

## Reproducible Research Project 2

### Data Processing

Download data processing set and load it to R.

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "FStormData.csv.bz2")
data <- read.csv("FStormData.csv.bz2")
dim(data)

## [1] 902297      37

head(data)

## STATE__      BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
EVTYPE BGN_RANGE BGN_AZI
## 1      1 4/18/1950 0:00:00    0130     CST    97    MOBILE
AL TORNADO      0
## 2      1 4/18/1950 0:00:00    0145     CST     3    BALDWIN
AL TORNADO      0
## 3      1 2/20/1951 0:00:00    1600     CST    57    FAYETTE
AL TORNADO      0
## 4      1  6/8/1951 0:00:00    0900     CST    89    MADISON
AL TORNADO      0
## 5      1 11/15/1951 0:00:00    1500     CST    43    CULLMAN
AL TORNADO      0
## 6      1 11/15/1951 0:00:00    2000     CST    77 LAUDERDALE
AL TORNADO      0
      BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN END_RANGE END_AZI
END_LOCATI LENGTH WIDTH F MAG
## 1      14.0    100 3    0      0      NA      0
## 2      2.0    150 2    0      0      NA      0
## 3      0.1    123 2    0      0      NA      0
## 4      0.0    100 2    0      0      NA      0
## 5      0.0    150 2    0      0      NA      0
## 6      1.5    177 2    0      0      NA      0
      FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO
STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1      3040      8812      15    25.0      K      0
## 2      3042      8755      0     2.5      K      0
## 3      3042      8755      2    25.0      K      0
```

3340	8742				
## 4	0	2	2.5	K	0
3458	8626				
## 5	0	2	2.5	K	0
3412	8642				
## 6	0	6	2.5	K	0
3450	8748				
	LATITUDE_E	LONGITUDE_	REMARKS	REFNUM	
## 1	3051	8806		1	
## 2	0	0		2	
## 3	0	0		3	
## 4	0	0		4	
## 5	0	0		5	
## 6	0	0		6	

## Compile Data on Fatalities and Calculate the Impact of People

Only looking at the seven events on fatalities pertaining to weather.

```
fatal_set <- aggregate(data$FATALITIES, by=list(data$EVTYPE), sum)
names(fatal_set) <- c("EVTYPE", "FATALITIES")
ordered_fatalities <- fatal_set[order(fatal_set$FATALITIES, decreasing
= TRUE),]
top_fatal <- head(ordered_fatalities, 7)
```

## Injuries on the Weather Event

Only looking at the seven events on injuries pertaining to weather.

```
inj_set <- aggregate(data$INJURIES, by=list(data$EVTYPE), sum)
names(inj_set) <- c("EVTYPE", "INJURIES")
ordered_injuries <- inj_set[order(inj_set$INJURIES, decreasing =
TRUE),]
top_inj <- head(ordered_injuries, 7)
```

## Calulation of Impact to Property

Only looking at the seven events that causes impact to property.

