NOAA Storm Database analysis for U.S. healt and economic consequences due to severe weather events

Health and Economic Impacts of Severe Weather Events in the United States

Synopsis

This analysis explores the NOAA Storm Database (1950-2011) to identify which types of weather events are most harmful to population health and have the greatest economic consequences. The data reveals that tornadoes cause the highest number of fatalities and injuries nationwide. From an economic perspective, floods, hurricanes/typhoons, and drought have the most substantial impact on property and crop damage combined. This information can help emergency managers and municipal authorities effectively allocate resources for disaster preparedness and response.

Data Processing

```r
# Load necessary packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(tidyr)
library(knitr)

# Set options
knitr::opts_chunk$set(echo = TRUE, cache = TRUE, fig.width = 10, fig.height = 6)
```

Loading the Data

```r
# Download the file if it doesn't exist locally
if (!file.exists("StormData.csv.bz2")) {
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                "StormData.csv.bz2")
}

# Read the data
storm_data <- read.csv("StormData.csv.bz2")

# View the structure of the data
str(storm_data)
```

```
## '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  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Data Preparation

```r
# Convert event types to uppercase for consistency
storm_data$EVTYPE <- toupper(storm_data$EVTYPE)

# Create a subset with the columns relevant to our analysis
storm_subset <- storm_data %>%
  select(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

# Check the first few rows
head(storm_subset)
```

```
##    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
```

Handling Missing or Inconsistent Values

```r
# Check for missing values
colSums(is.na(storm_subset))
```

```
##     EVTYPE FATALITIES   INJURIES    PROPDMG PROPDMGEXP    CROPDMG CROPDMGEXP
##          0          0          0          0          0          0          0
```

```r
# Check unique values in damage exponent columns
unique(storm_subset$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```r
unique(storm_subset$CROPDMGEXP)
```

```
## [1] ""  "M" "K" "m" "B" "?" "0" "k" "2"
```

Converting Damage Values

```r
# Function to convert exponent symbols to actual multipliers
convert_exponent <- function(e) {
  if (e %in% c("K", "k")) return(1000)
  else if (e %in% c("M", "m")) return(1000000)
  else if (e %in% c("B", "b")) return(1000000000)
  else if (e %in% c("H", "h")) return(100)
  else if (e == "") return(1)
  else if (e %in% as.character(0:9)) return(10 ^ as.numeric(e))
  else return(0)
}

# Apply the conversion function
storm_subset <- storm_subset %>%
  mutate(
    PROP_DAMAGE = PROPDMG * sapply(PROPDMGEXP, convert_exponent),
    CROP_DAMAGE = CROPDMG * sapply(CROPDMGEXP, convert_exponent),
    TOTAL_DAMAGE = PROP_DAMAGE + CROP_DAMAGE
```

```
  )

# View the transformed data
head(storm_subset)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP PROP_DAMAGE
## 1 TORNADO          0       15    25.0          K       0                  25000
## 2 TORNADO          0        0     2.5          K       0                   2500
## 3 TORNADO          0        2    25.0          K       0                  25000
## 4 TORNADO          0        2     2.5          K       0                   2500
## 5 TORNADO          0        2     2.5          K       0                   2500
## 6 TORNADO          0        6     2.5          K       0                   2500
##   CROP_DAMAGE TOTAL_DAMAGE
## 1           0        25000
## 2           0         2500
## 3           0        25000
## 4           0         2500
## 5           0         2500
## 6           0         2500
```

Results

Question 1: Events Most Harmful to Population Health

To determine which types of events are most harmful to population health, we'll analyze fatalities and injuries.

