Impact of Severe Weather Events on Public Health and Economy in the United States

Ulises González

2026-02-27

```
knitr::opts_chunk$set(echo = TRUE, cache = TRUE, warning = FALSE, message = FALSE)
```

Synopsis

This analysis explores the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database to identify the most harmful severe weather events in terms of public health and economic impact. The data spans from 1950 to November 2011 and includes information on fatalities, injuries, property damage, and crop damage across the United States. After cleaning and processing the raw data, we found that **tornados** are by far the most harmful weather events to population health, causing the highest number of both fatalities and injuries. In terms of economic consequences, **floods** cause the greatest total damage when combining property and crop losses, followed by hurricanes/typhoons and storm surges. These findings can help government and municipal managers prioritize resources and preparedness efforts for the most impactful weather events.

Data Processing

Loading Required Libraries

```
library(dplyr)
library(ggplot2)
library(tidyr)
library(scales)
```

Downloading and Reading the Data

We start from the raw `.csv.bz2` file. The data is downloaded directly from the course website and read into R.

```
data_url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
data_file <- "StormData.csv.bz2"

if (!file.exists(data_file)) {
    download.file(data_url, destfile = data_file, method = "curl")
}

storm_data <- read.csv(data_file)
```

Exploring the Dataset

```
dim(storm_data)
```

```
## [1] 902297     37
```

```
str(storm_data, list.len = 10)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##   [list output truncated]
```

```
head(storm_data[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP
```

```
##     EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

Processing Health Impact Data

We aggregate fatalities and injuries by event type to determine which events are most harmful to population health.

```
health_data <- storm_data %>%
    group_by(EVTYPE) %>%
    summarise(
        fatalities = sum(FATALITIES, na.rm = TRUE),
        injuries = sum(INJURIES, na.rm = TRUE),
        total_health_impact = sum(FATALITIES, na.rm = TRUE) + sum(INJURIES, na.rm = TRUE)
    ) %>%
    arrange(desc(total_health_impact))

top_health <- head(health_data, 10)
top_health
```

```
## # A tibble: 10 x 4
##    EVTYPE          fatalities injuries total_health_impact
```

```
##    <chr>                  <dbl>  <dbl>          <dbl>
##  1 TORNADO                 5633  91346          96979
##  2 EXCESSIVE HEAT          1903   6525           8428
##  3 TSTM WIND                504   6957           7461
##  4 FLOOD                    470   6789           7259
##  5 LIGHTNING                816   5230           6046
##  6 HEAT                     937   2100           3037
##  7 FLASH FLOOD              978   1777           2755
##  8 ICE STORM                 89   1975           2064
##  9 THUNDERSTORM WIND        133   1488           1621
## 10 WINTER STORM             206   1321           1527
```

Processing Economic Impact Data

Property and crop damage values use exponent codes (`PROPDMGEXP` and `CROPDMGEXP`) to indicate magnitude. We convert these codes to numeric multipliers and compute actual dollar amounts.

```r
convert_exp <- function(exp) {
    exp <- toupper(as.character(exp))
    case_when(
        exp == "H" ~ 1e2,
        exp == "K" ~ 1e3,
        exp == "M" ~ 1e6,
        exp == "B" ~ 1e9,
        exp %in% as.character(0:9) ~ 10^as.numeric(exp),
        TRUE ~ 1
    )
}

economic_data <- storm_data %>%
    mutate(
        prop_multiplier = convert_exp(PROPDMGEXP),
        crop_multiplier = convert_exp(CROPDMGEXP),
        property_damage = PROPDMG * prop_multiplier,
        crop_damage = CROPDMG * crop_multiplier,
        total_damage = property_damage + crop_damage
    )

economic_summary <- economic_data %>%
    group_by(EVTYPE) %>%
    summarise(
        property_damage = sum(property_damage, na.rm = TRUE),
        crop_damage = sum(crop_damage, na.rm = TRUE),
        total_damage = sum(total_damage, na.rm = TRUE)
    ) %>%
    arrange(desc(total_damage))

top_economic <- head(economic_summary, 10)
top_economic
```

```
## # A tibble: 10 x 4
##    EVTYPE          property_damage crop_damage  total_damage
##    <chr>                     <dbl>       <dbl>         <dbl>
```

