Severe Weather Impact

Joe Okelly

2024-12-30

1. Introduction

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

2. Data Processing

The data for this analysis is sourced from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. The dataset, provided in a compressed CSV file format, includes records of major storms and severe weather events in the United States. For this project, we focus on identifying patterns in population health impacts (fatalities and injuries) and economic impacts (property and crop damage).

The dataset spans from 1950 to November 2011. However, early records (1950-1995) predominantly document tornadoes, thunderstorm winds, and hail, with more comprehensive data available starting in January 1996. To ensure a balanced and unbiased analysis, we limit our study to events from 1996 onward. The following subsections detail the preprocessing steps applied to the data.

- Storm Data [47Mb]
- National Weather Service Storm Data Documentation
- National Climatic Data Center Storm Events FAQ

2.1. Loading Data

```r
# Load libraries

library(scales)
library(tidyverse)
library(lubridate)

library(knitr)

# Download data

if(!file.exists("repdata_data_StormData.csv.bz2")) {
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                "repdata_data_StormData.csv.bz2") }

# Read data

noaa <- read.csv("repdata_data_StormData.csv.bz2",
                 stringsAsFactors = F,
                 sep = ",",
                 strip.white = T,
                 na.strings = "")
```

2.2. Subseting Data

Population health impacts are measured as the fatalities and injuries resulting from severe weather, and economic impacts as property and crop damage. Variables of interest for this analysis are:

- **EVTYPE**: Severe weather event type
- **BGN_DATE**: Start date of severe weather event
- **FATALITIES**: Number of deaths resulting from the severe weather event
- **INJURIES**: Number of injuries resulting from the severe weather event
- **PROPDMG**: Property damage (base amount)
- **PROPDMGEXP**: Property damage multiplier (e.g. K: 1,000; M: 1,000,000, etc.)
- **CROPDMG**: Crop damage (base amount)
- **PROPDMGEXP**: Crop damage multiplier (e.g. K: 1,000; M: 1,000,000, etc.)

The events in the database start in 1950 and end in November 2011. Earlier records only report tornados, thunderstorm wind, and hail. Records beginning from January 1996 report more comprehensive severe weather event types and are considered more complete. We restrict the analysis to weather events from January 1996 and onwards to prevent bias from earlier records limited only to tornados, thunderstorm wind, and hail events.

```r
# Subset weather type, date, health, and economic variables

noaa.subset <- select(noaa,
                      EVTYPE,
                      BGN_DATE,
                      FATALITIES,
                      INJURIES,
                      PROPDMG,
                      PROPDMGEXP,
                      CROPDMG,
                      CROPDMGEXP) %>%

  # Filter weather events from 1996 onwards

  mutate(BGN_DATE = mdy_hms(BGN_DATE)) %>%
  filter(year(BGN_DATE) >= 1996)
```

2.3. Recoding Severe Weather Types

The EVTYPE variable contains over 400 unique weather event types, many of which are inconsistent or redundant. To address this, we recoded the event types to align with the 48 standardized storm event categories specified in the NOAA Storm Data Documentation (Table 2.1.1). This process involved:

Identifying allowed event types: These include categories such as "Flood," "Hurricane (Typhoon)," and "Winter Storm."

Splitting the dataset: Separating records into those that match the allowed event types and those requiring recoding.

Recoding based on patterns: Using regular expressions to map similar or misspelled event types to their corresponding allowed categories (e.g., "HURRIC" mapped to "Hurricane (Typhoon)").

