Analysis of NOAA Storm Data in terms of Health/Economic Consequences

Sofia

February 28, 2016

Synopsis

The goal of this analysis was to explore the NOAA Storm Database and answer 2 research questions concerning the effects of severe weather events: 1. Across the United States, which types of events are most harmful with respect to population health? 2. Across the United States, which types of events have the greatest economic consequences? The analysis has revealed that the top 5 severe weather events that are most harmful in terms of population health are: (1) Tornado, (2) Excessive Heat, (3) TSTM Wind, (4) Flood, and (5) Lightning. The top 5 severe weather events that have the most harmful economic consequences have been shown to be: (1) Thunderstorm Winds, (2) Tornadoes, TSTM Wind, Hail, (3) Heavy Rain, (4) Extreme Cold, and (5) Severe Thunderstorm.

Data Processing

The first step of data processing was to read the source data from the compressed csv file, as shown below.

```
## read from raw data file
data <- read.csv(bzfile("../data/repdata-data-StormData.csv.bz2"))
```

Once the data has been loaded into R, a series of transformations has been performed on the data frame in order to answer the research questions. (Note: During the transformations, several modified data frames (with distinct names) have been generated for the sake of better comprehensibility.)

First, only those variables that are relevant to answering the research questions have been selected.

```
library(dplyr)

## select variables relevant to answering the research questions
## EVTYPE: event type
## FATALITIES: number of fatalities
## INJURIES: number of injuries
## PROPDMG: amount of property damages incurred
## PROPDMGEXP: explanation on property damage amount (indication of unit)
## CROPDMG: amount of crop damages incurred
## CROPDMGEXP: explanation on crop damage amount (indication of unit)
selected_data <- select(data,
                        EVTYPE, FATALITIES, INJURIES,
                        PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

Next, the columns have been renamed with more descriptive names that clearly indicate their meaning.

```
## rename columns using more descriptive names
renamed_data <- selected_data %>%
    rename(event.type = EVTYPE,
```

```
           fatalities = FATALITIES,
           injuries = INJURIES,
           property.damage = PROPDMG,
           property.damage.unit = PROPDMGEXP,
           crop.damage = CROPDMG,
           crop.damage.unit = CROPDMGEXP)
```

Next, the data frame has been filtered to exclude those rows that have the value of zero (0) for all variables that indicate health or economic damages (i.e., fatalities, injuries, property.damage, crop.damage).

```
## filter data to exclude rows that have no info on health/economic damages
filtered_data <- renamed_data %>%
    filter(!(fatalities == 0 && injuries == 0 &&
                 property.damage == 0 && crop.damage == 0))
```

Next, the two factor variables that contain unit information on the property and crop damages (i.e., property.damage.unit and crop.damage.unit) have been converted to character variables in order to facilitate computation of actual damage values.

```
## convert factor variables to character variables to compute actual damage values
filtered_data$property.damage.unit <- as.character(filtered_data$property.damage.unit)
filtered_data$crop.damage.unit <- as.character(filtered_data$crop.damage.unit)
```

Next, two additional columns (i.e., property.damage.val and crop.damage.val) have been added to store property and crop damage currency amounts (in $).

```
## add 2 additional columns to store economic damage values,
## taking into account the specified unit values
## (i.e., H/h = 100, K/k = 1000, M/m = 1000000, B/b = 1000000000)
mutated_data <- filtered_data %>%
    mutate(property.damage.val = property.damage,
           crop.damage.val = crop.damage)
```

Next, using the information in property.damage.unit and crop.damage.unit columns, property and crop damage values have been computed (in $).