```
# Aggregate fatalities and injuries by event type
health_impact <- storm_subset %>%
  group_by(EVTYPE) %>%
  summarize(
    TOTAL_FATALITIES = sum(FATALITIES, na.rm = TRUE),
    TOTAL_INJURIES = sum(INJURIES, na.rm = TRUE),
    TOTAL_CASUALTIES = TOTAL_FATALITIES + TOTAL_INJURIES
  ) %>%
  arrange(desc(TOTAL_CASUALTIES))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# Display top 10 most harmful events for health
top_10_health <- head(health_impact, 10)
kable(top_10_health, caption = "Top 10 Weather Events by Total Casualties")
```

Table 1: Top 10 Weather Events by Total Casualties

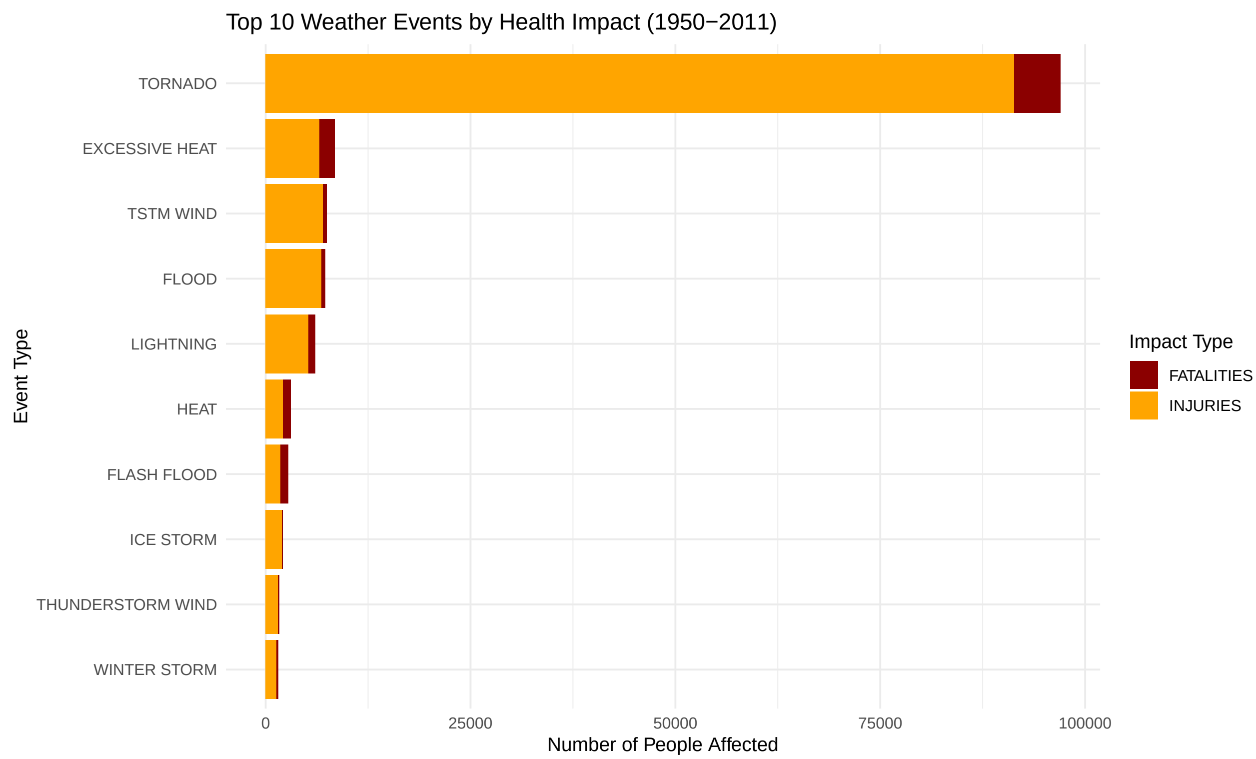
EVTYPE	TOTAL_FATALITIES	TOTAL_INJURIES	TOTAL_CASUALTIES
TORNADO	5633	91346	96979
EXCESSIVE HEAT	1903	6525	8428
TSTM WIND	504	6957	7461
FLOOD	470	6789	7259
LIGHTNING	816	5230	6046
HEAT	937	2100	3037
FLASH FLOOD	978	1777	2755
ICE STORM	89	1975	2064
THUNDERSTORM WIND	133	1488	1621
WINTER STORM	206	1321	1527

```r
# Create a combined plot for fatalities and injuries
top_10_events <- head(health_impact, 10)$EVTYPE

