Health and Economic Cost of US Weather Events

Al Anderson

July 17, 2015

Synopsis

This report processes and analyses data from the United States National Oceanic and Atmospheric Administration (NOAA) and reports on the top 10 type of weather events that cause the most health issues, injuries and fatalities, and the most economic cost, property damage and crop damage. The data used for this report is available in the file located at https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2. The worst events that impact health are tornado and heat events. The worst events that impact economics are flooding and hurricane related events. Only the top 10 events in both categories because they have the greatest impact by far. Also, this report only looked at events since 2001 since building codes, evacuation, and weather forecasting have significantly improved.

Data Processing

This data set is somewhat messy due to it having quite old data and reporting standards have changed over time. This was another reason the report focuses on data since 2001. Several steps are involved to process the data to ready it for analysis. RStudio and the R language where uses to process the data and to produce the analysis.

Data Processing

Several R libraries are used with the processing and analysis. This code loads those libraries.

```
library(readr)
suppressWarnings(library(dplyr))
library(lubridate)
```

The next step is to read the data in. The data needs to located in the current R working directory for this to work.

```
storm_data <- read_csv("repdata-data-StormData.csv",
                       col_types = c("ccccccccccccccccccccccccccccccccccccccc"),
                       progress = FALSE)
```

The next code converts the columns of the data set to the proper types. It also converts the beginning date column to a proper date type for R. The code also modifies some of the weather event type to allow for better grouping of similar weather events. In other words all events involving non-thunderstorm wind are group together. It also helps with differences in reporing event types. The code then creates multiplier columns for the crop and property damage columns. Those columns are reported at integers with an associated column representing the exponent multiplier. We these use those columns along with the property damage and crop damage columns to add another column which represents the total economic damage caused by the weather event.

```
storm_data$BGN_DATE <- mdy_hms(storm_data$BGN_DATE)
storm_data$FATALITIES <- as.numeric(storm_data$FATALITIES)
storm_data$INJURIES <- as.numeric(storm_data$INJURIES)
storm_data$PROPDMG <- as.numeric(storm_data$PROPDMG)
storm_data$PROPDMGEXP <- as.character(storm_data$PROPDMGEXP)
storm_data$CROPDMG <- as.numeric(storm_data$CROPDMG)
storm_data$CROPDMGEXP <- as.character(storm_data$CROPDMGEXP)
storm_data$EVTYPE <- as.character(storm_data$EVTYPE)

storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="HIGH WIND",
                            "WIND", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="TSTM WIND",
                            "THUNDERSTORM WIND", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="HURRICANE/TYPHOON",
                            "HURRICANE", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="EXCESSIVE HEAT",
                            "HEAT", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="TSTM WIND/HAIL",
                            "WIND-HAIL", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="WILD/FOREST FIRE",
                            "WILDFIRE", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="URBAN/SML STREAM FLD",
                            "URBAN SML STREAM FLD", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="RECORD HEAT",
                            "HEAT", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="WND", "WIND", storm_data$EVTYPE)
storm_data$EVTYPE <- ifelse(storm_data$EVTYPE=="STRONG WIND",
                            "WIND", storm_data$EVTYPE)

dmg_exponent <- sort(unique(c(storm_data$PROPDMGEXP, storm_data$CROPDMGEXP)))
dmg_multipler <- c(0,0,0,0,1,10,100,1000,10000,100000,1000000,
                   10000000,100000000,1000000000, 1000000000,
                   100, 100, 1000, 1000000, 1000000)
dmg_multipler_df = data.frame(dmg_exponent, dmg_multipler)

storm_data <- storm_data %>%
        mutate(CROPDMG_MULT =
                        dmg_multipler_df$dmg_multipler[match(CROPDMGEXP,
                                                             dmg_multipler_df$dmg_exponent)])
storm_data <- storm_data %>%
        mutate(PROPDMG_MULT =
                        dmg_multipler_df$dmg_multipler[match(PROPDMGEXP,
                                                             dmg_multipler_df$dmg_exponent)])
storm_data$PROPDMG <- storm_data$PROPDMG * storm_data$PROPDMG_MULT
storm_data$CROPDMG <- storm_data$CROPDMG * storm_data$CROPDMG_MULT
storm_data <- storm_data %>% mutate(ECONDMG=PROPDMG + CROPDMG)
```

