Health and Economic Impact of Severe Weather Conditions

Abstract/Synopsis

Data released by the National Oceanic and Atmospheric Administration were examined to determine weather conditions that lead to the greatest hazards to human health and sources of economic damage. This report describes the methods used to gather and clean the data set, transform the data, and an analysis of the impacts of various weather conditions. In this analysis, it is demonstrated that tornados have the largest impact to human health in terms of fatalities and injuries, exceeding the damage toll on human health of the next several types of weather conditions combined. Tornado activity led to 97,043 injuries and deaths compared to 155,673 from all sources combined.

Economic impact of weather patterns was assessed on the basis of reported property and crop damage combined. The total reported economic damage reached $49 billion from all sources combined. Flooding had the largest economic impact, followed closely by hurricanes, and tornados. The severity and impact of weather conditions must be assessed on both grounds to mitigate the effects when planning responses as there are substantial differences between the severity of their health and economic impact.

Assignment

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. You must use the database to answer the questions below and show the code for your entire analysis. Your analysis can consist of tables, figures, or other summaries. You may use any R package you want to support your analysis.

Questions

Your data analysis must address the following questions:

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? Across the United States, which types of events have the greatest economic consequences? Consider writing your report as if it were to be read by a government or municipal manager who might be responsible for preparing for severe weather events and will need to prioritize resources for different types of events. However, there is no need to make any specific recommendations in your report.

Requirements

For this assignment you will need some specific tools

- RStudio: You will need RStudio to publish your completed analysis document to RPubs. You can also use RStudio to edit/write your analysis.
- knitr: You will need the knitr package in order to compile your R Markdown document and convert it to HTML

Document Layout

- Language: Your document should be written in English.
- Title: Your document should have a title that briefly summarizes your data analysis
- Synopsis: Immediately after the title, there should be a synopsis which describes and summarizes your analysis in at most 10 complete sentences.
- There should be a section titled Data Processing which describes (in words and code) how the data were loaded into R and processed for analysis. In particular, your analysis must start from the raw CSV file containing the data. You cannot do any preprocessing outside the document. If preprocessing is time-consuming you may consider using the cache = TRUE option for certain code chunks.
- There should be a section titled Results in which your results are presented.
- You may have other sections in your analysis, but Data Processing and Results are required.
- The analysis document must have at least one figure containing a plot.
- Your analysis must have no more than three figures. Figures may have multiple plots in them (i.e. panel plots), but there cannot be more than three figures total.
- You must show all your code for the work in your analysis document. This may make the document a bit verbose, but that is okay. In general, you should ensure that echo = TRUE for every code chunk (this is the default setting in knitr).

Data Processing

Reading in the datset

Weather data were obtained from the following url:
https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
(https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)

```
#Read in file if it is not already saved in the work environment
setwd(getwd())

if(exists("stormData")==FALSE){
  OriginalstormData<-read.csv("repdata%2Fdata%2FStormData.csv.bz2")
  stormData<-OriginalstormData
  }else{
      stormData<-OriginalStormData #Resets original data if we rerun the progra
m
  }
```

Cleaning the Dataset: Weather Categories

An initial look at the dataset using unique(stormData$EVTYPE) reveals numerous inconsistencies in data input for weather condition types. This includes different naming conventions, capitalziation usage, and spelling errors. The first step in the data processing will be an attempt to reduce these errors. In addition, related weather conditions will be further grouped into broader categories. The method used creates broad changes without examining every statment. This has the possibility of miscategorizing some of the terms or simply assigning an event that had multiple conditions listed into a single category based on the order the replacements were processed. There is a tradeoff in accuracy and time spent, and for this initial exploration of data, the faster approach is sufficient.

```
#Make Everything lowercase
stormData$EVTYPE<-tolower(stormData$EVTYPE)

#This section attempts to better group the categories as slight variations in similar weather phenomena drastically change the results
stormData$EVTYPE[grepl("thunderstorm|tstm|lightning",stormData$EVTYPE)]<-"thunderstorm"
stormData$EVTYPE[grepl("tornado",stormData$EVTYPE)]<-"tornado"
stormData$EVTYPE[grepl("flood",stormData$EVTYPE)]<-"flood"
stormData$EVTYPE[grepl("snow",stormData$EVTYPE)]<-"snow"
stormData$EVTYPE[grepl("ice|freeze|frost|sleet|hail|freezing|blizzard|snow|winter storm",stormData$EVTYPE)]<-"ice/snow"
stormData$EVTYPE[grepl("wind",stormData$EVTYPE)]<-"wind"
stormData$EVTYPE[grepl("heat|hot|warm",stormData$EVTYPE)]<-"heat"
stormData$EVTYPE[grepl("cold",stormData$EVTYPE)]<-"cold"
stormData$EVTYPE[grepl("rain|precipitation|wet",stormData$EVTYPE)]<-"rain"
stormData$EVTYPE[grepl("fog",stormData$EVTYPE)]<-"fog"
stormData$EVTYPE[grepl("fire",stormData$EVTYPE)]<-"fire"
stormData$EVTYPE[grepl("current|surf|tides|seas|swells|waterspout|surge|wave",stormData$EVTYPE)]<-"current/tides"
stormData$EVTYPE[grepl("dry|drought",stormData$EVTYPE)]<-"drought"
stormData$EVTYPE[grepl("mud|slide",stormData$EVTYPE)]<-"mud"
stormData$EVTYPE[grepl("hurricane|tropical storm",stormData$EVTYPE)]<-"hurricane"
```

