Impacts of Environmental Events from 1991 to 2011

Synopsis

The objective of this report is to draw conclusions about the impact of severe weather events in terms of both humn health and economic damage. To complete this analysis, data is drawn from the National Oceanic and Atmospheric Administration(NOAA) Storm Database.

Specifically, this report will seek to investigate the following questions:

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 have the greatest economic consequences?

The data is originally taken from the National Weather Service who collects the information from sources which include, but are not limited to: county, state and federal emergency management officials, local law enforcement officials, National Weather Service damage surveys, and the general public.

For more detailed information, refer to:

Storm Data DOcumentation

FAQ

This report will attempt to illuminate current trends by narrowing the scope of analysis to a timeframe of twenty years, spanning from 1991 - 2011 (the most recent year of published data in the dataset).

The findings suggest that tornados, excessive heat and flooding were the most harmful to human health during this period, while flooding hurricanes and storm surges caused the most monetary damage to property and crops.

Data Processing

Loading and Processing Raw Data

The following link houses a copy of the NOAA storm data set which is used throughout this analysis.

NOAA Storm Data

Reading in the data

```
storm <- read.csv(bzfile("repdata_data_StormData.csv.bz2"), stringsAsFactors = FALSE)
# bz command extracts the bz file.

# Load necesary packages.
library(dplyr)
library(ggplot2)
library(reshape2)
library(lubridate)
library(stringr)
```

We are interested in the human and economics costs associated with the weather events, namely:

FATALITIES - Number of direct deaths attributed to the weather event.

INJURIES - Number of direct injuries attributed to the weather event.

PROPDMG - Monetary amount of property damage associated with event.

CROPDMG - Monetary damage to crops associated with event.

Cleaning The Data

Removing Missing/Miscoded Values

Therefore, we will remove all instances where no human or economic impacts are recorded.

```
stormFilt <- storm %>%
  filter(FATALITIES != 0 |
           INJURIES != 0 |
           PROPDMG  != 0 |
           CROPDMG  != 0)
```

We are interested in events within the United States, so we will also remove entries where the states are not coded as a US State. This is done will the built in data from the state_abb vector in R.

```
stormFilt$STATE <- toupper(stormFilt$STATE)
stormFilt <- stormFilt[stormFilt$STATE %in% state.abb,]
```

Creating Usable Dates

Next, we will want to convert the date columns into usable formats. I have elected to use the BGN_DATE column to determine time of event. Since we do not need extreme granularity for this analysis, we will just isolate the year from this column and convert to 'date' class.

```
firstElement <- function(x){x[1]} # function for extracting first element of list.
thirdElement <- function(x){x[3]} # function for extracting thrid element of list

splitDates <- str_split(stormFilt$BGN_DATE, "/") # split dates by /
splitDates <- sapply(splitDates, thirdElement) # apply to splitDates

splitDates <- str_split(splitDates, " ")
firstElement <- function(x){x[1]}
splitDates <- sapply(splitDates, firstElement)
stormFilt$Date <- splitDates
```

It is now possible to isolate just entries from 1991 until 2011.

```
stormFilt <- stormFilt[stormFilt$Date >= 1991,]
# we will observe 20 years of data (1991 - 2011). This will show recent trends and make use of more com
```

Ensuring clean, consistent events types

At this point in the cleaning process, there are 471 unique EVTYPEs listed in the dataset. The original documentation list only 48. This difference is due to miscoding, misspelling, and coding multiple EVTYPEs in one entry and other input anomalies. To ensure an accurate analysis, it is critical to consolidate these EVTYPES to ensure that discrete categories are representative and meaningful.

The analysis below does the following: - converts all text to lower case - removes digits and symbols for simplicity - corrects common spelling errors - codes each EVTYPE as only a single event.