```
## 1 FLOOD               144657709807   5661968450 150319678257
## 2 HURRICANE/TYPHOON    69305840000   2607872800  71913712800
## 3 TORNADO              56947380676.   414953270  57362333946.
## 4 STORM SURGE          43323536000         5000  43323541000
## 5 HAIL                 15735267513.  3025954473  18761221986.
## 6 FLASH FLOOD          16822673978.  1421317100  18243991078.
## 7 DROUGHT               1046106000  13972566000  15018672000
## 8 HURRICANE            11868319010   2741910000  14610229010
## 9 RIVER FLOOD           5118945500   5029459000  10148404500
## 10 ICE STORM            3944927860   5022113500   8967041360
```

We verify the exponent codes present in the data:

```r
table(storm_data$PROPDMGEXP)
```

```
##
##                  -       ?       +       0       1       2       3       4       5       6
## 465934           1       8       5     216      25      13       4       4      28       4
##      7           8       B       h       H       K       m       M
##      5           1      40       1       6  424665       7   11330
```

```r
table(storm_data$CROPDMGEXP)
```

```
##
##                  ?       0       2       B       k       K       m       M
## 618413           7      19       1       9      21  281832       1    1994
```

Results

Question 1: Events Most Harmful to Population Health

The following figure shows the top 10 weather event types by their impact on population health, including both fatalities and injuries.

```r
# Prepare data for side-by-side bar plot
health_long <- top_health %>%
    select(EVTYPE, fatalities, injuries) %>%
    pivot_longer(cols = c(fatalities, injuries), names_to = "type", values_to = "count") %>%
    mutate(
        EVTYPE = factor(EVTYPE, levels = rev(top_health$EVTYPE)),
        type = factor(type, levels = c("fatalities", "injuries"),
                      labels = c("Fatalities", "Injuries"))
    )

p1 <- ggplot(health_long, aes(x = EVTYPE, y = count, fill = type)) +
    geom_bar(stat = "identity", position = "dodge") +
    coord_flip() +
    scale_fill_manual(values = c("Fatalities" = "#D32F2F", "Injuries" = "#FF8F00")) +
    scale_y_continuous(labels = comma) +
    labs(
        title = "A) Fatalities and Injuries by Weather Event Type",
```

```
            x = NULL,
            y = "Count",
            fill = "Impact Type"
        ) +
        theme_minimal(base_size = 12) +
        theme(legend.position = "top")

# Total health impact
total_health_plot <- top_health %>%
    mutate(EVTYPE = factor(EVTYPE, levels = rev(EVTYPE)))

p2 <- ggplot(total_health_plot, aes(x = EVTYPE, y = total_health_impact)) +
    geom_bar(stat = "identity", fill = "#C62828") +
    coord_flip() +
    scale_y_continuous(labels = comma) +
    labs(
        title = "B) Total Health Impact (Fatalities + Injuries)",
        x = NULL,
        y = "Total Count"
    ) +
    theme_minimal(base_size = 12)

gridExtra::grid.arrange(p1, p2, ncol = 1)
```

Key finding: Tornados are overwhelmingly the most harmful weather event type to population health,
causing **5,633** fatalities and **91,346** injuries over the period of record.

Question 2: Events with Greatest Economic Consequences

The following figure shows the top 10 weather event types by their economic impact, split into property
damage and crop damage.