Cleaning the Dataset: Economic Damage Values

The economic impact of various hazards needs to be transformed to make the columns directly comparable. For both the crops and property damage, there is a second column with an factor that is listed in several different formats. Two approaches could be taken here: 1. Sort the data only by the largest exponentiation term 2. Convert all terms to the same format

Option 1 is the fastest approach, but has the potential to miss large items (A billion impact may have been entered with a thousands factor). This could be useful for an intial look at the data, but is not accurate enough. The following section converts all the format types into a consistent numeric output of economic damage that can be directly compared.

```r
###Function to replace arrays of data. This will be used to standardize all of the terms in the exponetial columns.
recoderFunc <- function(data, oldvalue, newvalue) {

  # convert any factors to characters

  if (is.factor(data))     data     <- as.character(data)
  if (is.factor(oldvalue)) oldvalue <- as.character(oldvalue)
  if (is.factor(newvalue)) newvalue <- as.character(newvalue)

  # create the return vector

  newvec <- data

  # put recoded values into the correct position in the return vector

  for (i in unique(oldvalue)) newvec[data == i] <- newvalue[oldvalue == i]

  newvec

}

#Clean up crop and property damage data
#Determine factors for exponential column. Assigning a "1" to exp values that have an unclear values like "?".

#levels(stormData$PROPDMGEXP) Used this function to determine all of the factors in both "exp" datasets
# The next to functions list the exponentiation levels and the factors that they should be converted to.
a<-c("",  "-", "?", "+", "0", "1", "2", "3", "4", "5", "6", "7", "8", "B", 
"h", "H", "K", "m", "M")
b<-c(1,  1, 1, 1, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^2, 10^2, 10^3, 10^6, 10^6)

stormData$PROPDMGEXP2<-stormData$PROPDMGEXP #new variable to store the recoded data
#Replace data based on economic terms
stormData$PROPDMGEXP2<-recoderFunc(stormData$PROPDMGEXP,a,b)
stormData$CROPDMGEXP2<-recoderFunc(stormData$CROPDMGEXP,a,b)

#Multiple prop damage by the factor term to get standardized values
stormData$PROPDMGFixed<-as.numeric(stormData$PROPDMGEXP2)*stormData$PROPDMG
stormData$CROPDMGFixed<-as.numeric(stormData$CROPDMGEXP2)*stormData$CROPDMG
```

```
## Warning: NAs introduced by coercion
```

