RmarkdownAssign2

Weather events in United States and it impact.

Synopsis

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.

The objective of the study: To explore the NOAA Storm Database and answer some basic questions about severe weather events.

```
library(dplyr)
library(ggplot2)
library(data.table)
library(ggpubr)

setwd("C:/Users/sgpoh/Documents")
```

Read-in csv dataset from link

```
storm.url =
"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
storm.bz2 = "storm-data.csv.bz2"
if (! file.exists(storm.bz2)) {
    download.file(url=storm.url, destfile=storm.bz2)
}
storm_data <- read.csv(storm.bz2)

require(data.table)

## Loading required package: data.table

storm_dataDT <- as.data.table(storm_data)
```

Subsetting data into only variables required for analysis

```
cols2Remove <- colnames(storm_dataDT[, !c("EVTYPE"
  , "FATALITIES"
  , "INJURIES"
  , "PROPDMG"
  , "PROPDMGEXP"
  , "CROPDMG"
  , "CROPDMGEXP")])

storm_dataDT[, c(cols2Remove) := NULL]

storm_dataDT <- storm_dataDT[(EVTYPE != "?" &
             (INJURIES > 0 | FATALITIES > 0 | PROPDMG > 0 | CROPDMG > 0)),
c("EVTYPE"
                                                                            ,
"FATALITIES"
                                                                            ,
"INJURIES"
                                                                            ,
"PROPDMG"
                                                                            ,
"PROPDMGEXP"
                                                                            ,
"CROPDMG"
                                                                            ,
"CROPDMGEXP") ]
```

