Health and Economic Impacts of Storms in the United States

1/12/2022

Synopsis

This investigation aims to identify which types of storm events are most harmful to population health and which types of storm events have the greatest economic impact, across the entire United States. Injuries and fatalities are used to assess population health while property damage and crop damage costs are used to assess economic impact.

Data from the US National Oceanic and Atmospheric Administration are used for this investigation. Data cover the time period from January 1950 through November 2011. Note that earlier years have fewer entries and that more recent years are considered to be more complete.

Results from this investigation show that the most expensive storm event category is Flooding, which cost 180 billion dollars over the observation period. Hurricanes/Typhoons (91 billion), Tornados (57 billion), and Storm Surges (43 billion) were the next most expensive storm categories. Property damage accounted for the majority of the cost in most of the storm event categories, with the exception of drought which had expenses driven by crop damage.

In terms of population health, the most harmful storm events were Tornados, which injured or killed ~97,000 people during the observation period. Heat, Thunderstorm Wind, and Flooding were the next most harmful events, each injuring or killing between 10,000 and 12,500 people.

Data Processing

After downloading the storm data files and loading the necessary analytic packages, the data was processed as follows to streamline analysis:

1. Property damage and crop damage amounts were converted from hundreds, thousands, millions, or billions of dollars to dollars to enable direct comparisons. Note that property or crop damage amounts could not be identified for approximately 0.04% of all storm events. Those events were not included in the economic portion of this analysis.

2. Total number of persons harmed was calculated by summing together the number of injuries and fatalities from any given storm episode.

3. Total number of dollars spent on damage was calculated by summing together the property damage and crop damage amounts.

4. The storm event identifier variable (i.e. "EVTYPE") was cleaned to create more consistent categorizations. This included removing leading/trailing spaces, removing hyphens and back slashes, and accounting for minor differences in spelling or phrasing (e.g., 'Floods' and 'Flooding' were combined into a single category).

```
#Load required packages
library(tidyverse)
library(scales)
library(reshape2)
```

```r
#Download data file and read into R
download.file(url =
↪  "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile
↪  = "./stormdata.csv.bz2", method = "curl")
storm_data <- read.csv("stormdata.csv.bz2")

#Create a cleaned data file that standardizes property damage and crop damage amounts to
↪  dollars, and calculates total damage amount and total persons harmed for each event
storm_data_clean <- storm_data %>%
    mutate(
        PROPDMGDOLLAR = case_when(
            PROPDMGEXP %in% c('B', 'b') ~ PROPDMG * 10^9,
            PROPDMGEXP %in% c('M', 'm') ~ PROPDMG * 10^6,
            PROPDMGEXP %in% c('K', 'k') ~ PROPDMG * 10^3,
            PROPDMGEXP %in% c('H', 'h') ~ PROPDMG * 10^2,
            PROPDMGEXP == '' ~ PROPDMG
        ),
        CROPDMGDOLLAR = case_when(
            CROPDMGEXP %in% c('B', 'b') ~ CROPDMG * 10^9,
            CROPDMGEXP %in% c('M', 'm') ~ CROPDMG * 10^6,
            CROPDMGEXP %in% c('K', 'k') ~ CROPDMG * 10^3,
            CROPDMGEXP %in% c('H', 'h') ~ CROPDMG * 10^2,
            CROPDMGEXP == '' ~ CROPDMG
        ),
        TOTAL_DOLLAR_DAMAGE = PROPDMGDOLLAR + CROPDMGDOLLAR,
        TOTAL_PERSON_HARMED = INJURIES + FATALITIES
    )

#Remove leading and trailing spaces, hyphens, and back slashes from event names
storm_data_clean$EVTYPE_CLN <- toupper(storm_data_clean$EVTYPE) %>% trimws(which =
↪  "both")
storm_data_clean$EVTYPE_CLN <- gsub("-|/", " ", storm_data_clean$EVTYPE_CLN)

#Create standard naming conventions for most common events and eliminate differences in
↪  event categorization due to spelling or minor phrasing discrepancies
storm_data_clean$EVTYPE_CLN <- gsub("^THUNDERSTORM(.*)|^TSTM WIND(.*)", "THUNDERSTORM
↪  WIND", storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("^MARINE THUNDERSTORM(.*)|^MARINE TSTM WIND(.*)",
↪  "MARINE THUNDERSTORM WIND", storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)HURRICANE(.*)|(.*)TYPHOON(.*)", "HURRICANE
↪  (TYPHOON)", storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)BLIZZARD(.*)", "BLIZZARD",
↪  storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)HAIL(.*)", "HAIL", storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)TROPICAL STORM(.*)", "TROPICAL STORM",
↪  storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)HIGH WIND(.*)", "HIGH WINDS",
↪  storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)TORNADO(.*)", "TORNADO",
↪  storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)RIP CURRENT(.*)", "RIP CURRENT",
↪  storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)FLOOD(.*)", "FLOOD",
↪  storm_data_clean$EVTYPE_CLN)
```

```
storm_data_clean$EVTYPE_CLN <- gsub("(.*)COLD(.*)", "COLD", storm_data_clean$EVTYPE_CLN)
storm_data_clean$EVTYPE_CLN <- gsub("(.*)HEAT(.*)", "HEAT", storm_data_clean$EVTYPE_CLN)
↪
```