```
# Prepare data for stacked bar plot
econ_long <- top_economic %>%
    select(EVTYPE, property_damage, crop_damage) %>%
    pivot_longer(cols = c(property_damage, crop_damage), names_to = "type", values_to = "amount") %>%
    mutate(
        EVTYPE = factor(EVTYPE, levels = rev(top_economic$EVTYPE)),
        type = factor(type, levels = c("property_damage", "crop_damage"),
                      labels = c("Property Damage", "Crop Damage"))
    )

p3 <- ggplot(econ_long, aes(x = EVTYPE, y = amount / 1e9, fill = type)) +
    geom_bar(stat = "identity", position = "stack") +
    coord_flip() +
    scale_fill_manual(values = c("Property Damage" = "#1565C0", "Crop Damage" = "#2E7D32")) +
    scale_y_continuous(labels = dollar_format(suffix = "B")) +
    labs(
        title = "A) Property and Crop Damage by Weather Event Type",
        x = NULL,
        y = "Damage (Billions USD)",
        fill = "Damage Type"
    ) +
```

A) Fatalities and Injuries by Weather Event Type

Impact Type ■ Fatalities ■ Injuries

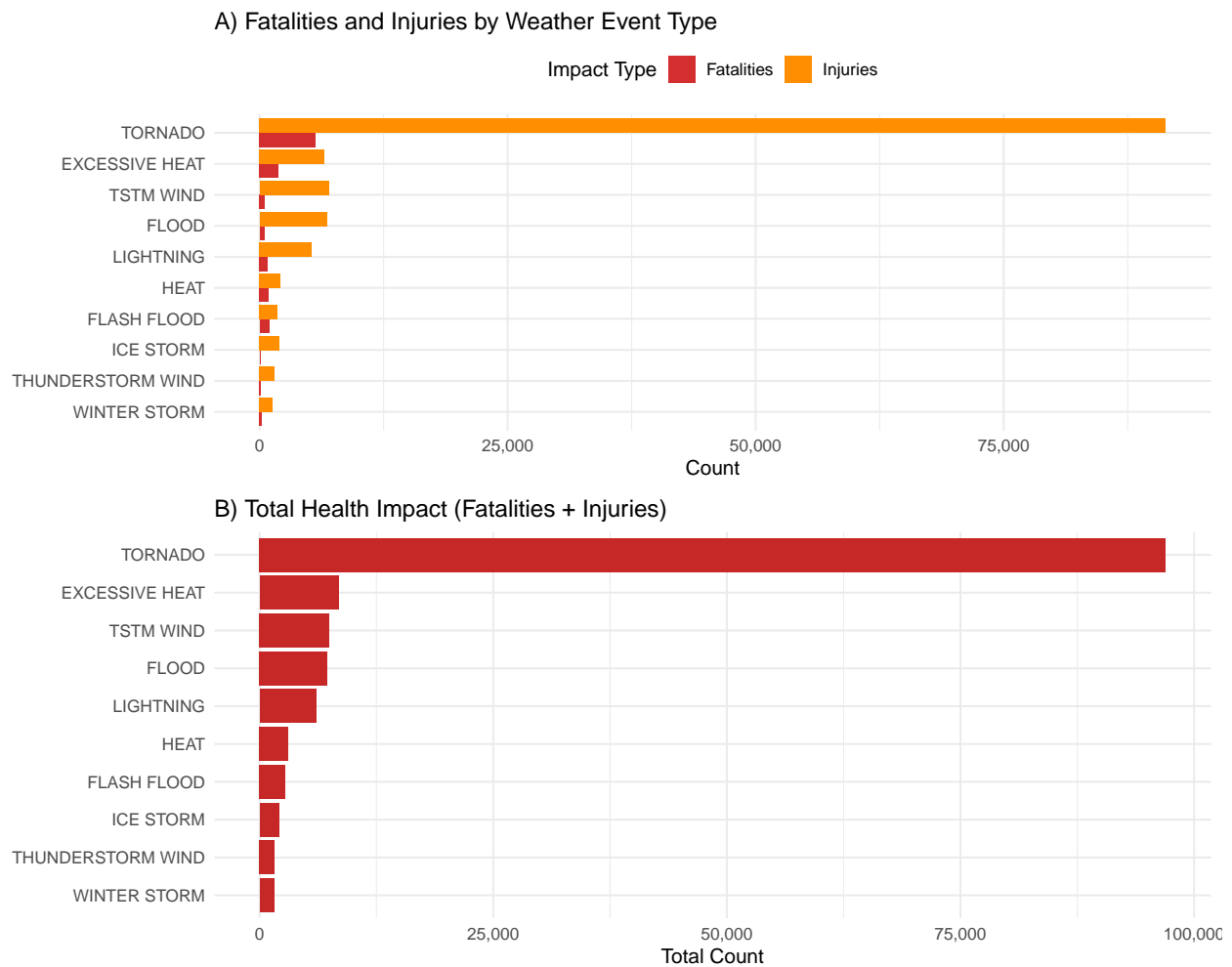B) Total Health Impact (Fatalities + Injuries)

Figure 1: Figure 1: Top 10 weather events most harmful to population health in the United States (1950–2011). Panel A shows fatalities and injuries side by side. Panel B shows the combined health impact. Tornados dominate both categories by a wide margin.