```
## compute economic damage values (in $)
mutated_data[mutated_data$property.damage.unit == "H", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "H", ]$property.damage.val *
    10^2
mutated_data[mutated_data$property.damage.unit == "h", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "h", ]$property.damage.val *
    10^2
mutated_data[mutated_data$property.damage.unit == "K", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "K", ]$property.damage.val *
    10^3
mutated_data[mutated_data$property.damage.unit == "k", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "k", ]$property.damage.val *
    10^3
mutated_data[mutated_data$property.damage.unit == "M", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "M", ]$property.damage.val *
    10^6
```

```r
mutated_data[mutated_data$property.damage.unit == "m", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "m", ]$property.damage.val *
    10^6
mutated_data[mutated_data$property.damage.unit == "B", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "B", ]$property.damage.val *
    10^9
mutated_data[mutated_data$property.damage.unit == "b", ]$property.damage.val <-
    mutated_data[mutated_data$property.damage.unit == "b", ]$property.damage.val *
    10^9

mutated_data[mutated_data$crop.damage.unit == "H", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "H", ]$crop.damage.val *
    10^2
mutated_data[mutated_data$crop.damage.unit == "h", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "h", ]$crop.damage.val *
    10^2
mutated_data[mutated_data$crop.damage.unit == "K", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "K", ]$crop.damage.val *
    10^3
mutated_data[mutated_data$crop.damage.unit == "k", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "k", ]$crop.damage.val *
    10^3
mutated_data[mutated_data$crop.damage.unit == "M", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "M", ]$crop.damage.val *
    10^6
mutated_data[mutated_data$crop.damage.unit == "m", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "m", ]$crop.damage.val *
    10^6
mutated_data[mutated_data$crop.damage.unit == "B", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "B", ]$crop.damage.val *
    10^9
mutated_data[mutated_data$crop.damage.unit == "b", ]$crop.damage.val <-
    mutated_data[mutated_data$crop.damage.unit == "b", ]$crop.damage.val *
    10^9
```

Next, a new, trimmed data frame has been generated, selecting only those variables that will be used in answering the research questions (i.e., event.type, fatalities, injuries, property.damage.val, crop.damage.val).

```r
## select variables indicating health/economic damages
adjusted_data <- mutated_data %>%
    select(event.type, fatalities, injuries, property.damage.val, crop.damage.val)
```

Next, the data has been grouped by event type, and summary stats representing the magnitude of health/economic damages have been generated.

```r
## group data by event type and summarize total damages
grouped_by_event_type <- adjusted_data %>%
    group_by(event.type) %>%
    summarise(total.fatalities = as.integer(sum(fatalities)),
              total.injuries = as.integer(sum(injuries)),
              total.health.damage = total.fatalities + total.injuries,
              total.property.damage = as.integer(sum(property.damage.val)),
              total.crop.damage = as.integer(sum(crop.damage.val)),
              total.economic.damage = total.property.damage + total.crop.damage)
```

Next, the grouped data has been sorted according to total health damage and total economic damage, respectively.

```
## sort data in terms of total health damages and total economic damages,
## respectively
sorted_by_health_damages <- grouped_by_event_type %>%
    arrange(desc(total.health.damage)) %>%
    select(event.type, total.fatalities, total.injuries,
           total.health.damage)

sorted_by_economic_damages <- grouped_by_event_type %>%
    arrange(desc(total.economic.damage)) %>%
    select(event.type, total.property.damage, total.crop.damage,
           total.economic.damage)
```

Finally, event.type variable has been converted to a factor variable according to the sorting order so as to faciliate plotting the results.

```
## make event type an ordered factor column for x-axis on plots
sorted_by_health_damages$event.type <-
    factor(sorted_by_health_damages$event.type,
           levels = sorted_by_health_damages$event.type)