Ensure consistency of values in variables

```
storm_dataDT$EVTYPE2 <- trimws(storm_dataDT$EVTYPE)
storm_dataDT$EVTYPE2 <- toupper(storm_dataDT$EVTYPE2)
storm_dataDT$PROPDMGEXP2 <- trimws(storm_dataDT$PROPDMGEXP)
storm_dataDT$PROPDMGEXP2 <- toupper(storm_dataDT$PROPDMGEXP2)
storm_dataDT$CROPDMGEXP2 <- trimws(storm_dataDT$CROPDMGEXP)
storm_dataDT$CROPDMGEXP2 <- toupper(storm_dataDT$CROPDMGEXP2)

storm_dataDT$EVTYPE2 <- gsub("( ){1,}", " ", gsub("[^A-Z0-9 ]", " ",
storm_dataDT$EVTYPE2))

# Recognise flood events
storm_dataDT$EVTYPE2[grepl("COASTAL|STORM SURGE", storm_dataDT$EVTYPE2)] <-
"COASTAL FLOOD"
storm_dataDT$EVTYPE2[grepl("FLASH", storm_dataDT$EVTYPE2)] <- "FLASH FLOOD"
storm_dataDT$EVTYPE2[!grepl("FLASH|COASTAL", storm_dataDT$EVTYPE2) &
grepl("FLOOD", storm_dataDT$EVTYPE2)] <- "FLOOD"
storm_dataDT$EVTYPE2[grepl("STREAM|URBAN", storm_dataDT$EVTYPE2)] <- "FLOOD"

# Recognise heat events
storm_dataDT$EVTYPE2[grepl("HEAT|DRY", storm_dataDT$EVTYPE2)] <- "EXCESSIVE
HEAT"
storm_dataDT$EVTYPE2[grepl("HOT|WARM", storm_dataDT$EVTYPE2)] <- "EXCESSIVE
HEAT"
storm_dataDT$EVTYPE2[grepl("RECORD (HIGH|.*TEMP)|HIGH TEMPERA",
storm_dataDT$EVTYPE2)] <- "EXCESSIVE HEAT"

# Recognise cold events
storm_dataDT$EVTYPE2[grepl("SLEET", storm_dataDT$EVTYPE2)] <- "SLEET"
```

```
storm_dataDT$EVTYPE2[grepl("BLIZZARD", storm_dataDT$EVTYPE2)] <- "BLIZZARD"
storm_dataDT$EVTYPE2[grepl("EXTREME", storm_dataDT$EVTYPE2) &
grepl("CHILL|COLD", storm_dataDT$EVTYPE2)] <- "EXTREME COLD/WIND CHILL"
storm_dataDT$EVTYPE2[!grepl("EXTREME", storm_dataDT$EVTYPE2) &
grepl("CHILL|COLD", storm_dataDT$EVTYPE2)] <- "COLD/WIND CHILL"
storm_dataDT$EVTYPE2[grepl("LAKE", storm_dataDT$EVTYPE2) & grepl("SNOW",
storm_dataDT$EVTYPE2)] <- "LAKE-EFFECT SNOW"
storm_dataDT$EVTYPE2[!grepl("LAKE", storm_dataDT$EVTYPE2) & grepl("SNOW",
storm_dataDT$EVTYPE2)] <- "HEAVY SNOW"
storm_dataDT$EVTYPE2[grepl("FROST|FREEZE", storm_dataDT$EVTYPE2)] <-
"FROST/FREEZE"
storm_dataDT$EVTYPE2[!grepl("FROST", storm_dataDT$EVTYPE2) & grepl("FREEZE",
storm_dataDT$EVTYPE2)] <- "SLEET"
storm_dataDT$EVTYPE2[grepl("FREEZ", storm_dataDT$EVTYPE2) & grepl("RAIN",
storm_dataDT$EVTYPE2)] <- "SLEET"
storm_dataDT$EVTYPE2[grepl("DRIZZLE", storm_dataDT$EVTYPE2)] <- "SLEET"
storm_dataDT$EVTYPE2[grepl("(RECORD LOW|LOW TEMP)", storm_dataDT$EVTYPE2)] <-
"EXTREME COLD/WIND CHILL"
storm_dataDT$EVTYPE2[grepl("GLAZE", storm_dataDT$EVTYPE2)] <- "EXTREME
COLD/WIND CHILL"
storm_dataDT$EVTYPE2[grepl("ICE", storm_dataDT$EVTYPE2)] <- "ICE STORM"
storm_dataDT$EVTYPE2[grepl("WINT", storm_dataDT$EVTYPE2)] <- "WINTER STORM"
storm_dataDT$EVTYPE2[grepl("HAIL", storm_dataDT$EVTYPE2)] <- "HAIL"

# Recognise and collect lightning, wind and rain events
storm_dataDT$EVTYPE2<- gsub("WINDS", "WIND", storm_dataDT$EVTYPE2)
storm_dataDT$EVTYPE2[!grepl("DERSTORM WIND", storm_dataDT$EVTYPE2) &
grepl("THUN|TSTM", storm_dataDT$EVTYPE2)] <- "LIGHTNING"
storm_dataDT$EVTYPE2[grepl("LIGHT|LIGN", storm_dataDT$EVTYPE2)] <-
"LIGHTNING"
storm_dataDT$EVTYPE2[grepl("DERSTORM WIND", storm_dataDT$EVTYPE2)] <-
"THUNDERSTORM WIND"
storm_dataDT$EVTYPE2[grepl("TORN", storm_dataDT$EVTYPE2)] <- "TORNADO"
storm_dataDT$EVTYPE2[grepl("SPOUT", storm_dataDT$EVTYPE2)] <- "WATERSPOUT"
storm_dataDT$EVTYPE2[grepl("HURRICANE|TYPHOON", storm_dataDT$EVTYPE2)] <-
"HURRICANE (TYPHOON)"
storm_dataDT$EVTYPE2[grepl("FIRE", storm_dataDT$EVTYPE2)] <- "WILDFIRE"
storm_dataDT$EVTYPE2[!grepl("MARINE", storm_dataDT$EVTYPE2) & grepl("HIGH
WIND", storm_dataDT$EVTYPE2)] <- "HIGH WIND"
storm_dataDT$EVTYPE2[grepl("GUST", storm_dataDT$EVTYPE2)] <- "STRONG WIND"
storm_dataDT$EVTYPE2[!grepl("COLD|MARINE|THUNDER|STRONG|HIGH",
storm_dataDT$EVTYPE2) & grepl("WIND", storm_dataDT$EVTYPE2)] <- "STRONG WIND"
storm_dataDT$EVTYPE2[grepl("FUNNEL", storm_dataDT$EVTYPE2)] <- "FUNNEL CLOUD"
storm_dataDT$EVTYPE2[grepl("TROPICAL STORM", storm_dataDT$EVTYPE2)] <-
"TROPICAL STORM"
storm_dataDT$EVTYPE2[!grepl("FREEZIN", storm_dataDT$EVTYPE2) &
grepl("FOG|VOG", storm_dataDT$EVTYPE2)] <- "DENSE FOG"
storm_dataDT$EVTYPE2[grepl("WET|RAIN|SHOWER|PRECIP", storm_dataDT$EVTYPE2)]
<- "HEAVY RAIN"
```

```
# Find dust events
storm_dataDT$EVTYPE2[grepl("DUST DEVEL", storm_dataDT$EVTYPE2)] <- "DUST
DEVIL"
storm_dataDT$EVTYPE2[!grepl("DEVIL", storm_dataDT$EVTYPE2) & grepl("DUST",
storm_dataDT$EVTYPE2)] <- "DUST STORM"

# All on-sea events (marine)
storm_dataDT$EVTYPE2[grepl("RIP CURRENT", storm_dataDT$EVTYPE2)] <- "RIP
CURRENT"
storm_dataDT$EVTYPE2[!grepl("LOW", storm_dataDT$EVTYPE2) &
grepl("TIDE|WAVE|SWELL", storm_dataDT$EVTYPE2)] <- "STORM SURGE/TIDE"
storm_dataDT$EVTYPE2[grepl("SURF", storm_dataDT$EVTYPE2)] <- "HIGH SURF"

# A few odd, miscellaneuosu events
storm_dataDT$EVTYPE2[grepl("VOLCAN", storm_dataDT$EVTYPE2)] <- "VOLCANIC ASH"

# Hmm, sort of landslides so reclassify
storm_dataDT$EVTYPE2[grepl("(MUD|LAND|ROCK).*SLIDE", storm_dataDT$EVTYPE2)]
<- "LANDSLIDE"
```