In cases where multiple EVTYPEs are listed, they were recoded to the event which was more likely to have caused human injury and monetary damage. In some cases, event types were consolidated when they were similar enough for the purposes of this report.

```
stormFilt$EVTYPE <- tolower(stormFilt$EVTYPE)
# convert all text to lower case.

stormFilt$EVTYPE <- gsub("//d", "", stormFilt$EVTYPE)
# remove digits
stormFilt$EVTYPE <- gsub("/^[a-z ]", "", stormFilt$EVTYPE)
# removes anything that is not a letter or space.

stormFilt$EVTYPE <- gsub("?", "", stormFilt$EVTYPE)
# removes '?'
stormFilt$EVTYPE <- gsub("other", "", stormFilt$EVTYPE) # removes 'other'
stormFilt$EVTYPE <- gsub("apache county", "", stormFilt$EVTYPE) # remove 'apache county'

# the following commands seek out particular events
# which are explained the original documentation,
# then renames the event so that no extraneous text
# remains. In this way some multiple events are reduced
# to a single event. For example, 'blizzard/snow' is
# reduced to 'blizzard' (the component which was likely
# to have caused harm and damage).
stormFilt$EVTYPE <- gsub(".*(astronomical).*", "astronomical low tide", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(avalanche).*", "avalanche", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(blizzard).*", "blizzard", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(coastal).*", "coastal flood", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(cold/wind).*", "cold/wind chill", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(cold/wind).*", "cold/wind chill", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(fog).*", "fog", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(heat).*", "excessive heat", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(hyperthermia).*", "extreme heat", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(fire).*", "wildfire", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(cold).*", "extreme cold", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(frost).*", "extreme cold", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(winter).*", "extreme cold", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(flash flood).*", "flash flood", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(flood).*", "flood", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(fld).*", "flood", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(freez).*", "ice", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(landslide).*", "landslide", stormFilt$EVTYPE)
```

```
stormFilt$EVTYPE <- gsub(".*(hail).*", "hail", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(ice).*", "ice", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(icy).*", "ice", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(hurricane).*", "hurricane", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(tropical storm).*", "tropical storm", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(storm surge).*", "storm surge", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(rain).*", "heavy rain", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(wind).*", "wind", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(snow).*", "snow", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(lightning).*", "lightning", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(thunderstorm).*", "thunderstorm wind", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(tstm).*", "thunderstorm wind", stormFilt$EVTYPE)

stormFilt$EVTYPE <- gsub(".*(tornado).*", "tornado", stormFilt$EVTYPE)
stormFilt$EVTYPE <- gsub(".*(torndao).*", "tornado", stormFilt$EVTYPE)

stormFilt$EVTYPE <- str_trim(stormFilt$EVTYPE) # remove trailing and leading spaces
```

This brings the total unique EVTYPEs to 89. We will consider this sufficientlty clean to continue with analysis since any other anaomolies should be small enough that they will not appear significantly in final tabulation.

Cleaning Monetary Damage Columns

The raw dataset uses a letter coding system to denote mtrhe magnitude of dollar figures. This coding system has not been strictly adhered to, but some of the majors categories include:

- h = hundred
- k = thousand
- m = million
- b = billion

Therefore, we will convert these to numeric values and multiply them by corresponding damage estimates. For entries which use small number instead of coded letters (such as 2,3,4), will we assume that they were intended to act as some type of multiplier. This is not specified, but it should have minimal bearing on final results.

```
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "h"] <- 100
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "H"] <- 100
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "k"] <- 1000
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "K"] <- 1000
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "m"] <- 1000000
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "M"] <- 1000000
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "B"] <- 1000000000
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP ==  ""] <- 1
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "-"] <- 1
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "+"] <- 1
stormFilt$PROPDMGEXP[stormFilt$PROPDMGEXP == "0"] <- 1
```

```r
stormFilt$PROPDMGEXP <- as.numeric(stormFilt$PROPDMGEXP) # convert to numeric
stormFilt$PROPDMG <- as.numeric(stormFilt$PROPDMG)       # convert to numeric

stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "h"] <- 100
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "H"] <- 100
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "k"] <- 1000
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "K"] <- 1000
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "m"] <- 1000000
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "M"] <- 1000000
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "B"] <- 1000000000
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP ==  ""] <- 1
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "-"] <- 1
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "+"] <- 1
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "0"] <- 1
stormFilt$CROPDMGEXP[stormFilt$CROPDMGEXP == "?"] <- 1
stormFilt$CROPDMGEXP <- as.numeric(stormFilt$CROPDMGEXP) # convert to numeric
stormFilt$CROPDMG <- as.numeric(stormFilt$CROPDMG)       # convert to numeric
```

Results

Question 1 - Human Harm

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

To answer this questions, we will define 'harmful to human health' as the total number of injuries and fatalities caused by a given event.