Following data cleaning, two summary datasets (one summarizing economic effects and one summarizing health effects) were created for final analysis. The steps for creating these summary files were as follows:

1. The property damage cost, crop damage cost, total damage cost, injury count, fatality count, and total person harmed count were aggregated by storm event type. (i.e., each of these six variables was summed across all storm episodes corresponding to a given storm event type to get a total amount for that event.)

2. The 10 storm events with the highest total damage cost were identified and the data were manipulated into a 'long' format which could be more readily plotted. (Summary dataset #1)

3. The 10 storm events with the highest total number of persons harmed were identified and the data were manipulated into a 'long' format which could be more readily plotted. (Summary dataset #2)

```
#Creates a summary dataset with cost and person harmed variables aggregated by storm
↪  event type
storm_data_sum <- aggregate(
    storm_data_clean[,c("TOTAL_DOLLAR_DAMAGE","PROPDMGDOLLAR","CROPDMGDOLLAR",
↪  "TOTAL_PERSON_HARMED", "FATALITIES","INJURIES")],
    by = list(storm_data_clean$EVTYPE_CLN),
    FUN = sum,
    na.rm = TRUE)

#Identifies the 10 storm event categories with the highest cost (sum of property damage
↪  and crop damage). Transforms data into 'long' format.
economic_storm_data <-
↪  storm_data_sum[order(-storm_data_sum$TOTAL_DOLLAR_DAMAGE),][1:10,1:4]
economic_pivot <- economic_storm_data %>%
                    pivot_longer(cols=3:4,
                        names_to = "CATEGORY",
                        values_to = "DAMAGE_AMT") %>%
                    rename(EVTYPE_CLN = Group.1)

#Identifies the 10 storm event categories with the highest number of persons harmed (sum
↪  of fatalities and injuries). Transforms data into 'long' format.
health_storm_data <-
↪  storm_data_sum[order(-storm_data_sum$TOTAL_PERSON_HARMED),][1:10,c(1,5:7)]
health_pivot <- health_storm_data %>%
                pivot_longer(cols=3:4,
                             names_to = "CATEGORY",
                             values_to = "PERSON_HARMED")  %>%
                rename(EVTYPE_CLN = Group.1)
```