Convert exponential columns into numerical format

```
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="\"\""] <- 10^0
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="-"] <- 10^0
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="+"] <- 10^0
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="0"] <- 10^0
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="1"] <- 10^1
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="2"] <- 10^2
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="3"] <- 10^3
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="4"] <- 10^4
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="5"] <- 10^5
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="6"] <- 10^6
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="7"] <- 10^7
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="8"] <- 10^8
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="9"] <- 10^9
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="H"] <- 10^2
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="K"] <- 10^3
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="M"] <- 10^6
storm_dataDT$PROPDMGEXP2[storm_dataDT$PROPDMGEXP2=="B"] <- 10^8

storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="\"\""] <- 10^0
storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="?"] <- 10^0
storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="0"] <- 10^0
storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="K"] <- 10^3
storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="M"] <- 10^6
storm_dataDT$CROPDMGEXP2[storm_dataDT$CROPDMGEXP2=="B"] <- 10^9
```

Calculate Economic Cost Columns

```
storm_dataDT$PROPDMG = as.numeric(storm_dataDT$PROPDMG)
storm_dataDT$PROPDMGEXP2 = as.numeric(storm_dataDT$PROPDMGEXP2)
```

```
storm_dataDT$CROPDMG = as.numeric(storm_dataDT$CROPDMG)
storm_dataDT$CROPDMGEXP2 = as.numeric(storm_dataDT$CROPDMGEXP2)
storm_dataDT$propCost = ifelse(storm_dataDT$PROPDMG!=0,storm_dataDT$PROPDMG *
storm_dataDT$PROPDMGEXP2,0)
storm_dataDT$cropCost = ifelse(storm_dataDT$CROPDMG!=0,storm_dataDT$CROPDMG *
storm_dataDT$CROPDMGEXP2,0)
```

Calcuating Total Property and Crop Cost

```
storm_dataDT$totalCost <- storm_dataDT$propCost + storm_dataDT$cropCost
TotalCostDT = aggregate(totalCost~EVTYPE2, storm_dataDT, sum)
MeanCostDT = aggregate(totalCost~EVTYPE2, storm_dataDT, mean)
TotalCostDT2 = TotalCostDT[order(-TotalCostDT$totalCost),]
MeanCostDT2 = MeanCostDT[order(-MeanCostDT$totalCost),]

#Top10 Total Property and Crop Cost
TotalCostDT2 <- TotalCostDT2[1:10, ]
#Top10 Mean Property and Crop Cost
MeanCostDT2 <- MeanCostDT2[1:10, ]
```