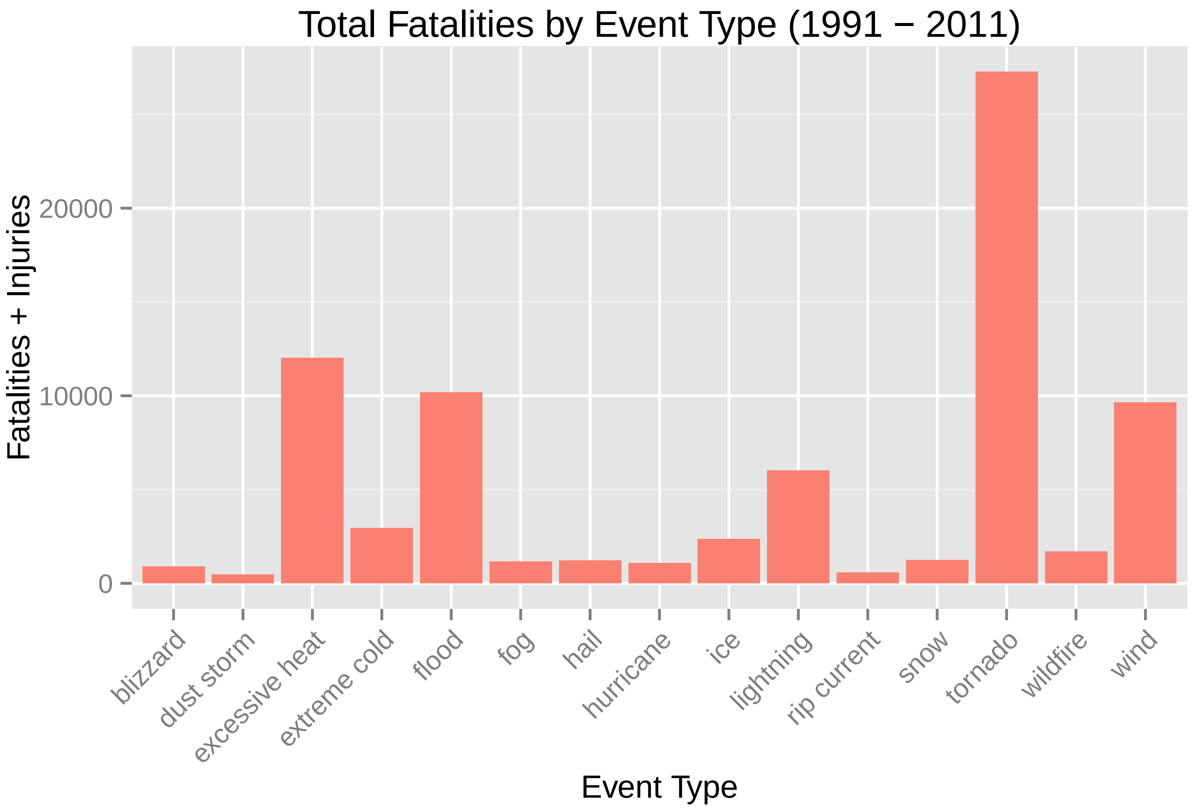
```r
# manipulating the dataframe to show total health cost by EVTYPE, arranged in descending order.
healthImpact <- stormFilt %>%
  group_by(EVTYPE) %>%
  mutate(healthCost = FATALITIES + INJURIES) %>%
  summarize(totalHealthCost = sum(healthCost),
            injuries = sum(INJURIES),
            fatalities = sum(FATALITIES)) %>%
  arrange(desc(totalHealthCost))

Top15 <- healthImpact[1:15, ] # Filtering for the top 15 most harmful event types.
```

To illustrate the findings, we will use a barplot to chart total injuries and fatalities from each type of event.

This information can be leveraged to create new policies and initiatives aimed as protecting citizens from particularly harmful event types.

```r
g <- ggplot(Top15, aes(x = EVTYPE, y = fatalities + injuries))
g <- g + geom_bar(stat="identity", fill = "salmon") +
  xlab("Event Type") +
  ylab("Fatalities + Injuries") +
  ggtitle("Total Fatalities by Event Type (1991 - 2011)")
g + theme(axis.text.x =
              element_text(size  = 10,
                           angle = 45,
                           hjust = 1,
                           vjust = 1))
```



As we can see, tornados proved to be the most costly to human health during the twenty year period analysed. Excessive heat, flooding and wind were the next three most damaging, though they are associated with significantly smaller figures than tornados.