Merging datasets: Reintegrating the recoded and already-matching records into a single dataset for subsequent analysis.

```r
# Permittted storm events as listed in data documentation

permitted.storm.events <- toupper(c("Astronomical Low Tide",    "Avalanche",
                                    "Blizzard",                 "Coastal Flood",
                                    "Cold/Wind Chill",          "Debris Flow",
                                    "Dense Fog",                "Dense Smoke",
                                    "Drought",                  "Dust Devil",
                                    "Dust Storm",               "Excessive Heat",
                                    "Extreme Cold/Wind Chill",  "Flash Flood",
                                    "Flood",                    "Frost/Freeze",
                                    "Funnel Cloud",             "Freezing Fog",
                                    "Hail",                     "Heat",
                                    "Heavy Rain",               "Heavy Snow",
                                    "High Surf",                "High Wind",
                                    "Hurricane (Typhoon)",      "Ice Storm",
                                    "Lake-Effect Snow",         "Lakeshore Flood",
                                    "Lightning",                "Marine Hail",
                                    "Marine High Wind",         "Marine Strong Wind",
                                    "Marine Thunderstorm Wind", "Rip Current",
                                    "Seiche",                   "Sleet",
                                    "Storm Surge/Tide",         "Strong Wind",
                                    "Thunderstorm Wind",        "Tornado",
                                    "Tropical Depression",      "Tropical Storm",
                                    "Tsunami",                  "Volcanic Ash",
                                    "Waterspout",               "Wildfire",
                                    "Winter Storm",             "Winter Weather"))

# Split data based on allowed and not allowed event types

noaa.subset.event.allowed <- filter(noaa.subset,  EVTYPE %in% permitted.storm.events)
noaa.subset.event.recode  <- filter(noaa.subset, !EVTYPE %in% permitted.storm.events)

# Recode event types with more than 100 occurances into allowed event types

noaa.subset.event.recode <- mutate(noaa.subset.event.recode,
  EVTYPE = case_when(
    grepl("MARINE",              EVTYPE) ~ "MARINE THUNDERSTORM WIND",
    grepl("TSTM|THUNDE",         EVTYPE) ~ "THUNDERSTORM",
    grepl("FIRE",                EVTYPE) ~ "WILDFIRE",
    grepl("COASTAL",             EVTYPE) ~ "COASTAL FLOOD",
    grepl("FLD|FLOOD",           EVTYPE) ~ "FLOOD",
    grepl("WINTER|SNOW",         EVTYPE) ~ "WINTER WEATHER",
    grepl("COLD|CHILL",          EVTYPE) ~ "COLD/WIND CHILL",
    grepl("FOG",                 EVTYPE) ~ "DENSE FOG",
    grepl("HURRIC|TYPHOO",       EVTYPE) ~ "HURRICANE (TYPHOON)",
    grepl("WARM|HEAT|HOT",       EVTYPE) ~ "HEAT",
    grepl("WIND",                EVTYPE) ~ "STRONG WIND",
    grepl("RIP ",                EVTYPE) ~ "RIP CURRENT",
    grepl("SURGE",               EVTYPE) ~ "STORM SURGE/TIDE",
    grepl("SURF",                EVTYPE) ~ "HIGH SURF",
    grepl("BLIZZ",               EVTYPE) ~ "BLIZZARD",
    grepl("FROST|FREEZ|ICY|ICE", EVTYPE) ~ "FROST/FREEZE",
                                    TRUE ~ "OTHER"))

# Join data frames

noaa.subset <- rbind(noaa.subset.event.allowed, noaa.subset.event.recode)
```

2.4. Calculating Total Economic Costs

Economic impacts are assessed by combining property and crop damages. The PROPDMGEXP and CROPDMGEXP variables indicate the magnitude of damage, coded as:

K: Thousands (1,000)

M: Millions (1,000,000)

B: Billions (1,000,000,000)

Numeric values (1-10): Power of ten (e.g., 3 = 10^3 = 1,000)

To calculate total damages:

The base damage values (PROPDMG and CROPDMG) were multiplied by their respective multipliers (PROPDMGEXP and CROPDMGEXP).

Any undefined or unrecognized multipliers were defaulted to the base damage values, assuming no multiplier.

The results were stored as new variables: PROPDMG.TOTAL and CROPDMG.TOTAL, representing the total property and crop damage in monetary terms.

These calculations allow us to quantify the economic toll of severe weather events and identify the most financially impactful event types.

```r
# Calculate economic costs as total dollars

noaa.subset <- mutate(noaa.subset,

  # Capitalize multiplier codes

  PROPDMGEXP = toupper(PROPDMGEXP),
  CROPDMGEXP = toupper(CROPDMGEXP),

  # Multiply base damage with appropriate multiplier

  PROPDMG.TOTAL = case_when(
             PROPDMGEXP == "H" ~ PROPDMG * 1e+02,          # Hundreds
             PROPDMGEXP == "K" ~ PROPDMG * 1e+03,          # Thousands
             PROPDMGEXP == "M" ~ PROPDMG * 1e+06,          # Millions
             PROPDMGEXP == "B" ~ PROPDMG * 1e+09,          # Billions
    grepl("[0-9]", PROPDMGEXP) ~ PROPDMG * 10^as.numeric(PROPDMGEXP),
                          TRUE ~ PROPDMG),
  CROPDMG.TOTAL = case_when(
             CROPDMGEXP == "H" ~ CROPDMG * 1e+02,          # Hundreds
             CROPDMGEXP == "K" ~ CROPDMG * 1e+03,          # Thousand
             CROPDMGEXP == "M" ~ CROPDMG * 1e+06,          # Millions
             CROPDMGEXP == "B" ~ CROPDMG * 1e+09,          # Billions
    grepl("[0-9]", CROPDMGEXP) ~ PROPDMG * 10^as.numeric(CROPDMGEXP),
                          TRUE ~ CROPDMG))
```

3. Results

3.1. Health Impacts

Across the United States from 1996 to 2011, out of all severe weather events, excessive heat caused the greatest number of fatalities. Excessive heat accounts for `20.6%` of all severe weather–related fatalities (`1,787` of `8,732`), followed closely by tornados, `17.3%`. However, tornados account for the greatest number of injuries by far: `42.2%` of all severe weather–related injuries (`20,667` of `48,917`).

