

Health and Economic Impact of Weather Events in the U.S.

JB

August, 2017

Synopsis

This document describes an analysis of weather event data taken from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. The data provides details about weather events that have taken place all over the U.S.A. from the years 1950 to 2011. This analysis will focus specifically on the impact of different weather events on population health and economic factors. The population health impact will be evaluated by determining the total number of fatalities and injuries from different types of events over the period of 1950 to 2011. The economic impact of different events will be evaluated by the amount of property and crop damage (in billions of U.S. dollars) caused by different event types over the same time period. The data will be processed and analyzed and the results will be presented using bar plots.

Data Processing

This section outlines steps taken to load and process the data, in order to generate the population health and economic impact data for different types of weather events.

Load the libraries needed for this work (ggplot2, dplyr and gridExtra) and also load the raw weather data into a data frame from a zipped .csv file.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.4.1
```

```
##
```

```
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      combine
```

```
storm_data <- read.csv("repdata_data_StormData.csv.bz2")
```

Convert the fatalities and injuries columns to numeric values. These were originally factors upon loading the data.

```
storm_data$FATALITIES <- as.numeric(as.character(storm_data$FATALITIES))
storm_data$INJURIES <- as.numeric(as.character(storm_data$INJURIES))
```

Create a smaller data frame containing only the event types, fatalities, injuries, crop damage, property damage and exponent columns. Only these columns are needed for the data analysis in question.

```
storm_economic_health_data <- select(storm_data, EVTYPE, FATALITIES, INJURIES,
                                     PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

There are two exponent columns in the data frame for property damage and crop damage. These exponents represent different powers of 10, which can be used to calculate a final dollar value for the total damage. K,k = kilo (or 1000), M,m = millions (10^6) and B = billions (10^9), H,h = hundreds (100). Digits 0 through 8 represent 10 to the power of the digit. Blanks, question marks assign a value of zero to the row in which they are found, since no property damage is present for these entries. Plus and minus signs are also assigned a value of zero, since they don't represent a valid exponent. Essentially, these erroneous data points should be skipped. This code shows the different exponent levels that were found and then replaces them with their corresponding number values. The exponent columns are converted to numeric values and blanks (or NAs in this case) are replaced with ones.

```
levels(storm_economic_health_data$PROPDMGEXP)
```

```
## [1] "" "- " "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8" "B" "h" "H" "K"
## [18] "m" "M"
```

```
storm_economic_health_data$PROPDMGEXP <- recode(storm_economic_health_data$PROPDMGEXP,
'K' = '1000', 'm' = '1000000', 'M' = '1000000', 'B' = '1000000000', '?' = '0',
'-' = '0', '+' = '0', 'H' = '100', 'h' = '100', '0' = '1', '1' = '10', '2' = '100',
'3' = '1000', '4' = '10000', '5' = '100000', '6' = '1000000', '7' = '10000000', '8' = '100000000')
```

```
storm_economic_health_data$PROPDMGEXP <-
as.numeric(as.character(storm_economic_health_data$PROPDMGEXP))
storm_economic_health_data$PROPDMGEXP[is.na(storm_economic_health_data$PROPDMGEXP)] <- 1
```

```
levels(storm_economic_health_data$CROPDMGEXP)
```

```
## [1] "" "?" "0" "2" "B" "k" "K" "m" "M"
```

```
storm_economic_health_data$CROPDMGEXP <- recode(storm_economic_health_data$CROPDMGEXP,
'K' = '1000', 'k' = '1000', 'm' = '1000000', 'M' = '1000000', 'B' = '1000000000',
'?' = '0', '0' = '1', '2' = '100')
```

```
storm_economic_health_data$CROPDMGEXP <-
as.numeric(as.character(storm_economic_health_data$CROPDMGEXP))
storm_economic_health_data$CROPDMGEXP[is.na(storm_economic_health_data$CROPDMGEXP)] <- 1
```

Now add two extra columns to the data frame showing the total dollar value of the crop and property damage for each event. These columns are obtained by multiplying the crop and property damage number columns with their corresponding, newly converted exponent columns.

```
storm_economic_health_data <- mutate(storm_economic_health_data,
PROPDMGDOLLARS = PROPDMG*PROPDMGEXP, CROPDMGDOLLARS = CROPDMG*CROPDMGEXP)
```