Results

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

The following section groups the data together by event types. For health hazards, injuries and fatalities are combined. This of course ignores other long term health affects that could be associated with homelessness, food shortages, mold, etc. that may be associated with the events. Economic damage is determined by combining property and crop damage for each weather condition. The Following figures present rankings of the top 10 most harmful conditions in each of the two categories.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
hazards<-aggregate(cbind(stormData$INJURIES+stormData$FATALITIES), by=list(stormData$EVTYPE),FUN=sum, na.rm=TRUE) #groups data based on hazard type and sums personal injures deaths by that weather condition
names(hazards)<-c("EVTYPE","hazardCount")
rankHazards<-hazards[order(hazards$hazardCount, decreasing=TRUE),]
ggplot(data=rankHazards[1:10,], aes(y=hazardCount,x=reorder(EVTYPE,-hazardCount))) +
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  labs(y="Injuries and Fatalities",x="")
```

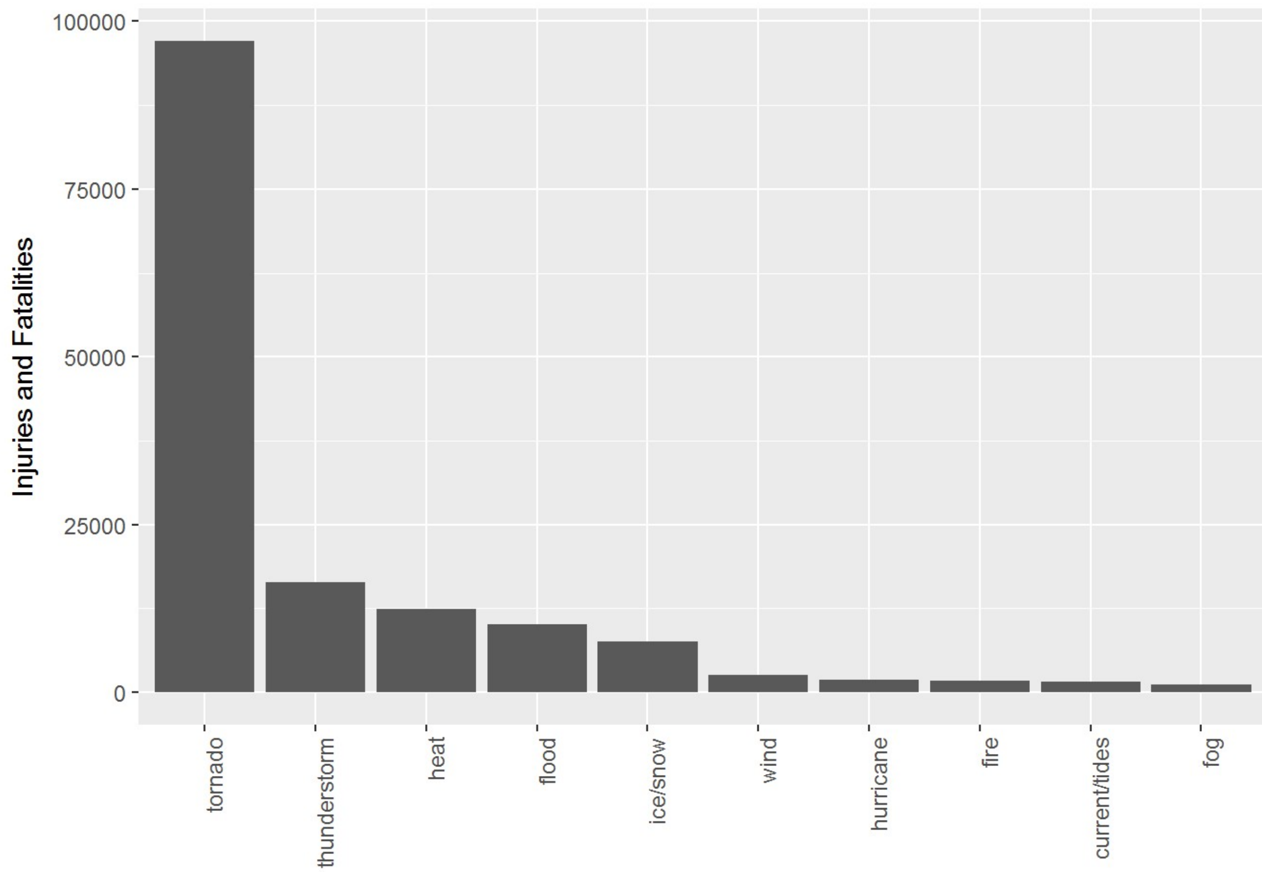

Figure 1: Personal injuries and deaths ranked in descending order by weather pattern type

```
hazardsEcon<-aggregate(cbind(stormData$PROPDMGFixed+stormData$CROPDMGFixed), by
=list(stormData$EVTYPE),FUN=sum, na.rm=TRUE) #groups data based on hazard type
and sums personal injures deaths by that weather condition
names(hazardsEcon)<-c("EVTYPE","hazardCount")
rankHazardsEcon<-hazardsEcon[order(hazardsEcon$hazardCount, decreasing=TRUE),]

