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]
```

3.6 Adding Columns Representing Economic Cost to the Dataset

```
# Adding Columns Representing Economic Cost to the Dataset
subsetdf[, PropCost := PROPDMG * PROPDMGEXP]
subsetdf[, CropCost := CROPDMG * CROPDMGEXP]

# Replace NA values with 0 in economic cost columns
subsetdf[is.na(PropCost), PropCost := 0]
subsetdf[is.na(CropCost), CropCost := 0]
```

```
# Convert economic cost columns to numeric
subsetdf[, c("PropCost", "CropCost") := lapply(.SD, as.numeric),
    .SDcols = c("PropCost", "CropCost")]
```

3.7 Calculating Total Damage

```
totalCostDT <- subsetdf[, .(PropCost = sum(PropCost), CropCost = sum(CropCost),
    Total_Cost = sum(PropCost) + sum(CropCost)), by = .(EVTYPE)]
totalCostDT <- totalCostDT[order(-Total_Cost), ]
totalCostDT <- totalCostDT[1:10, ]
head(totalCostDT, 5)
```

```
##               EVTYPE    PropCost   CropCost  Total_Cost
##               <char>       <num>      <num>       <num>
## 1:           TORNADO 41885103105  141477000 42026580105
## 2: HURRICANE/TYPHOON 32747770000 2273120800 35020890800
## 3:             FLOOD  6632851640  127602500  6760454140
## 4:    TROPICAL STORM  6560156000  157265000  6717421000
## 5:      WINTER STORM  5221162500   10283000  5231445500
```

3.8 Calculating Total Fatalities and Injuries

```
totalInjuriesDT <- subsetdf[, .(FATALITIES = sum(FATALITIES),
    INJURIES = sum(INJURIES), totals = sum(FATALITIES) + sum(INJURIES)),
    by = .(EVTYPE)]

totalInjuriesDT <- totalInjuriesDT[order(-FATALITIES), ]

totalInjuriesDT <- totalInjuriesDT[1:10, ]

head(totalInjuriesDT, 5)
```

```
##            EVTYPE FATALITIES INJURIES totals
##            <char>      <num>    <num>  <num>
## 1:        TORNADO       5633    91346  96979
## 2: EXCESSIVE HEAT       1903     6525   8428
## 3:    FLASH FLOOD        978     1777   2755
## 4:           HEAT        937     2100   3037
## 5:      LIGHTNING        816     5230   6046
```

4. Results

4.1 Identifying Events most harmful to Population Health

Before making bar graph, we will melt data.table for easy transformation

```
bad_stuff <- melt(totalInjuriesDT, id.vars = "EVTYPE", variable.name = "bad_thing")
head(bad_stuff, 5)
```