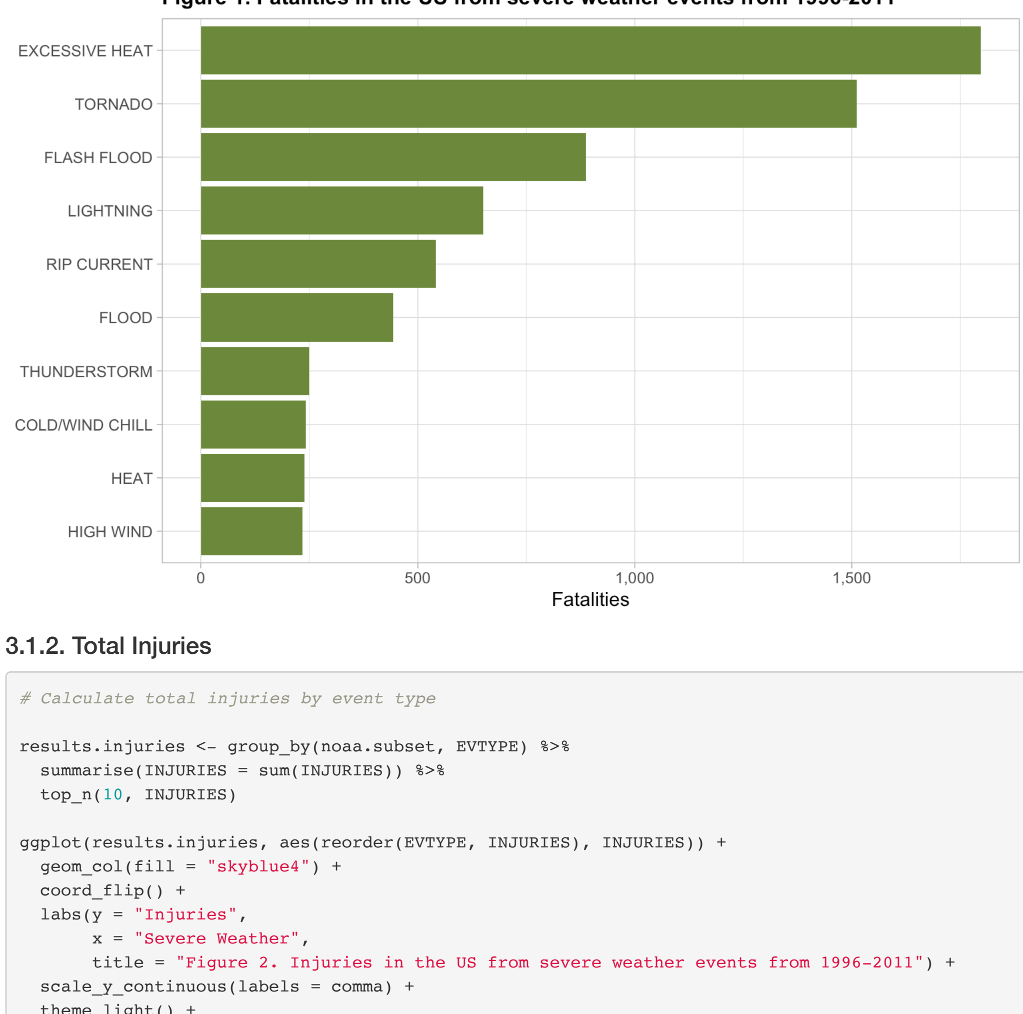
3.1.1. Total Fatalities

```r
# Calculate total fatalies by event type

results.fatalities <- group_by(noaa.subset, EVTYPE) %>%
  summarise(FATALITIES = sum(FATALITIES)) %>%
  top_n(10, FATALITIES)

ggplot(results.fatalities, aes(reorder(EVTYPE, FATALITIES), FATALITIES)) +
  geom_col(fill = "darkolivegreen4") +
  coord_flip() +
  labs(y = "Fatalities",
       x = "Severe Weather",
       title = "Figure 1. Fatalities in the US from severe weather events from 1996-2011") +
  scale_y_continuous(labels = comma) +
  theme_light()+
  theme(axis.title.y = element_blank(),
        plot.title   = element_text(size = 12,
                                    face = "bold"))
```