Aggregate the total sum of the number of fatalities and injuries, and the total \$US billion value of crop and property damage, for each type of event. Arrange the data in descending order after aggregation. After this step, it will then be feasible to plot the data.

```

fatalities_data <-
aggregate(storm_economic_health_data$FATALITIES~storm_economic_health_data$EVTYPE,
storm_economic_health_data,sum)
colnames(fatalities_data) <-c("EVTYPE", "FATALITIES")
fatalities_data <- arrange(fatalities_data,desc(FATALITIES))

injuries_data <-
aggregate(storm_economic_health_data$INJURIES~storm_economic_health_data$EVTYPE,
storm_economic_health_data,sum)
colnames(injuries_data) <-c("EVTYPE", "INJURIES")
injuries_data <- arrange(injuries_data,desc(INJURIES))

properties_data <-
aggregate(storm_economic_health_data$PROPDMDOLLARS~storm_economic_health_data$EVTYPE,
storm_economic_health_data,sum)
colnames(properties_data) <-c("EVTYPE", "PROPDMDOLLARS")
properties_data <- arrange(properties_data,desc(PROPDMDOLLARS))
properties_data$PROPDMDOLLARS <- properties_data$PROPDMDOLLARS/1e9

crop_data <-
aggregate(storm_economic_health_data$CROPDMDOLLARS~storm_economic_health_data$EVTYPE,
storm_economic_health_data,sum)
colnames(crop_data) <-c("EVTYPE", "CROPDMDOLLARS")
crop_data <- arrange(crop_data,desc(CROPDMDOLLARS))
crop_data$CROPDMDOLLARS <- crop_data$CROPDMDOLLARS/1e9

```

Results

This section outlines the results of the data analysis. Two plot figures are generated. The first figure shows bar plots of the total number of fatalities and injuries that occurred in the period of 1950 to 2011 for different kinds of weather events. Only the top ten most severe weather events for fatalities and injuries are plotted. The second figure shows bar plots of crop and property damage, in billions of \$US, for the most damaging ten types of weather events.

Create two separate bar plots for the fatality and injury data and combine both plots into one figure.

```

plot_fatalities <- ggplot(fatalities_data[1:10,],
aes(reorder(EVTYPE,-FATALITIES),FATALITIES)) +
geom_bar(stat="identity", col = "black", fill = "limegreen") +
labs(title = "Total Fatalities by Event",x = "Event Type",y= "Total Fatalities") +
theme(text = element_text(size=16),axis.text.x = element_text(angle = 60, hjust = 1))

plot_injuries <- ggplot(injuries_data[1:10,], aes(reorder(EVTYPE,-INJURIES),INJURIES)) +
geom_bar(stat="identity", col = "black", fill = "#f4733b") +
labs(title = "Total Injuries by Event",x = "Event Type",y= "Total Injuries") +
theme(text = element_text(size=16),axis.text.x = element_text(angle = 60, hjust = 1))

plot_health <- grid.arrange(plot_fatalities, plot_injuries, ncol= 2)