```
##            EVTYPE  bad_thing value
##            <char>     <fctr> <num>
## 1:        TORNADO FATALITIES  5633
## 2: EXCESSIVE HEAT FATALITIES  1903
## 3:    FLASH FLOOD FATALITIES   978
```

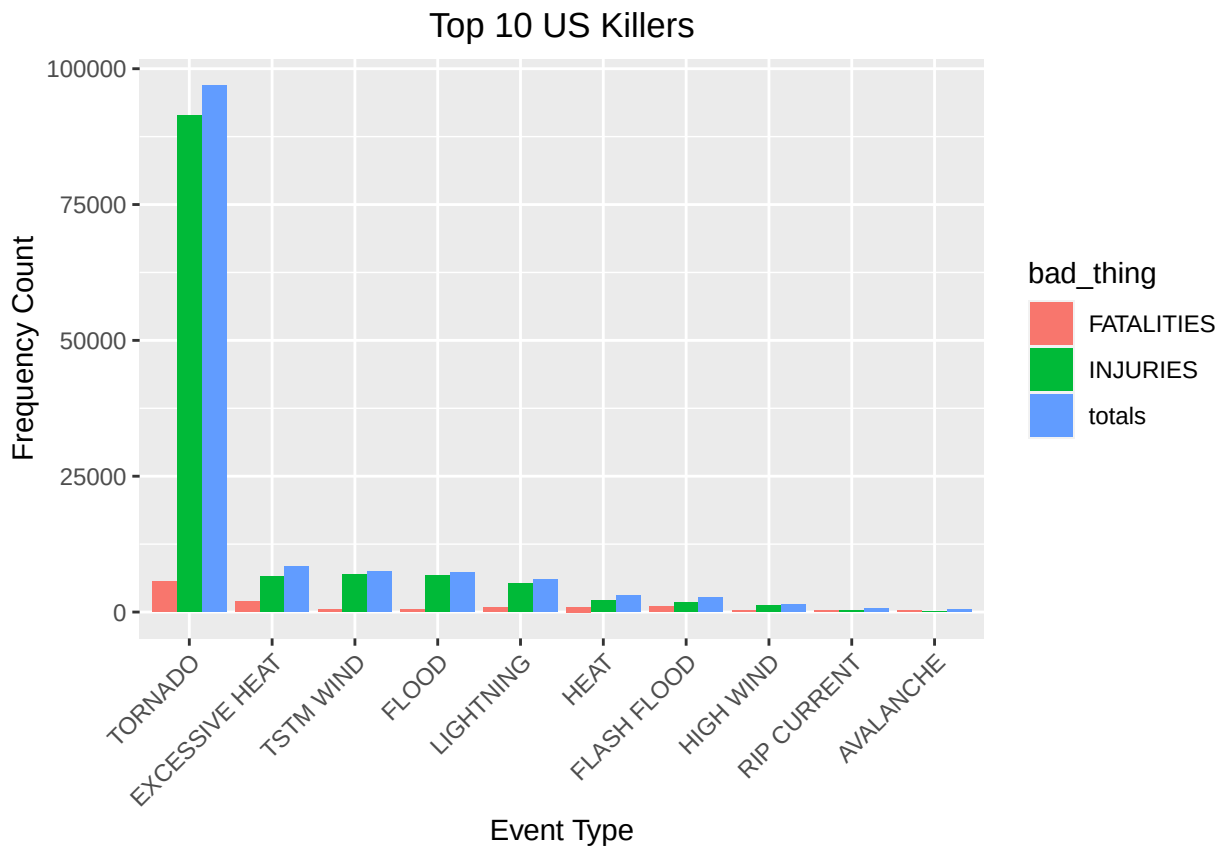
```
## 4:           HEAT FATALITIES   937
## 5:      LIGHTNING FATALITIES   816
```

Creating the Plot

```
library(ggplot2)
# Create a bar chart for top 10 US killers
healthChart <- ggplot(bad_stuff, aes(x = reorder(EVTYPE, -value),
    y = value))

# Plot data as bar chart
healthChart = healthChart + geom_bar(stat = "identity", aes(fill = bad_thing),
    position = "dodge")
# Format y-axis scale and set y-axis label
healthChart = healthChart + ylab("Frequency Count")
# Set x-axis label
healthChart = healthChart + xlab("Event Type")
# Rotate x-axis tick labels
healthChart = healthChart + theme(axis.text.x = element_text(angle = 45,
    hjust = 1))
# Set chart title and center it
healthChart = healthChart + ggtitle("Top 10 US Killers") + theme(plot.title = element_text(hjust = 0.5))