Figure 1. Fatalities in the US from severe weather events from 1996-2011



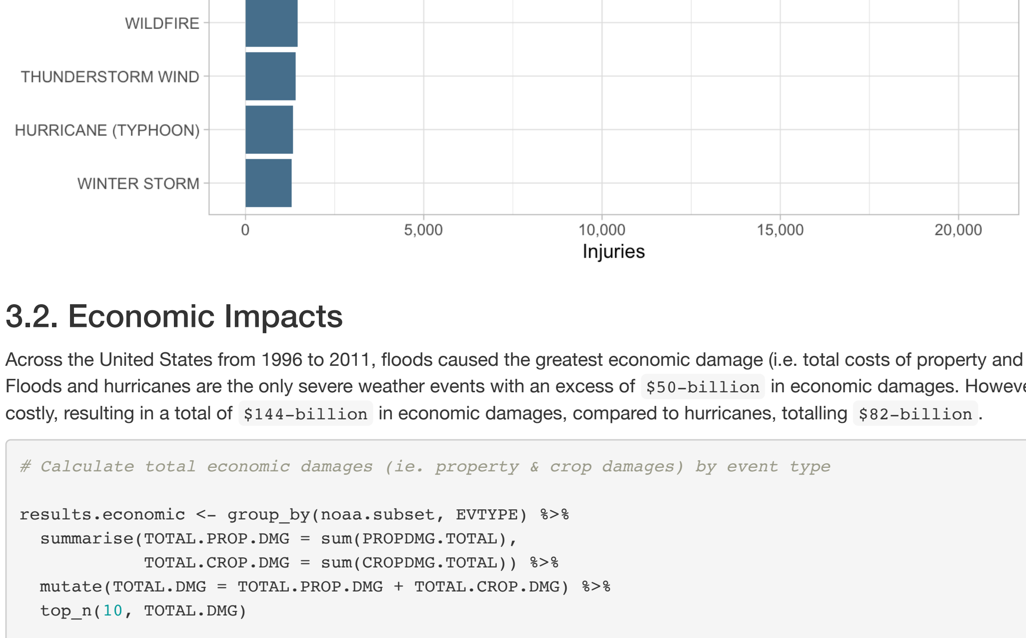
3.1.2. Total Injuries

```r
# Calculate total injuries by event type

results.injuries <- group_by(noaa.subset, EVTYPE) %>%
  summarise(INJURIES = sum(INJURIES)) %>%
  top_n(10, INJURIES)

ggplot(results.injuries, aes(reorder(EVTYPE, INJURIES), INJURIES)) +
  geom_col(fill = "skyblue4") +
  coord_flip() +
  labs(y = "Injuries",
       x = "Severe Weather",
       title = "Figure 2. Injuries in the US from severe weather events from 1996-2011") +
  scale_y_continuous(labels = comma) +
  theme_light() +
  theme(axis.title.y = element_blank(),
        plot.title   = element_text(size = 12,
                                    face = "bold"))
```

Figure 2. Injuries in the US from severe weather events from 1996-2011



3.2. Economic Impacts

Across the United States from 1996 to 2011, floods caused the greatest economic damage (i.e. total costs of property and crop damages). Floods and hurricanes are the only severe weather events with an excess of `$50-billion` in economic damages. However, floods are far more costly, resulting in a total of `$144-billion` in economic damages, compared to hurricanes, totalling `$82-billion`.

```r
# Calculate total economic damages (ie. property & crop damages) by event type

results.economic <- group_by(noaa.subset, EVTYPE) %>%
  summarise(TOTAL.PROP.DMG = sum(PROPDMG.TOTAL),
            TOTAL.CROP.DMG = sum(CROPDMG.TOTAL)) %>%
  mutate(TOTAL.DMG = TOTAL.PROP.DMG + TOTAL.CROP.DMG) %>%
  top_n(10, TOTAL.DMG)

# Prepare data for plotting and scale to billions

results.economic <- select(results.economic, -TOTAL.DMG) %>%
  gather("DMG.TYPE", "DMG.DOLLARS.BILLION", -EVTYPE) %>%
  mutate(DMG.DOLLARS.BILLION = round(DMG.DOLLARS.BILLION / 1e+09, 2),
         DMG.TYPE = recode_factor(DMG.TYPE,
                                  TOTAL.CROP.DMG = "Crop",
                                  TOTAL.PROP.DMG = "Property"))

ggplot(results.economic, aes(reorder(EVTYPE, DMG.DOLLARS.BILLION),
                             DMG.DOLLARS.BILLION, fill = DMG.TYPE)) +
  geom_col() +
  coord_flip() +
  labs(y = "Billions of Dollars",
       x = "Severe Weather",
       fill = "Damage",
       title = "Figure 3. Damages in the US from severe weather events from 1996-2011") +
  guides(fill = guide_legend(reverse=T)) +
  scale_y_continuous(labels = comma) +
  theme_light() +
  theme(legend.position = "bottom",
        axis.title.y    = element_blank(),
        plot.title      = element_text(size = 12,
                                       face = "bold"))
```

Figure 3. Damages in the US from severe weather events from 1996-2011