Analysis for Health Impacts of Weather Events

At this point the raw data set in the proper format to be analysed. We first tackle the health damage caused by weather events. First the code select events occurring on or after 01/01/2001. Then the code groups the data by weather event type. It then summarises the data by creating a column which sums the fatalities and injuries caused by the various weather events. Finally it arranges those events and select the top ten events.

This data is in the top10 data set.

```
health_storm_data <- storm_data %>%
        filter(as.Date(BGN_DATE) >= as.Date("2001/01/01")) %>%
        select(EVTYPE, FATALITIES, INJURIES)

health_storm_data <- health_storm_data %>% mutate(fat_inj_sum = FATALITIES + INJURIES)
grouped_by_EVTYPE <- health_storm_data %>% group_by(EVTYPE)
sum_health <- summarise(grouped_by_EVTYPE, Sum = sum(fat_inj_sum))
arrange_data <- arrange(sum_health, desc(Sum))
top10health <- arrange_data[1:10,]
```

Analysis for Economic Impacts of Weather Events

We now tackle the economic damage caused by weather events. First the code select events occurring on or after 01/01/2001. Then the code groups the data by weather event type. It then summarises the data by creating a column which sums the economic damage caused by the various weather events. Finally it arranges those events and select the top ten events. This data is in the top10 data set.

```
economic_storm_data <- storm_data %>%
        filter(as.Date(BGN_DATE) >+ as.Date("2001/01/01")) %>%
        select(EVTYPE, ECONDMG)

grouped_by_EVTYPE_ECON <- economic_storm_data %>% group_by(EVTYPE)
sum_econ <- summarise(grouped_by_EVTYPE_ECON, Sum = sum(ECONDMG))

arrange_data_econ <- arrange(sum_econ, desc(Sum))
top10econ <- arrange_data_econ[1:10,]
```

Results

Health Impacts from Weather Events in the US

The weather event that most impacts health are tornadoes followed by heat events. But tornadoes cause a huge number of injuries and fatalities compared to other weather events. Figure 1 clearly demonstrates this.

```
par(mai=c(1.4,1.7,0.82,0.42)+.1, mgp=c(4,0.5,0))
barplot(top10health$Sum, names.arg=top10health$EVTYPE, las=2, cex.names =  .5,
        xlab = "Figure 1 - Injuries & Fatalities", ylab = "",
        main = "Injuries & Fatalities from \nTop 10 Weather Events 2001 - 2011",
        horiz = TRUE)
```