sorted_by_economic_damages$event.type <-
    factor(sorted_by_economic_damages$event.type,
           levels = sorted_by_economic_damages$event.type)
```

Results

The results of analysis indicate that the following 5 severe weather events are most harmful with respect to population health:

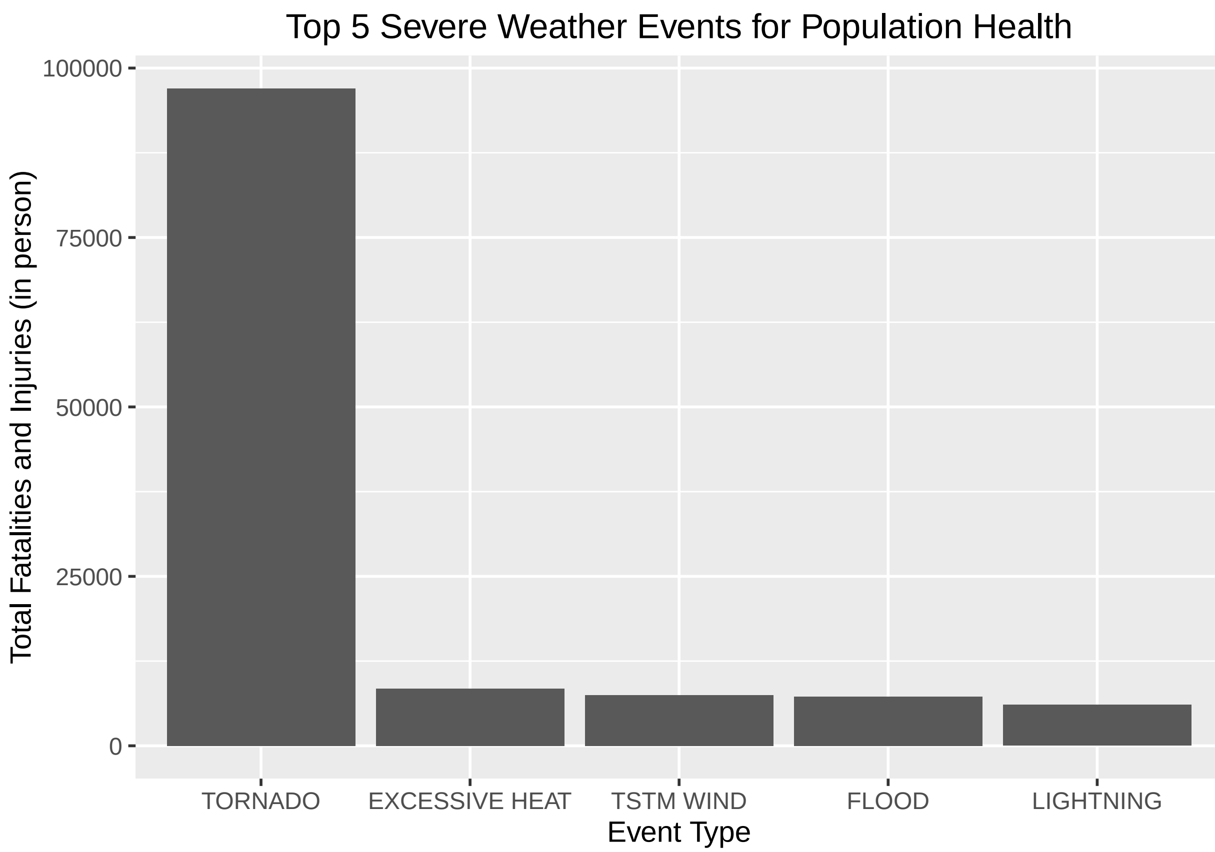
```
## display 5 events that are most harmful with respect to population health
head(sorted_by_health_damages, n=5)[, c("event.type", "total.health.damage")]
```

```
## Source: local data frame [5 x 2]
##
##       event.type total.health.damage
##           (fctr)               (int)
## 1        TORNADO               96979
## 2 EXCESSIVE HEAT                8428
## 3      TSTM WIND                7461
## 4          FLOOD                7259
## 5      LIGHTNING                6046
```

```
library(ggplot2)

## create a plot indicating top 5 severe weather events that are most harmful
## with respect to total population health
```