Question 2 - Economic Consequences

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

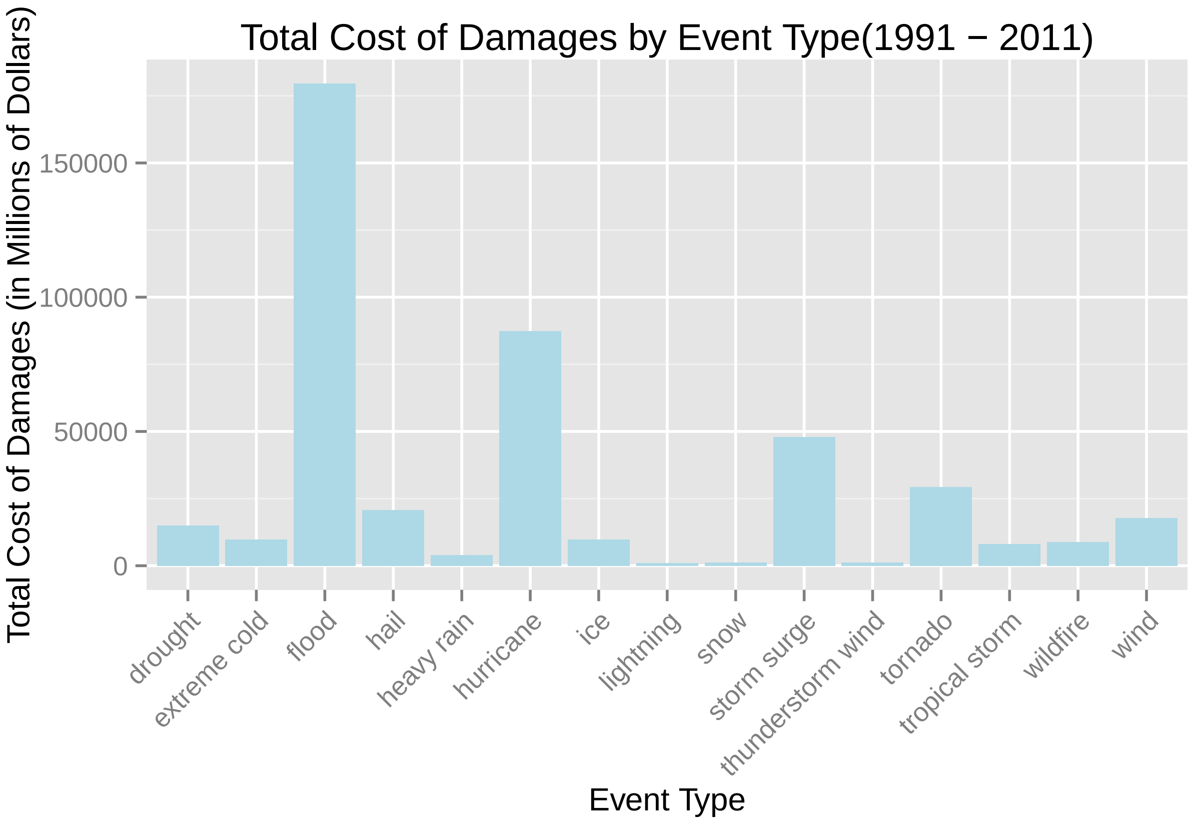
```
# manipulating the dataframe to show total dollar cost cost by EVTYPE, arranged in
# descending order.

dollarImpact <- stormFilt %>%
  group_by(EVTYPE) %>%
  mutate(dollarCost = (PROPDMG*PROPDMGEXP + CROPDMG*CROPDMGEXP)) %>%
  summarize(totalDollarCost = sum(dollarCost),
            propdamage = sum(PROPDMG*PROPDMGEXP),
            cropdamage = sum(CROPDMG*CROPDMGEXP)) %>%
  arrange(desc(totalDollarCost))

Top15Dollar <- dollarImpact[1:15, ] # Filtering for the top 15 most harmful event types.
```

To illustrate the findings, we will use a barplot to chart total cost of damages associated with each type of event.

```
g <- ggplot(Top15Dollar, aes(x = EVTYPE, y = (propdamage + cropdamage)/1000000))
g <- g + geom_bar(stat="identity", fill = "light blue") +
  xlab("Event Type") +
  ylab("Total Cost of Damages (in Millions of Dollars)") +
  ggtitle("Total Cost of Damages by Event Type(1991 - 2011)")
g + theme(axis.text.x =
               element_text(size  = 10,
                            angle = 45,
                            hjust = 1,
                            vjust = 1))
```



With respect to monetary damage to property, it appears that flooding causes the most extensive damage to property and crops, trailed by hurricanes and storm surges (which also amounts to flooding).

Understanding the monetary costs associated with such disasters can provide a guideline for the amount of national budgets which can and should be allocated to mitigating the effects of such events.