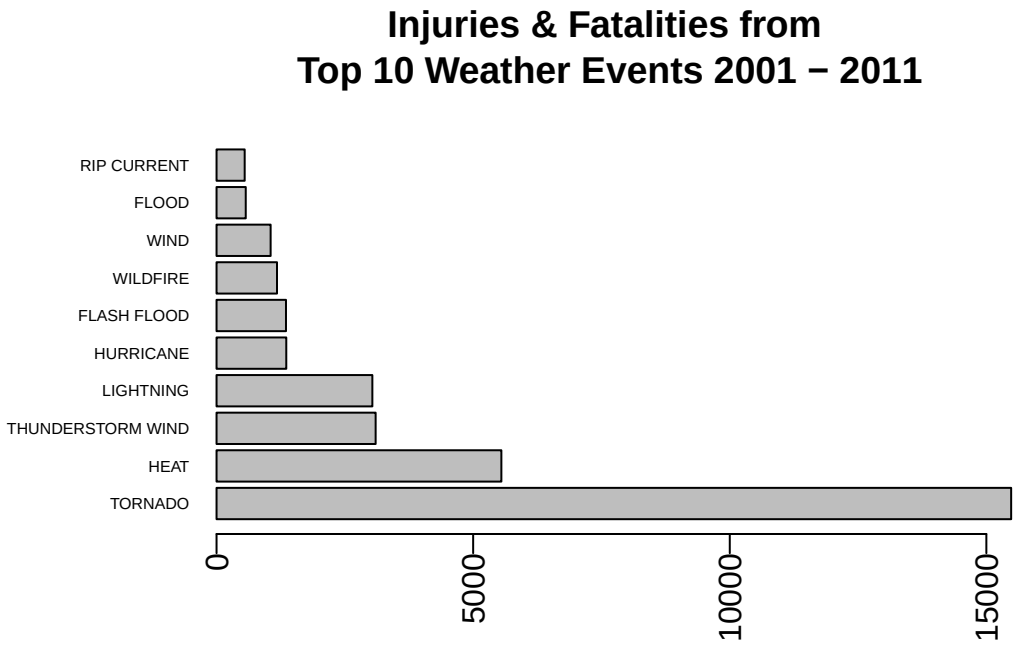

Figure 1 − Injuries & Fatalities

Economic Impacts from Weather Events in the US

The weather event that most impacts economics are floods, hurricanes, and storm surges. Figure 2 clearly demonstrates this.

```
par(mai=c(1.4,1.7,0.82,0.42)+.1, mgp=c(4,0.5,0))
barplot(top10econ$Sum, names.arg=top10econ$EVTYPE, las=2, cex.names =  .5,
       xlab = "Figure 2 - Economic Damage", ylab = "",
       main = "Economic Damage from \nTop 10 Weather Events 2001 - 2011",
       horiz = TRUE)
```

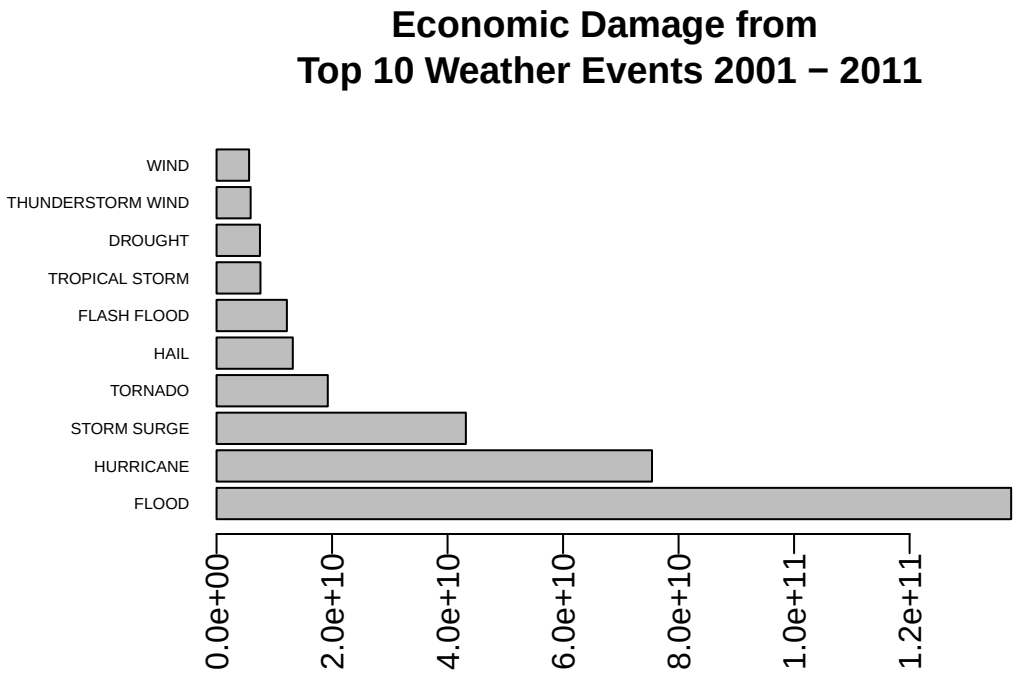

Figure 2 − Economic Damage