Analysis of the Adverse Health and Economic Impacts of US Storms

Farina Fayyaz

2024-02-12

1. Synopsis

This research delves into the NOAA Storm Database, spanning 1950 to November 2011, to assess the impact of severe weather events on human health, property, and crops. Through detailed analysis, we aim to identify which types of events pose the greatest threat to
1. Population 2. Economy measured by injuries, fatalities, and economic consequences.
For more information about Data: Documentation

2. Research Questions

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?

3. Data Processing

3.1. Downloading the Data

```
if (!dir.exists("Data")) {
    dir.create("Data")
    url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(url, destfile = "Data/data.csv")
}
```

3.2 Loading Data into R

```
# Loading dependencies
library(data.table)
# Reading Data from .csv file
data <- read.csv("Data/data.csv")
# Converting to data.table
df <- as.data.table(data)
```

3.3 Identifying Column Names

```
# Identifying Column Names
names(df)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
```

```
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

3.4 Subsetting the Data

We only need to keep columns "EVTYPE", "INJURIES","FATALITIES","PROPDMG","PROPDMGEXP", "CROPDMG" and "CROPDMGEXP" for our analysis. Also, we'll only use data where fatalities and injuries occurred.

```
# Keep only necessary columns
desired_columns <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG",
    "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
stormdata <- df[, ..desired_columns]
# Select rows where either fatalities or injuries are
# greater than 0
subsetdf <- stormdata[stormdata$FATALITIES > 0 | stormdata$INJURIES >
    0, ]
```

3.5 Data Cleaning

To calculate Property Damage and Crop Cost we need to clean relevant columns

```
# Change all damage exponents to uppercase.
cols <- c("PROPDMGEXP", "CROPDMGEXP")
subsetdf[, (cols) := lapply(.SD, toupper), .SDcols = cols]

# Map property damage alphanumeric exponents to numeric
# values.
propDmgKey <- c(`""` = 10^0, `-` = 10^0, `+` = 10^0, `0` = 10^0,
    `1` = 10^1, `2` = 10^2, `3` = 10^3, `4` = 10^4, `5` = 10^5,
    `6` = 10^6, `7` = 10^7, `8` = 10^8, `9` = 10^9, H = 10^2,
    K = 10^3, M = 10^6, B = 10^9)

# Map crop damage alphanumeric exponents to numeric values
cropDmgKey <- c(`""` = 10^0, `?` = 10^0, `0` = 10^0, K = 10^3,
    M = 10^6, B = 10^9)

# Apply mapping to columns
subsetdf[, PROPDMGEXP := propDmgKey[as.character(PROPDMGEXP)]]
subsetdf[, CROPDMGEXP := cropDmgKey[as.character(CROPDMGEXP)]]

# Replace NA values with 10^0
subsetdf[is.na(PROPDMGEXP), PROPDMGEXP := 10^0]
subsetdf[is.na(CROPDMGEXP), CROPDMGEXP := 10^0]
head(subsetdf)
```

```
##     EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
##     <char>      <num>    <num>   <num>      <num>   <num>      <num>
## 1: TORNADO          0       15    25.0       1000       0          1
## 2: TORNADO          0        2    25.0       1000       0          1
## 3: TORNADO          0        2     2.5       1000       0          1
## 4: TORNADO          0        2     2.5       1000       0          1
## 5: TORNADO          0        6     2.5       1000       0          1
## 6: TORNADO          0        1     2.5       1000       0          1
```