```
    theme_minimal(base_size = 12) +
    theme(legend.position = "top")

# Total economic damage
total_econ_plot <- top_economic %>%
    mutate(EVTYPE = factor(EVTYPE, levels = rev(EVTYPE)))

p4 <- ggplot(total_econ_plot, aes(x = EVTYPE, y = total_damage / 1e9)) +
    geom_bar(stat = "identity", fill = "#0D47A1") +
    coord_flip() +
    scale_y_continuous(labels = dollar_format(suffix = "B")) +
    labs(
        title = "B) Total Economic Damage (Property + Crop)",
        x = NULL,
        y = "Total Damage (Billions USD)"
    ) +
    theme_minimal(base_size = 12)

gridExtra::grid.arrange(p3, p4, ncol = 1)
```

Key finding: Floods cause the greatest total economic damage, with over **$150.3** billion in combined property and crop losses. Hurricanes/typhoons and storm surges follow as the next most costly event types.

Summary Table

```
top5_health <- head(health_data, 5) %>%
    mutate(rank = row_number()) %>%
    select(rank, event_health = EVTYPE, total_health_impact)

top5_econ <- head(economic_summary, 5) %>%
    mutate(rank = row_number()) %>%
    select(rank, event_economic = EVTYPE, total_damage)

# Create a comparison visualization
comparison <- bind_rows(
    head(health_data, 5) %>%
        mutate(
            category = "Health Impact",
            value = total_health_impact / max(total_health_impact),
            label = format(total_health_impact, big.mark = ","),
            rank = row_number()
        ) %>%
        select(EVTYPE, category, value, label, rank),
    head(economic_summary, 5) %>%
        mutate(
            category = "Economic Impact",
            value = total_damage / max(total_damage),
            label = paste0("$", round(total_damage / 1e9, 1), "B"),
            rank = row_number()
        ) %>%
        select(EVTYPE, category, value, label, rank)
```

A) Property and Crop Damage by Weather Event Type
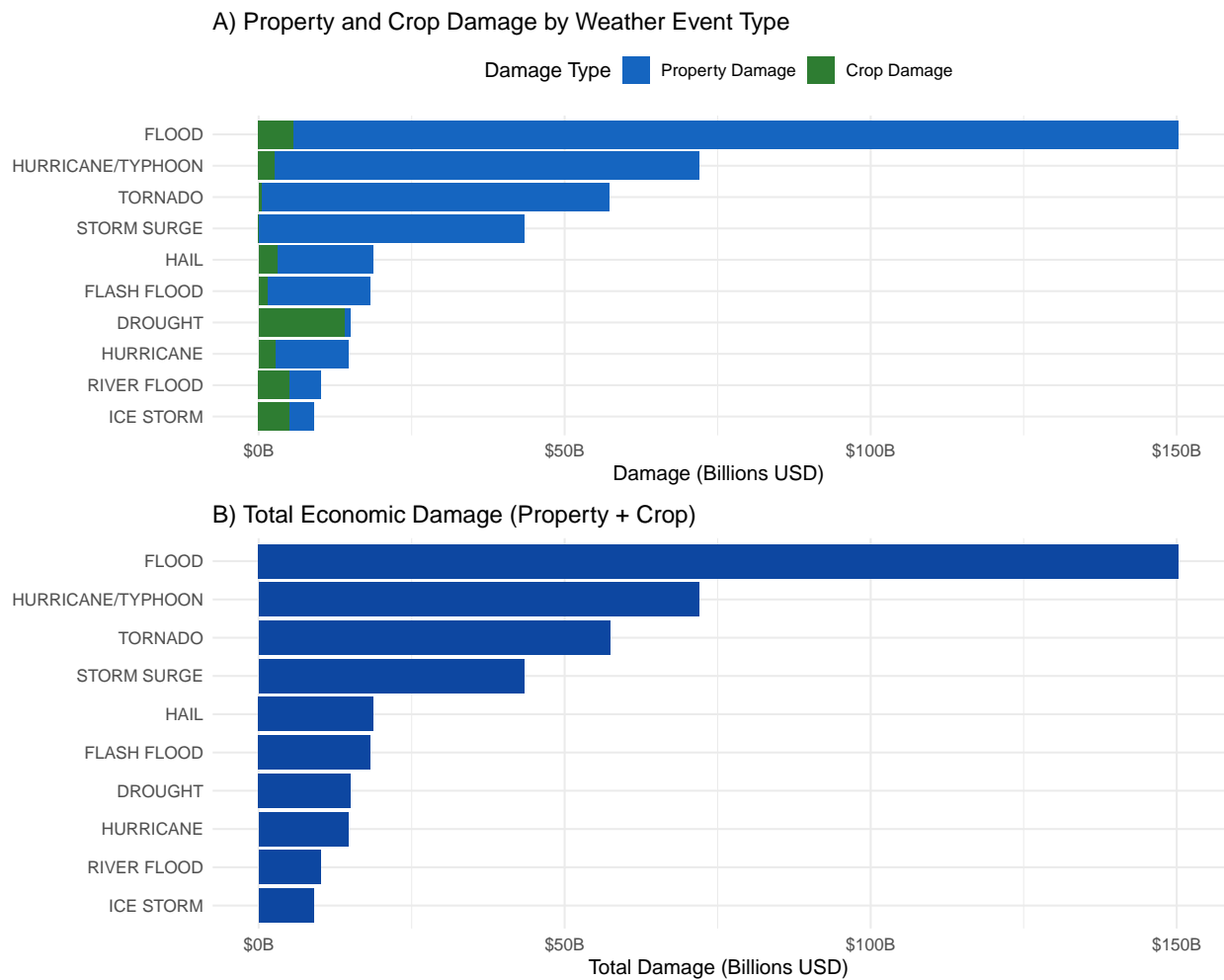
B) Total Economic Damage (Property + Crop)

Figure 2: Figure 2: Top 10 weather events with greatest economic consequences in the United States (1950–2011). Panel A shows property and crop damage separately. Panel B shows total economic damage. Floods cause the largest combined economic losses.

```
)