# Prepare data for visualization
health_plot_data <- storm_subset %>%
  filter(EVTYPE %in% top_10_events) %>%
  group_by(EVTYPE) %>%
  summarize(
    FATALITIES = sum(FATALITIES, na.rm = TRUE),
    INJURIES = sum(INJURIES, na.rm = TRUE)
  ) %>%
  pivot_longer(cols = c(FATALITIES, INJURIES),
               names_to = "IMPACT_TYPE",
               values_to = "COUNT")
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Create the plot
ggplot(health_plot_data, aes(x = reorder(EVTYPE, COUNT), y = COUNT, fill = IMPACT_TYPE)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  labs(
    title = "Top 10 Weather Events by Health Impact (1950-2011)",
    x = "Event Type",
    y = "Number of People Affected",
    fill = "Impact Type"
  ) +
  theme_minimal() +
  scale_fill_manual(values = c("FATALITIES" = "darkred", "INJURIES" = "orange"))
```



Based on the analysis, we can see that tornadoes have by far the most significant impact on population health, causing the highest number of both fatalities and injuries across the United States. Excessive heat, thunderstorm winds, and floods also rank high in terms of their impact on public health.

Question 2: Events with Greatest Economic Consequences

To determine which events have the greatest economic consequences, we'll analyze property damage and crop damage.

```r
# Aggregate property and crop damage by event type
economic_impact <- storm_subset %>%
  group_by(EVTYPE) %>%
  summarize(
    PROPERTY_DAMAGE = sum(PROP_DAMAGE, na.rm = TRUE),
    CROP_DAMAGE = sum(CROP_DAMAGE, na.rm = TRUE),
    TOTAL_DAMAGE = sum(TOTAL_DAMAGE, na.rm = TRUE)
  ) %>%
  arrange(desc(TOTAL_DAMAGE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Display top 10 most economically damaging events
top_10_economic <- head(economic_impact, 10)
kable(top_10_economic, caption = "Top 10 Weather Events by Total Economic Damage (USD)")
```

Table 2: Top 10 Weather Events by Total Economic Damage (USD)

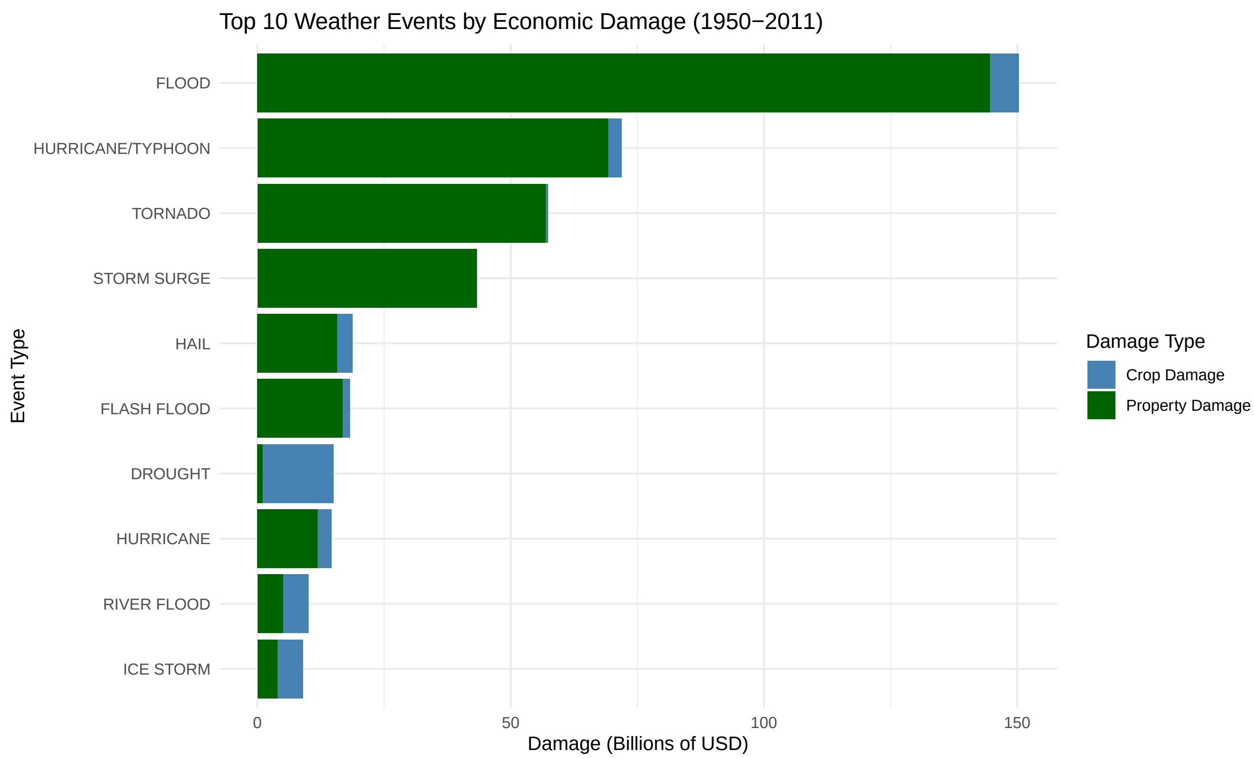
EVTYPE	PROPERTY_DAMAGE	CROP_DAMAGE	TOTAL_DAMAGE
FLOOD	144657709807	5661968450	150319678257
HURRICANE/TYPHOON	69305840000	2607872800	71913712800
TORNADO	56947380616	414953270	57362333886
STORM SURGE	43323536000	5000	43323541000
HAIL	15735267513	3025954473	18761221986
FLASH FLOOD	16822673978	1421317100	18243991078
DROUGHT	1046106000	13972566000	15018672000
HURRICANE	11868319010	2741910000	14610229010
RIVER FLOOD	5118945500	5029459000	10148404500
ICE STORM	3944927860	5022113500	8967041360