```
dmg_data <- data[, c("EVTYPE", "PROPDMG", "CROPDMG")]
dmg_data$total_dmg <- dmg_data$PROPDMG + dmg_data$CROPDMG
event_dmg_data <- aggregate(dmg_data$total_dmg,
by=list(dmg_data$EVTYPE), FUN=sum)
names(event_dmg_data) <- c("EVTYPE", "TOTAL_DMG")
ordered_event_dmg <- event_dmg_data[order(event_dmg_data$TOTAL_DMG,
decreasing = TRUE),]
top_5_dmg <- head(ordered_event_dmg, 7)
```

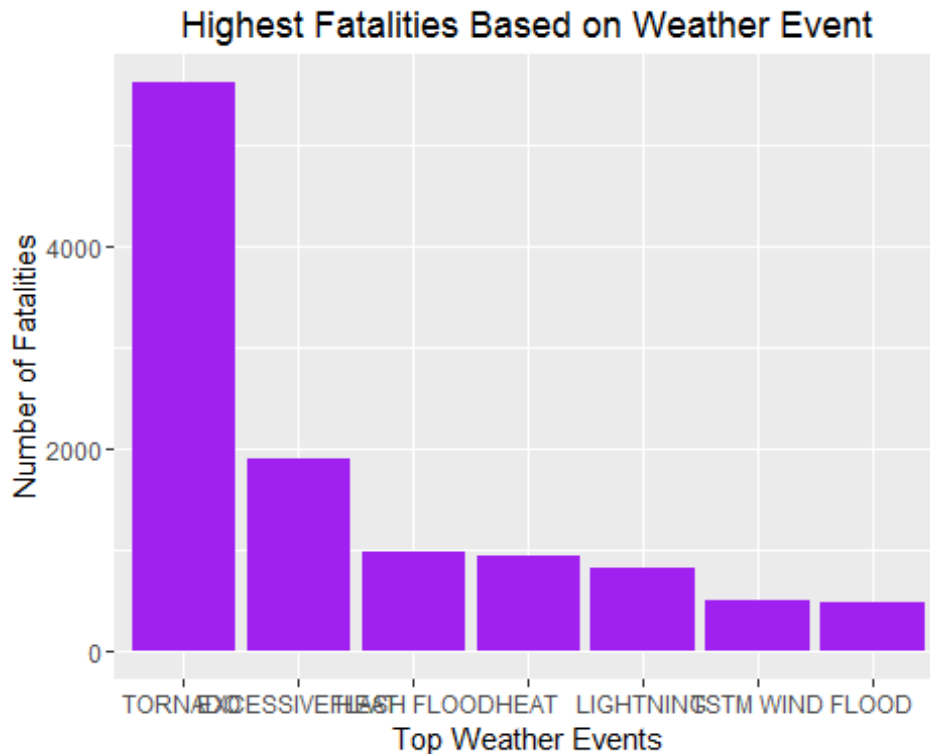
## Results / Summary

Load Necessary Library For Plotting Results

```
library(ggplot2)
```

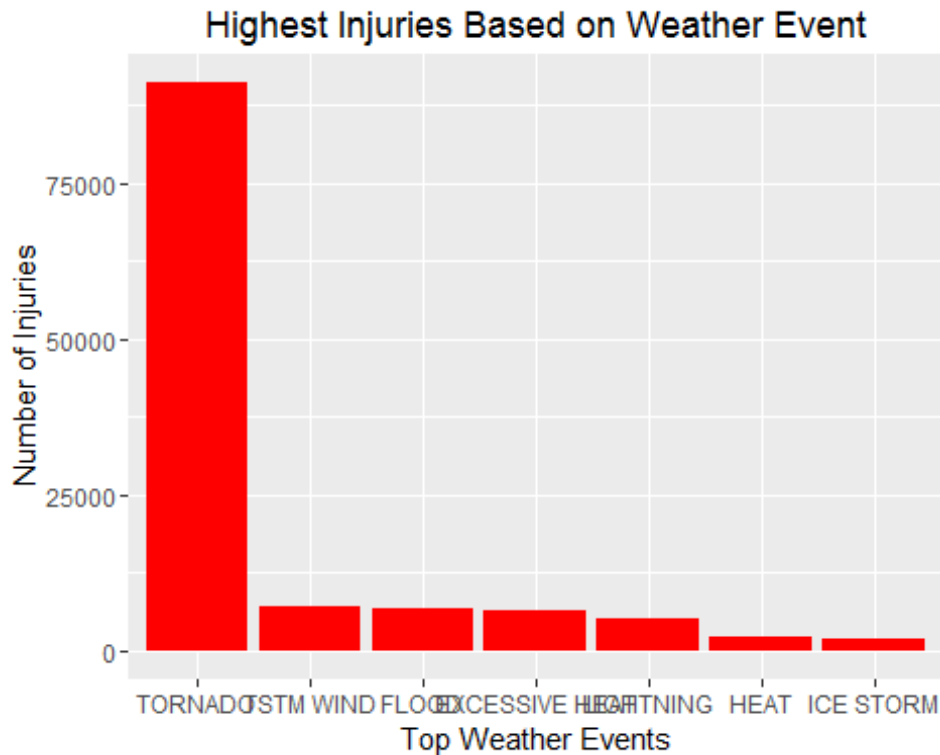
### Events that Impact the People

```
top_fatal$EVTYPE <- ordered(top_fatal$EVTYPE,  
levels=levels(top_fatal$EVTYPE)[unclass(top_fatal$EVTYPE)])  
ggplot(top_fatal, aes(x=EVTYPE,y=FATALITIES)) +  
geom_bar(fill="purple",stat="identity") + xlab("Top Weather Events") +  
ylab("Number of Fatalities") + ggtitle("Highest Fatalities Based on  
Weather Event")
```



According to the chart, tornado is the weather event that has the highest fatality.