ggplot(comparison, aes(x = reorder(EVTYPE, -rank), y = value, fill = category)) +
    geom_bar(stat = "identity", position = "dodge", width = 0.7) +
    geom_text(aes(label = label), position = position_dodge(width = 0.7),
              hjust = -0.1, size = 3) +
    coord_flip() +
    scale_fill_manual(values = c("Health Impact" = "#C62828", "Economic Impact" = "#0D47A1")) +
    scale_y_continuous(limits = c(0, 1.3), labels = percent) +
    facet_wrap(~category, scales = "free") +
    labs(
        title = "Top 5 Events: Health vs Economic Impact (Normalized)",
        subtitle = "Different weather events dominate health vs economic consequences",
        x = NULL,
        y = "Relative Impact (% of maximum)"
    ) +
    theme_minimal(base_size = 12) +
    theme(legend.position = "none",
          strip.text = element_text(face = "bold", size = 13))
```
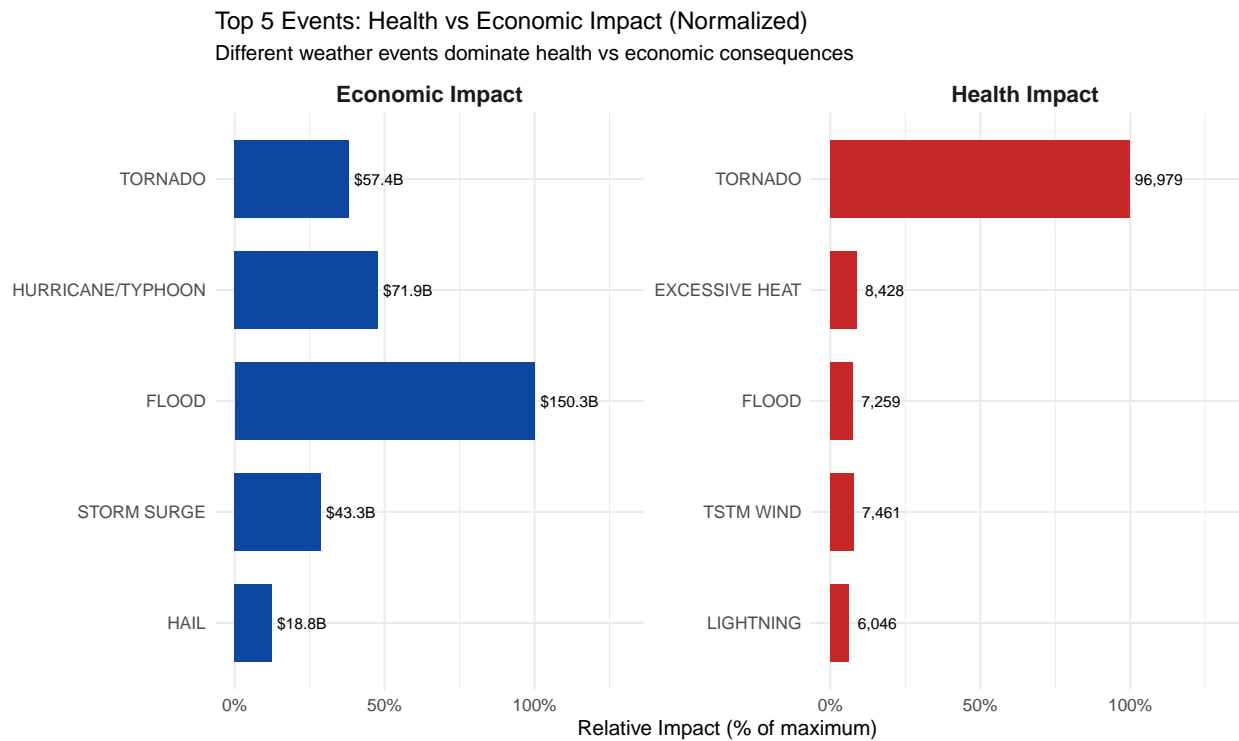

Figure 3: Figure 3: Comparison of the top 5 events for health impact vs economic impact, highlighting that different event types dominate each category.

Conclusions

1. **Population Health:** Tornados are the single most dangerous weather event, responsible for the most fatalities and injuries combined. Excessive heat ranks second for fatalities, while thunderstorm winds rank second for injuries.

2. **Economic Consequences:** Floods cause the greatest total economic damage, primarily through property destruction. Hurricanes/typhoons and storm surges are also major economic threats. Drought stands out as causing disproportionately high crop damage relative to property damage.

3. **Policy Implication:** Resources for public health preparedness should prioritize tornado warning systems and shelters, while economic preparedness should focus on flood mitigation infrastructure and hurricane resilience.