```

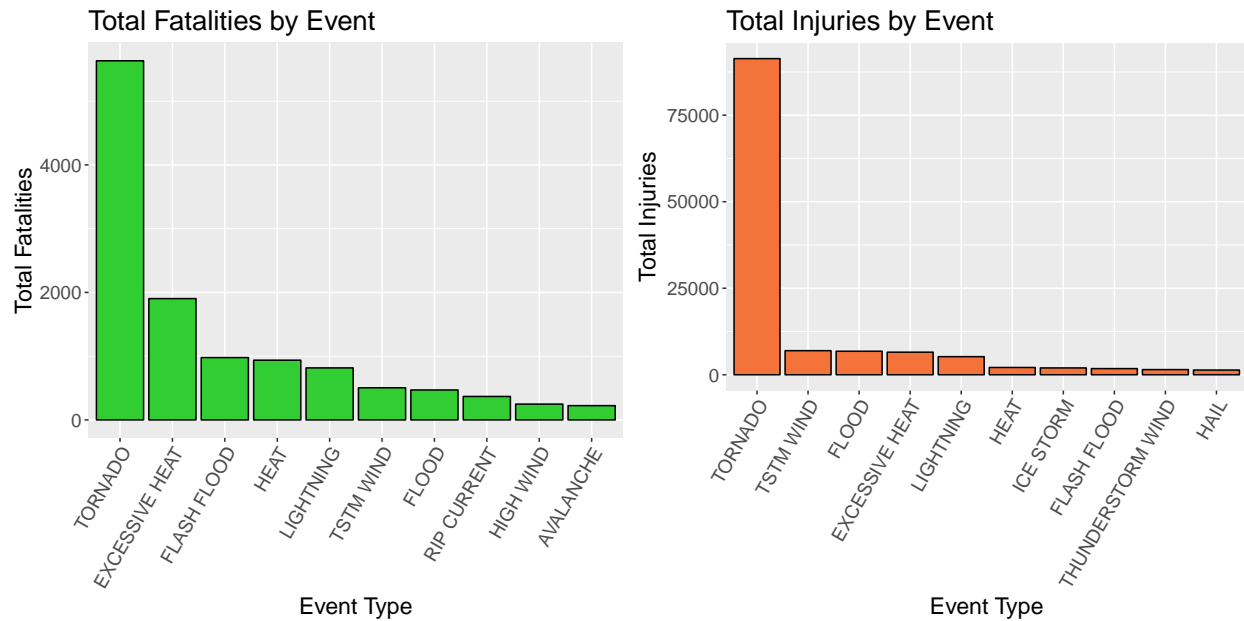


Figure 1: Total number of fatalities and injuries due to the ten most severe event types in each health category from 1950-2011.

Create two separate bar plots for the property and crop damage data and combine both plots into one figure.

```
plot_property <- ggplot(properties_data[1:10,],
  aes(reorder(EVTYPE,-PROPDMDOLLARS),PROPDMDOLLARS)) +
  geom_bar(stat="identity", col = "black", fill = "#ff5477") +
  labs(title = "Total Property Damage",
    x = "Event Type",y= "Total Property Damage ($US billions)") +
  theme(text = element_text(size=16),axis.text.x = element_text(angle = 60, hjust = 1))

plot_crops <- ggplot(crop_data[1:10,],
  aes(reorder(EVTYPE,-CROPDMDOLLARS),CROPDMDOLLARS)) +
  geom_bar(stat="identity", col = "black", fill = "#50bae8") +
  labs(title = "Total Crop Damage",x = "Event Type",
    y= "Total Crop Damage ($US billions)") +
  theme(text = element_text(size=16),axis.text.x = element_text(angle = 60, hjust = 1))

plot_economic <- grid.arrange(plot_property, plot_crops, ncol = 2)
```

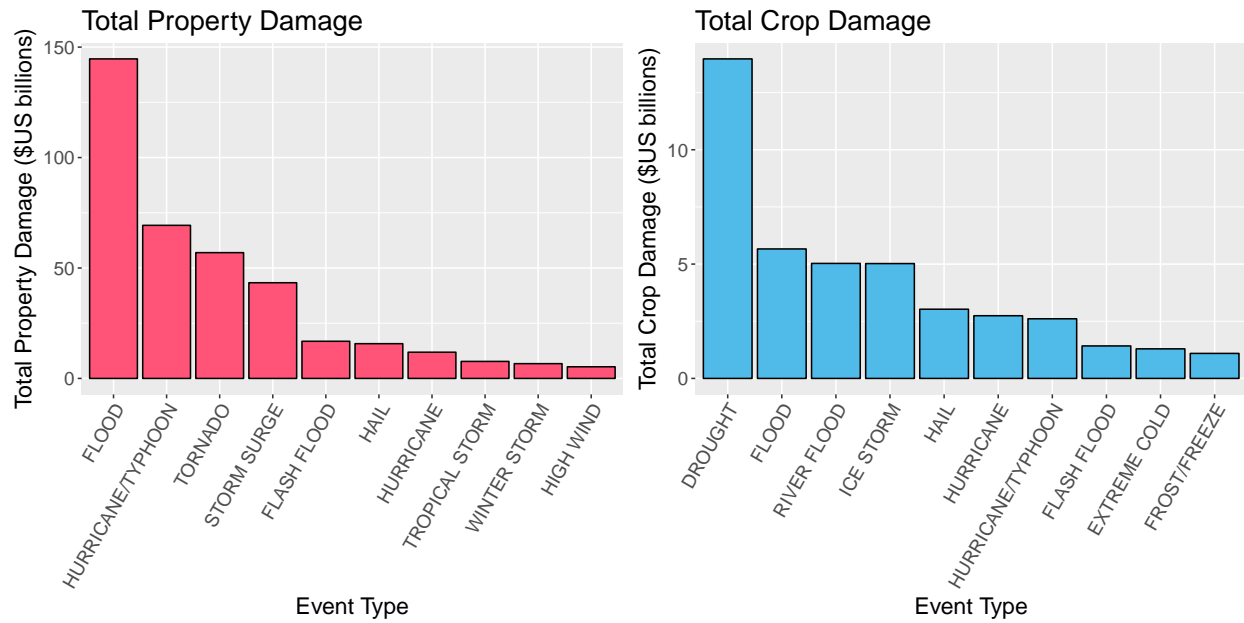


Figure 2: Total amount of damage in \$US billions to crops and property due to the ten most severe event types in each economic category from 1950-2011.

From these results, tornadoes appeared to cause the most fatalities and injuries over the specified time period. Floods appeared to do the most property damage and droughts caused the most crop damage over the same time period.