```r
# Format in billions for better readability
top_10_economic_billions <- top_10_economic %>%
  mutate(across(PROPERTY_DAMAGE:TOTAL_DAMAGE,
                ~round(. / 1e9, 2),
                .names = "{col}_BILLIONS"))

# Prepare data for visualization
economic_plot_data <- top_10_economic %>%
  select(EVTYPE, PROPERTY_DAMAGE, CROP_DAMAGE) %>%
  pivot_longer(cols = c(PROPERTY_DAMAGE, CROP_DAMAGE),
               names_to = "DAMAGE_TYPE",
               values_to = "AMOUNT")

# Create the plot
ggplot(economic_plot_data, aes(x = reorder(EVTYPE, AMOUNT), y = AMOUNT / 1e9, fill = DAMAGE_TYPE)) +
  geom_bar(stat = "identity", position = "stack") +
```

```r
  coord_flip() +
  labs(
    title = "Top 10 Weather Events by Economic Damage (1950-2011)",
    x = "Event Type",
    y = "Damage (Billions of USD)",
    fill = "Damage Type"
  ) +
  theme_minimal() +
  scale_fill_manual(values = c("PROPERTY_DAMAGE" = "darkgreen", "CROP_DAMAGE" = "steelblue"),
                    labels = c("Crop Damage", "Property Damage"))
```



The analysis shows that floods cause the greatest economic damage overall, followed by hurricanes/typhoons and tornadoes. When looking specifically at crop damage, drought has the most substantial impact, highlighting how different types of extreme weather events affect various sectors of the economy differently.