```
p1 <- ggplot(head(sorted_by_health_damages, n=5),
             aes(x=event.type, y=total.health.damage)) +
    geom_bar(stat="identity") +
    ggtitle("Top 5 Severe Weather Events for Population Health") +
    labs(x="Event Type", y="Total Fatalities and Injuries (in person)")
```



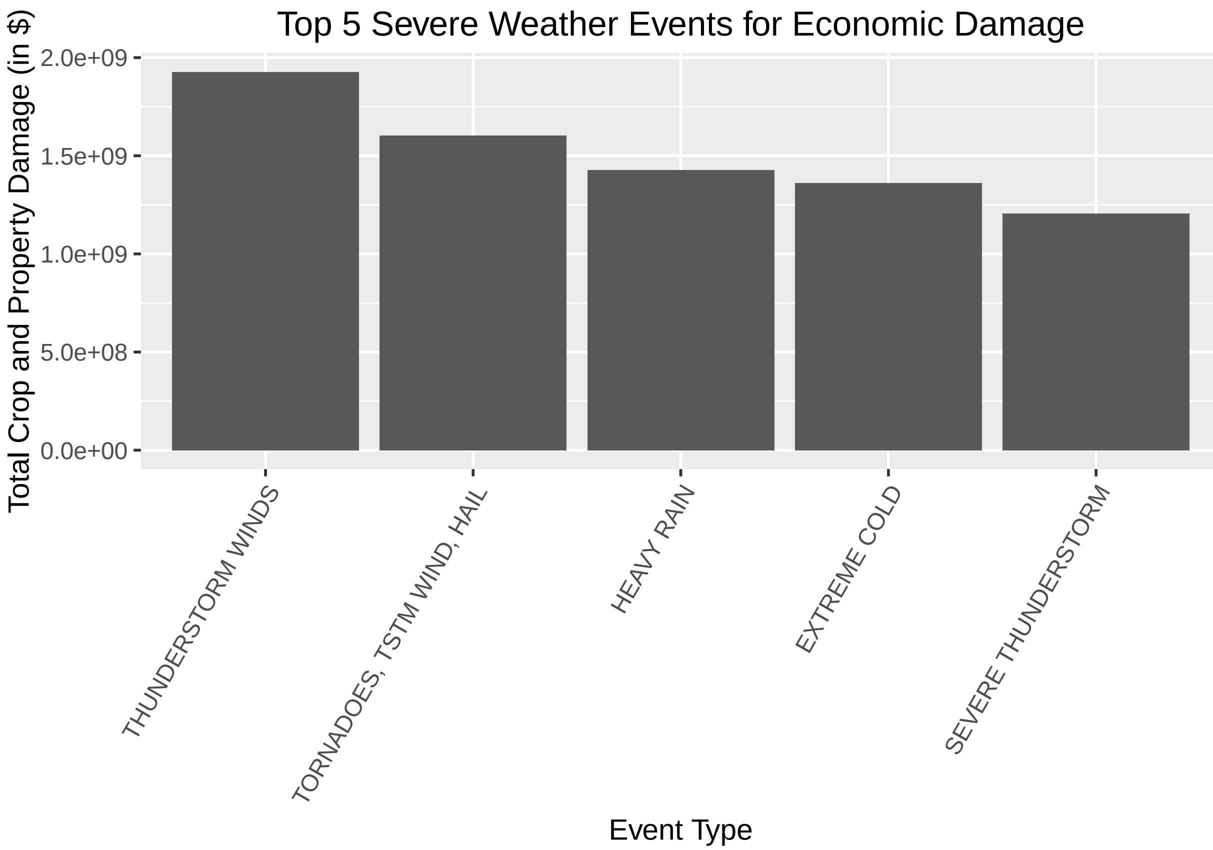
The result also indicate that the following 5 severe weather events are most harmful with respect to economic consequences:

```
## display 5 events that are most harmful with respect to economic consequences
head(sorted_by_economic_damages, n=5)[, c("event.type", "total.economic.damage")]
```

```
## Source: local data frame [5 x 2]
##
##                    event.type total.economic.damage
##                        (fctr)                 (int)
## 1          THUNDERSTORM WINDS            1926615791
## 2 TORNADOES, TSTM WIND, HAIL            1602500000
## 3                  HEAVY RAIN            1427647890
## 4                EXTREME COLD            1360710400
## 5         SEVERE THUNDERSTORM            1205560000
```

```
## create a plot indicating top 5 severe weather events that are most harmful
## with respect to economic consequences
p2 <- ggplot(head(sorted_by_economic_damages, n=5),
             aes(x=event.type, y=total.economic.damage)) +
```

```r
    geom_bar(stat="identity") +
    ggtitle("Top 5 Severe Weather Events for Economic Damage") +
    labs(x="Event Type", y="Total Crop and Property Damage (in $)") +
    theme(axis.text.x = element_text(angle = 60, hjust = 1))
```



Conclusion

The analysis described in this document has indicated that the Tornado event is most harmful, across the United States, in terms of total health damages (fatalities + injuries), whereas the Thunderstorm Winds event is most harmful in terms of total economic damages (property damages + crop damages).