healthChart
```



So, Tornado is the most harmful event with a lot of injuries.

4.3 Identifying Events with Most Economic Impact

Before making bar graph, we will melt data.table for easy transformation

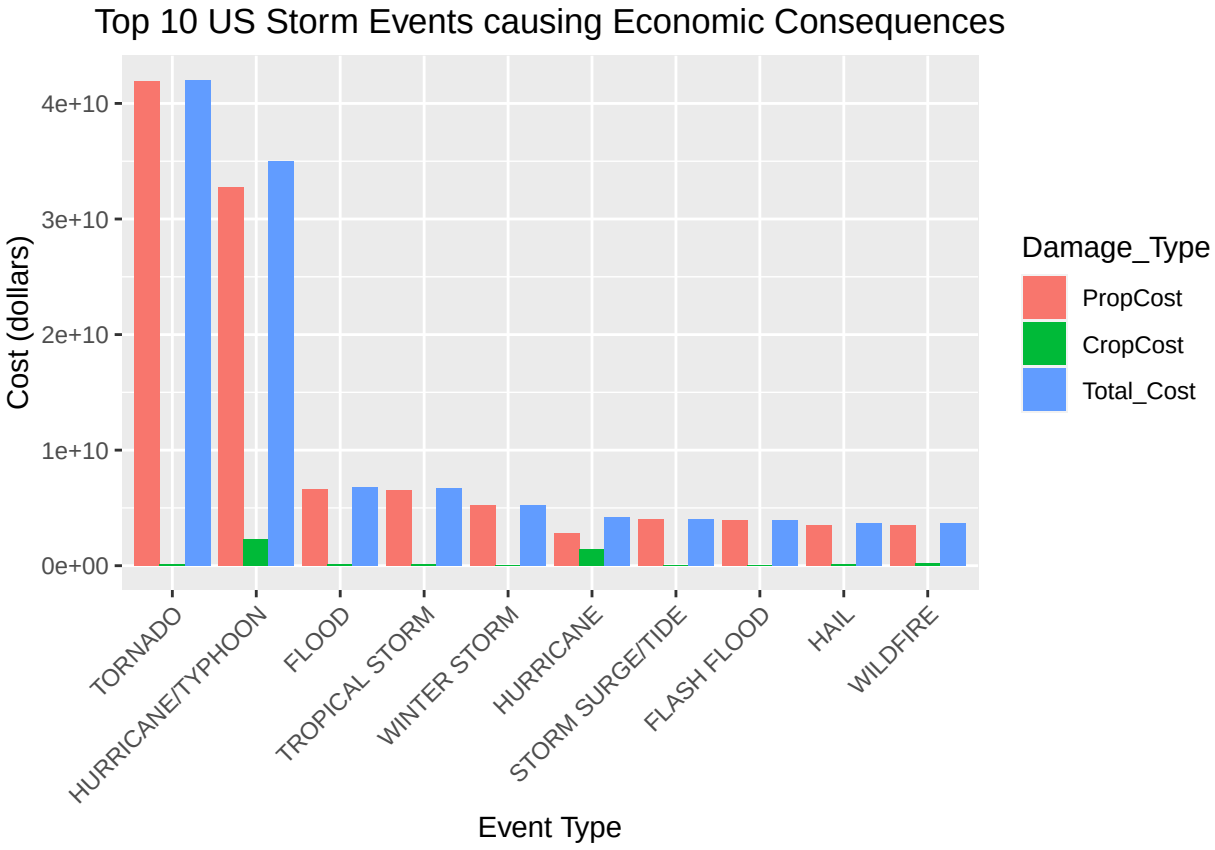
```
econ_consequences <- melt(totalCostDT, id.vars = "EVTYPE", variable.name = "Damage_Type")
head(econ_consequences, 5)
```

```
##                EVTYPE Damage_Type       value
##                <char>      <fctr>       <num>
## 1:            TORNADO    PropCost 41885103105
## 2: HURRICANE/TYPHOON    PropCost 32747770000
## 3:              FLOOD    PropCost  6632851640
## 4:    TROPICAL STORM    PropCost  6560156000
## 5:      WINTER STORM    PropCost  5221162500
```

Creating the Plot

```
# Create a bar chart for top 10 US storm events causing
# economic consequencs
econChart <- ggplot(econ_consequences, aes(x = reorder(EVTYPE,
    -value), y = value))

# Plot data as bar chart
econChart = econChart + geom_bar(stat = "identity", aes(fill = Damage_Type),
    position = "dodge")
# Format y-axis scale and set y-axis label
econChart = econChart + ylab("Cost (dollars)")
# Set x-axis label
econChart = econChart + xlab("Event Type")
# Rotate x-axis tick labels
econChart = econChart + theme(axis.text.x = element_text(angle = 45,
    hjust = 1))
# Set chart title and center it
econChart = econChart + ggtitle("Top 10 US Storm Events causing Economic Consequences") +
    theme(plot.title = element_text(hjust = 0.5))
econChart
```


Top 10 US Storm Events causing Economic Consequences
Cost (dollars)
4e+10
3e+10
2e+10
1e+10
0e+00
TORNADO
HURRICANE/TYPHOON
FLOOD
TROPICAL STORM
WINTER STORM
HURRICANE
STORM SURGE/TIDE
FLASH FLOOD
HAIL
WILDFIRE
Event Type
Damage_Type
PropCost
CropCost
Total_Cost