() + expand_limits(y=c(0,2500))

ggplot(datahealth_injuries, aes(TYPE, INJURIES, fill = INJURIES)) + geom_bar(stat = "identity") + labs(x = "Weather Event", y = "No. of Injuries") + ggtitle("Top 5 weather events causing injuries") + theme_update() + expand_limits(y=c(0,13000))





Economic Losses Related Analysis

Selecting required data for economic analysis

```
datacosts_prop <- subset(datacosts, datacosts$PROPDMGEXP=="K"| datacosts$PROPDMGEXP=="k"|
datacosts$PROPDMGEXP=="M"| datacosts$PROPDMGEXP=="m"| datacosts$PROPDMGEXP=="B"|
datacosts$PROPDMGEXP== "b")
```

```
datacosts_crop <- subset(datacosts, datacosts$CROPDMGEXP=="K"| datacosts$CROPDMGEXP=="k"|
datacosts$CROPDMGEXP=="M"| datacosts$CROPDMGEXP=="m"| datacosts$CROPDMGEXP=="B"|
datacosts$CROPDMGEXP== "b")
```

Converting values into numbers (because the data had non-numerical values)

Values for Property damage:

```
datacosts_prop$PROPDMGEXP <- gsub("m", 1e+06, datacosts_prop$PROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_prop$PROPDMGEXP <- gsub("k", 1000, datacosts_prop$PROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_prop$PROPDMGEXP <- gsub("b", 1e+09, datacosts_prop$PROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_prop$PROPDMGEXP <- as.numeric(datacosts_prop$PROPDMGEXP)
```

Values for Crop damage:

```
datacosts_crop$CROPDMGEXP <- gsub("m", 1e+06, datacosts_crop$CROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_crop$CROPDMGEXP <- gsub("k", 1000, datacosts_crop$CROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_crop$CROPDMGEXP <- gsub("b", 1e+09, datacosts_crop$CROPDMGEXP, ignore.case = TRUE)
```

```
datacosts_crop$CROPDMGEXP <- as.numeric(datacosts_crop$CROPDMGEXP)
```

Creating a total of the loss costs (crop + property)

```
datacosts$TOTALDMG <- (datacosts$CROPDMG * datacosts_crop$CROPDMGEXP) +
  (datacosts$PROPDMG * datacosts_prop$PROPDMGEXP)
```

```
total_final <- aggregate(datacosts$TOTAL_DMG, by = list(datacosts$EVTYPE), FUN = sum)
```

```
colnames(total_final) <- c("TYPE", "TOTAL_DMG")
```

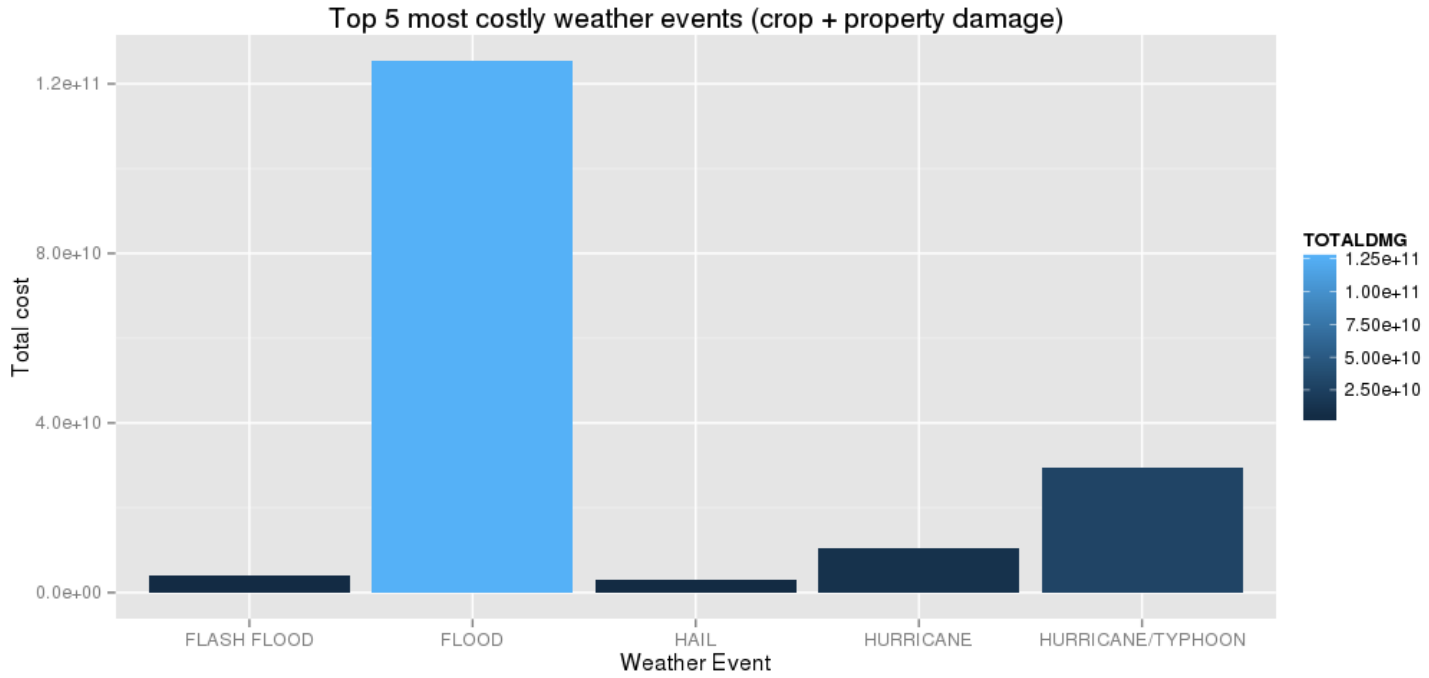
Ranking the top 5 events that cost the most in damage amounts

```
final_rank <- total_final[order(total_final$TOTAL_DMG, decreasing = TRUE), ][1:5, ]
```

Analysis Question 2: Across the United States, which types for events have the greatest economic impacts?

Creating plot of top 5 most costly types of weather events (crop + property damage)

ggplot(final_rank, aes(TYPE, TOTAL_DMG, fill = TOTAL_DMG)) + geom_bar(stat = "identity") + labs(x = "Weather Event", y = "Total cost") + ggtitle("Top 5 most costly weather events (crop + property damage)") + theme_update() + expand_limits(y=c(0,250000))



Summary

Different types of weather events can be costly both in terms of health and property.

This particular analysis revealed that tornadoes and excessive heat are the most impactful on the health of Americans. While, floods and hurricanes/typhoons are the most costly in terms of economic impacts.