```
top_inj$EVTYPE <- ordered(top_inj$EVTYPE, levels=levels(top_inj$EVTYPE)  
[unclass(top_inj$EVTYPE)])  
ggplot(top_inj, aes(x=EVTYPE,y=INJURIES)) +  
geom_bar(fill="red",stat="identity") + xlab("Top Weather Events") +  
ylab("Number of Injuries") + ggtitle("Highest Injuries Based on Weather  
Event")
```

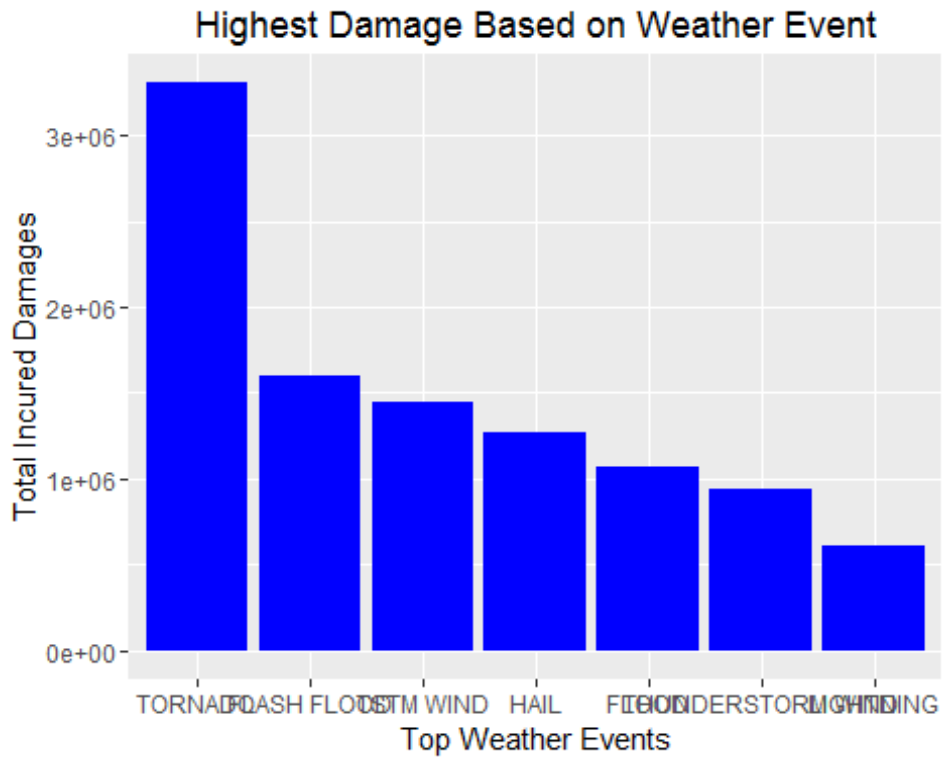


According to the chart, tornado again leads the high amount of injuries as of the weather event.

Therefore, on the three data sets that are provided below, it has determined that tornadoes are the leading cause for all damages by far. The second weather effects with excessive heat, thunderstorm wind, or flash floods are not coming close on causing the most damages.

#### Event That Impact Property

```
top_5_dmg$EVTYPE <- ordered(top_5_dmg$EVTYPE,
levels=levels(top_5_dmg$EVTYPE)[unclass(top_5_dmg$EVTYPE)])
ggplot(top_5_dmg, aes(x=EVTYPE,y=TOTAL_DMG)) +
geom_bar(fill="blue",stat="identity")+ xlab("Top Weather Events") +
ylab("Total Incured Damages") + ggtitle("Highest Damage Based on
Weather Event")
```



With the chart's analysis, it has determined that tornado has caused the most damages to the properties and crops around the country.