Calculate Total Fatalities and Injuries

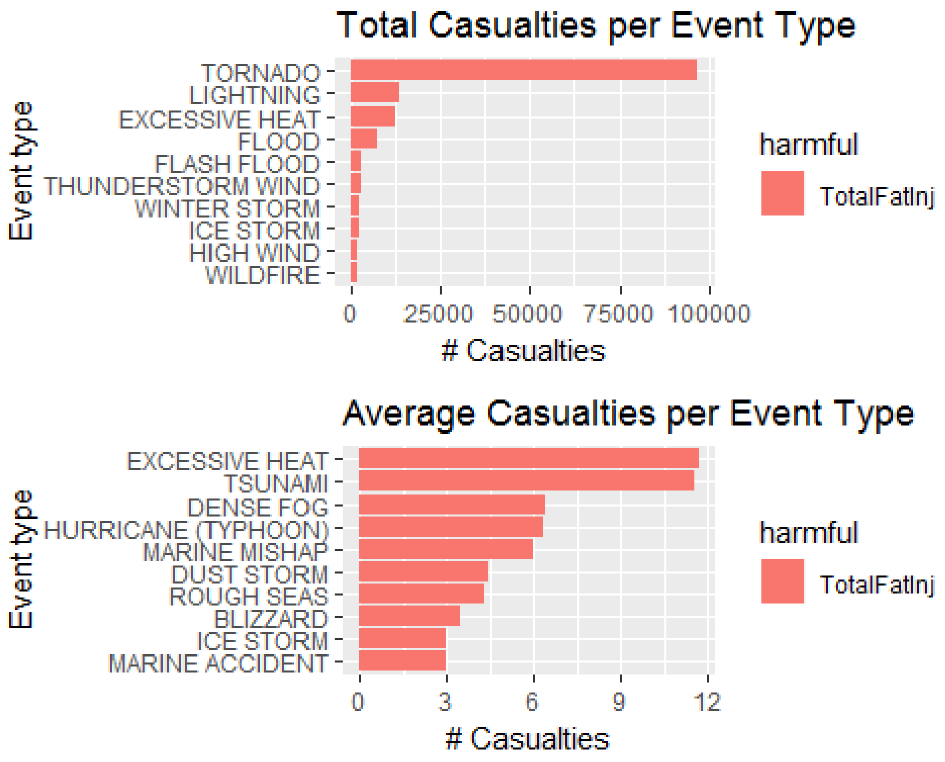
```
storm_dataDT$TotalFatInj = storm_dataDT$FATALITIES + storm_dataDT$INJURIES
TotalFatInjDT = aggregate(TotalFatInj~EVTYPE2, storm_dataDT, sum)
MeanFatInjDT = aggregate(TotalFatInj~EVTYPE2, storm_dataDT, mean)
TotalFatInjDT2 = TotalFatInjDT[order(-TotalFatInjDT$TotalFatInj),]
MeanFatInjDT2 = MeanFatInjDT[order(-MeanFatInjDT$TotalFatInj),]

#Top10 Total Fatalities and Injuries
TotalFatInjDT2 <- TotalFatInjDT2[1:10, ]
#Top10 Mean Fatalities and Injuries
MeanFatInjDT2 <- MeanFatInjDT2[1:10, ]
```

Results

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

```
library(reshape2)

##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##     dcast, melt

total_harmful_events= melt(TotalFatInjDT2, id.vars="EVTYPE2", variable.name =
"harmful")
mean_harmful_events= melt(MeanFatInjDT2, id.vars="EVTYPE2", variable.name =
"harmful")
```