ggplot(data=rankHazardsEcon[1:10,], aes(y=hazardCount/10^6,x=reorder(EVTYPE,-ha
zardCount))) +
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  labs(y="Economic Damage (Millions of $)",x="")
```

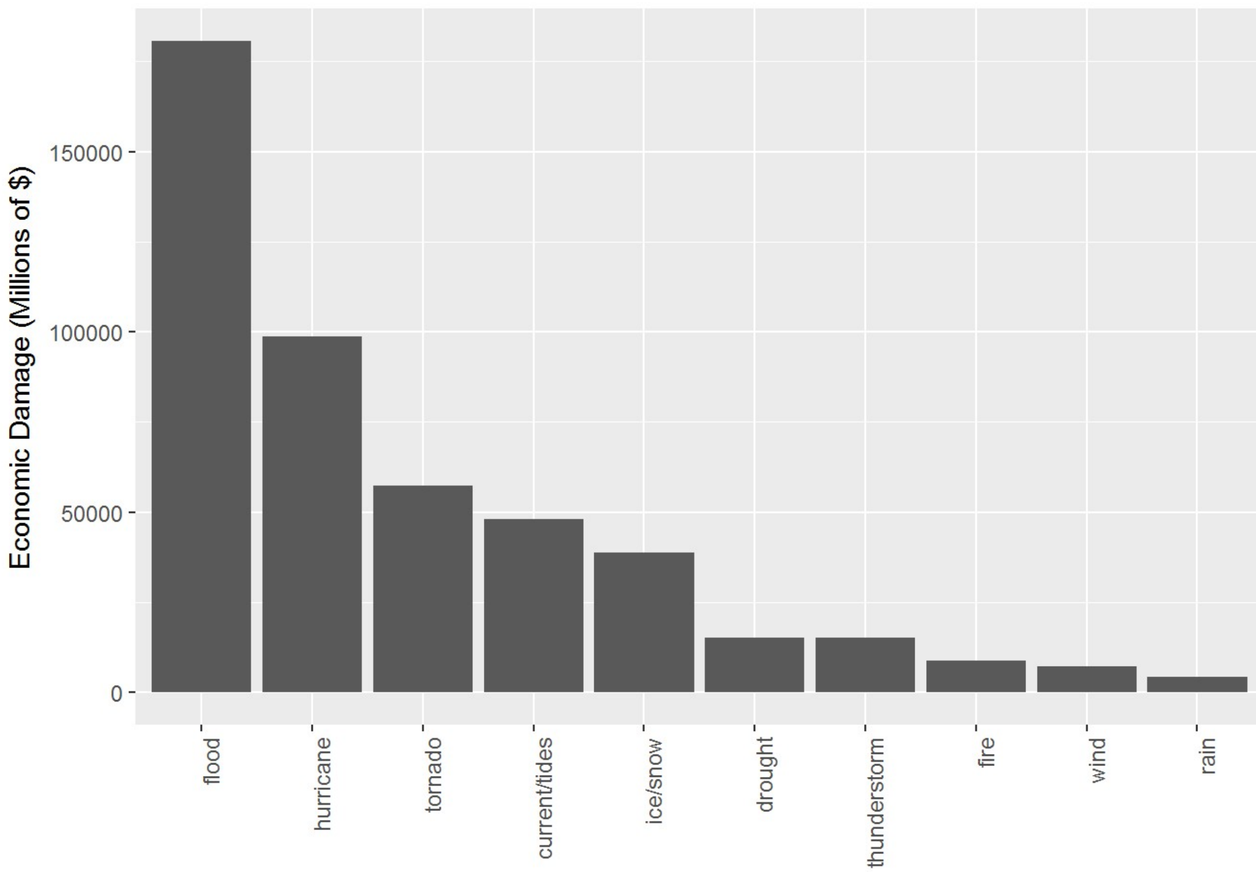

Figure 2: Weather related economic damage in descending order based on weather pattern type

```
#Summarizing some additional data to be used in the abstract
healthHazardCount<-sum(rankHazards$hazardCount)

#Total injuries and loss of life:
healthHazardCount
```

```
## [1] 155673
```

```
EconHazardCount<-sum(rankHazardsEcon$hazardCount)/10^9
#Total Economic Damage in Billions
EconHazardCount
```

```
## [1] 477.3283
```

Conclusions

The health and economic impact of various weather conditions were assessed based on reported data from the National Oceanic & Atmospheric Administration. Various weather conditions can be further grouped into categories to better understand their impact on human life and economic loss. On the basis of this data, tornados caused more deaths and injuries than the next several most hazardous weather conditions combined. Therefore, further attention should be focused on ensuring areas in high risk zones have risk mitigation plans in place, such as warning and alert systems and access to shelters capable of withstanding the storm. In terms of economic impact, flooding presented the highest cause of monetary loss, followed by hurricanes and tornados.