Results

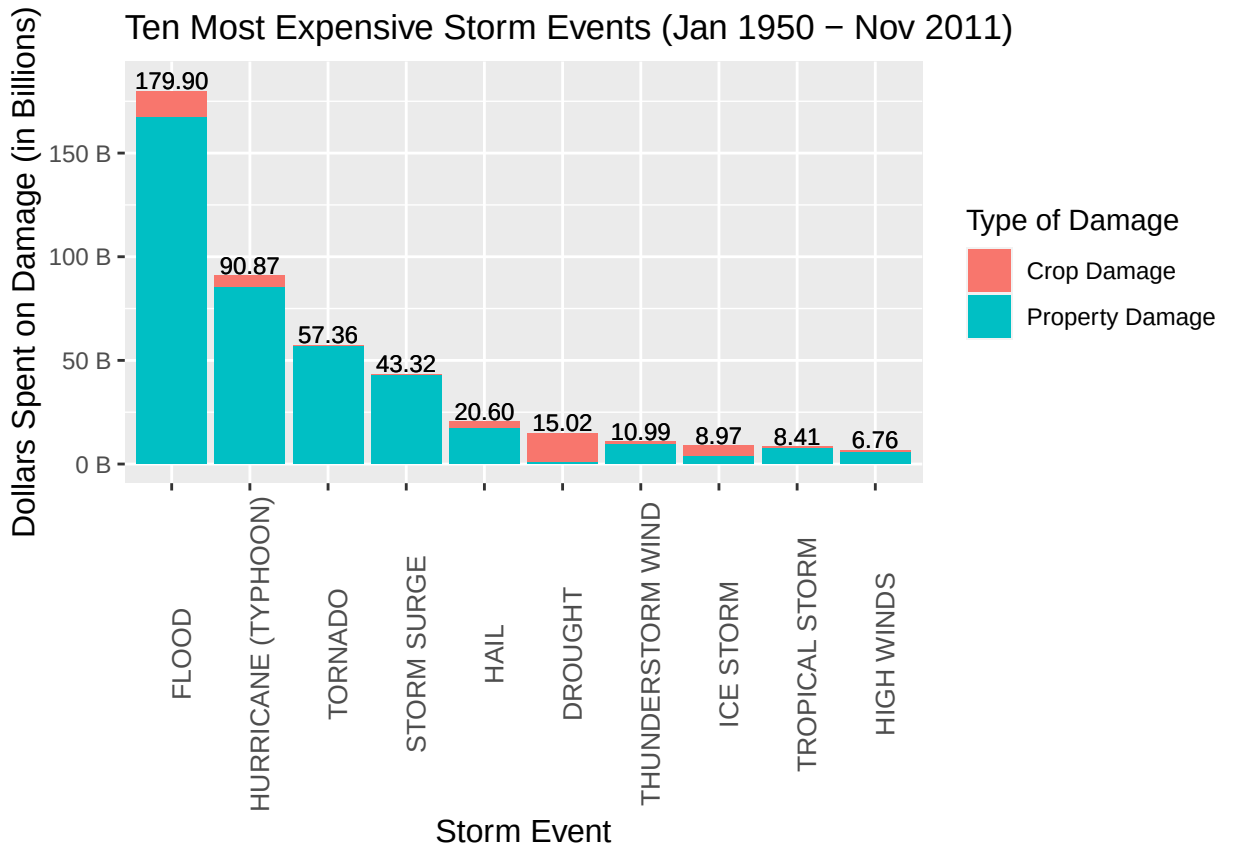
Two plots were created to summarize the results of this investigation:

1. The first plot displays the number of US dollars spent on storm-related damage for the 10 most expensive storm events experienced between Jan 1950 and Nov 2011. The total number of dollars (in billions) is displayed at the top of each bar, and color is used to indicate the amount spent on property damage and the amount spent on crop damage. The most expensive storm event by far is flooding (~180 billion

dollars), following by hurricane/typhoon (~91 billion dollars). For most storm events, the amount spent on property damage exceeds the amount spent on crop damage with the notable exception of drought.

2. The second plot displays the number of persons harmed for the 10 most harmful storm events experienced between Jan 1950 and Nov 2011. The total number of persons harmed is displayed at the top of each bar, and color is used to indicate the number of fatalities and the number of injuries. The most harmful storm event by far are tornados (~97,000 people injured or killed). Heat, thunderstorm winds, and flooding are the next three most harmful events to population health. Injuries are more common than fatalities for all storm events.

```
ggplot(economic_pivot, aes(x=reorder(EVTYPE_CLN, -TOTAL_DOLLAR_DAMAGE), y=DAMAGE_AMT)) +
    geom_col(aes(fill = CATEGORY)) +
    geom_text(aes(y = TOTAL_DOLLAR_DAMAGE + 5*10^9, label =
    ↪  scales::comma(TOTAL_DOLLAR_DAMAGE/10^9)), color = "black", size = 3) +
    labs (x="Storm Event", y = "Dollars Spent on Damage (in Billions)", title = "Ten Most
    ↪  Expensive Storm Events (Jan 1950 - Nov 2011)") +
    scale_y_continuous(labels = unit_format(unit = "B", scale = 1e-9)) +
    theme(axis.text.x = element_text(angle = 90, size = 10), axis.title =
    ↪  element_text(size = 12)) +
    scale_fill_discrete(name = "Type of Damage",labels = c("Crop Damage", "Property
    ↪  Damage"))
```




```
ggplot(health_pivot, aes(x=reorder(EVTYPE_CLN, -TOTAL_PERSON_HARMED), y=PERSON_HARMED)) +
            geom_col(aes(fill = CATEGORY)) +
            geom_text(aes(y = TOTAL_PERSON_HARMED + 3000, label =
            ↪  scales::comma(TOTAL_PERSON_HARMED)), color = "black", size = 3) +
            labs (x="Storm Event", y = "Number of Persons Injured or Killed", title =
            ↪  "Ten Most Harmful Storm Events (Jan 1950 to Nov 2011)") +
```

```
        scale_y_continuous(labels = comma) +
        theme(axis.text.x = element_text(angle = 90, size = 10), axis.title =
        ↪  element_text(size = 12)) +
        scale_fill_discrete(name = "",labels = c("Fatalities", "Injuries"))
```