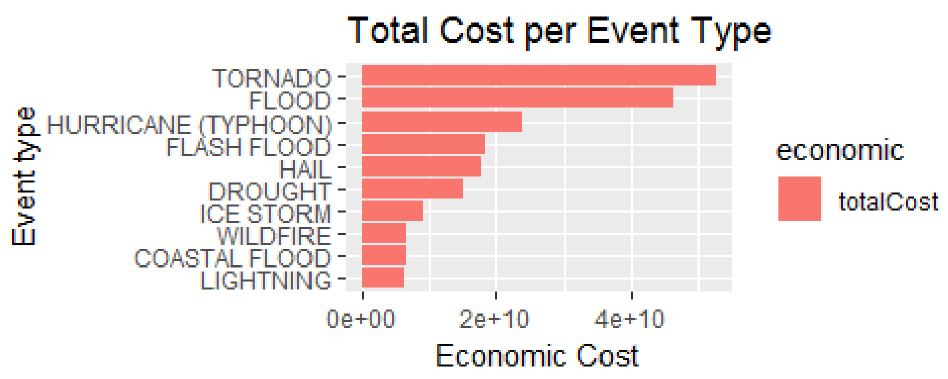
```
require(ggplot2)

## Loading required package: ggplot2

Total_Harmful = ggplot(total_harmful_events, aes(x=reorder(EVTYPE2,value),
y=value, fill=harmful)) + geom_bar(stat="identity") + coord_flip() +
ggtitle("Total Casualties per Event Type") + xlab("Event type") + ylab("#
Casualties")

Mean_Harmful = ggplot(mean_harmful_events, aes(x=reorder(EVTYPE2,value),
y=value, fill=harmful)) + geom_bar(stat="identity") + coord_flip() +
ggtitle("Average Casualties per Event Type") + xlab("Event type") + ylab("#
Casualties")

require(ggpubr)

## Loading required package: ggpubr

## Loading required package: magrittr

figure <- ggarrange(Total_Harmful,Mean_Harmful, ncol = 1, nrow = 2)
figure
```

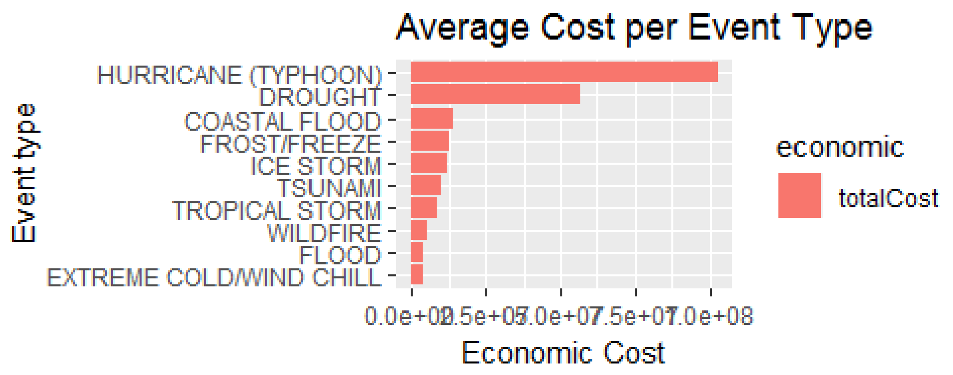


Tornado have the highest impact on health in terms of absolute numbers reported, while heat-related events have the highest number of casualties per event.

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

```
library(reshape2)
total_economic_cost= melt(TotalCostDT2, id.vars="EVTYPE2", variable.name =
"economic")
mean_economic_cost= melt(MeanCostDT2, id.vars="EVTYPE2", variable.name =
"economic")

require(ggplot2)
Total_Economic = ggplot(total_economic_cost, aes(x=reorder(EVTYPE2,value),
y=value, fill=economic)) + geom_bar(stat="identity") + coord_flip() +
ggtitle("Total Cost per Event Type") + xlab("Event type") + ylab("Economic
Cost")

Mean_Economic = ggplot(mean_economic_cost, aes(x=reorder(EVTYPE2,value),
y=value, fill=economic)) + geom_bar(stat="identity") + coord_flip() +
ggtitle("Average Cost per Event Type") + xlab("Event type") + ylab("Economic
Cost")

require(ggpubr)
figure <- ggarrange(Total_Economic,Mean_Economic, ncol = 1, nrow = 2)
figure
```





### Wind related events such as tornado and hurricane have the high economic consequences in terms of both absolute numbers reported, as well as the highest economic cost per event.