Comparative Analysis of Health and Economic Impacts

For a comprehensive understanding, let's examine the relationship between health impacts and economic consequences for the most severe event types.

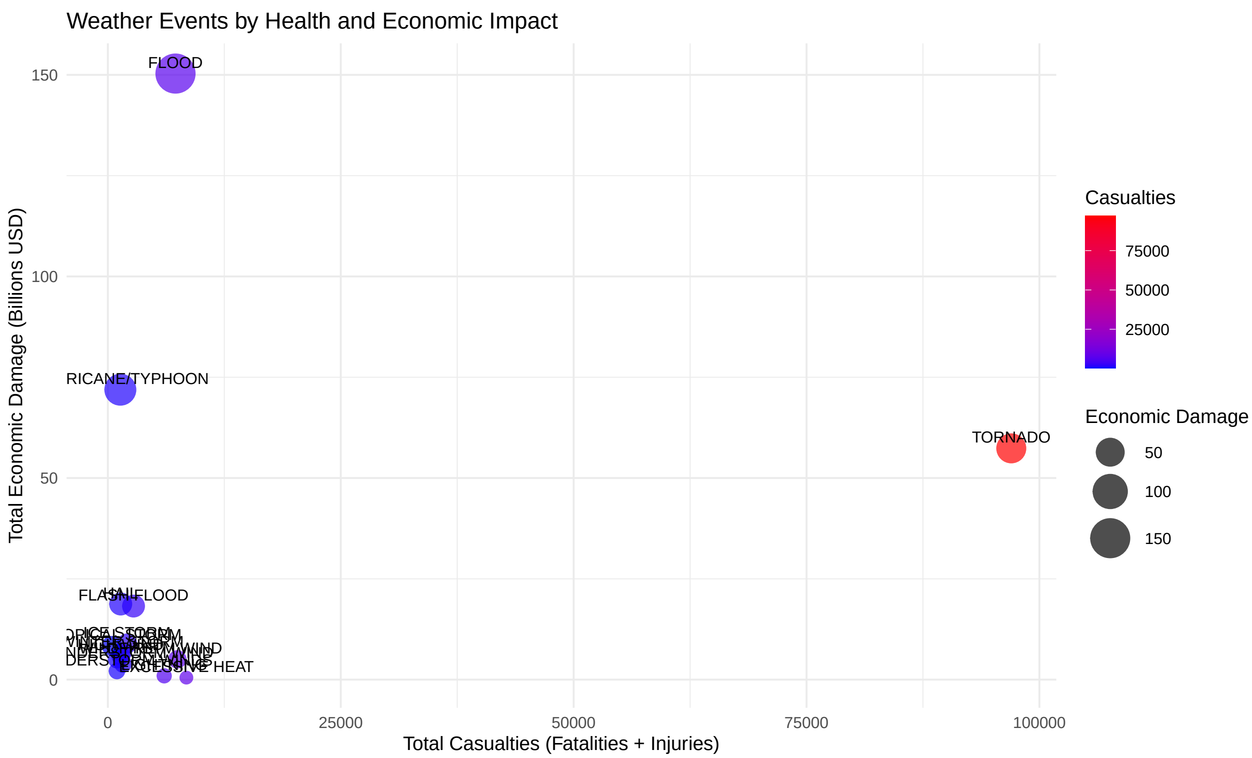
```r
# Join health and economic impact data
combined_impact <- inner_join(health_impact, economic_impact, by = "EVTYPE")

# Find top 15 events by combined ranking (casualties + damage)
combined_impact <- combined_impact %>%
  mutate(
    HEALTH_RANK = min_rank(desc(TOTAL_CASUALTIES)),
    ECONOMIC_RANK = min_rank(desc(TOTAL_DAMAGE)),
    COMBINED_RANK = HEALTH_RANK + ECONOMIC_RANK
  ) %>%
```

```
  arrange(COMBINED_RANK)

top_15_combined <- head(combined_impact, 15)

# Create a scatter plot
ggplot(top_15_combined, aes(x = TOTAL_CASUALTIES, y = TOTAL_DAMAGE / 1e9)) +
  geom_point(aes(size = TOTAL_DAMAGE / 1e9, color = TOTAL_CASUALTIES), alpha = 0.7) +
  geom_text(aes(label = EVTYPE), vjust = -0.5, hjust = 0.5, size = 3) +
  scale_size(range = c(3, 10)) +
  scale_color_gradient(low = "blue", high = "red") +
  labs(
    title = "Weather Events by Health and Economic Impact",
    x = "Total Casualties (Fatalities + Injuries)",
    y = "Total Economic Damage (Billions USD)",
    size = "Economic Damage",
    color = "Casualties"
  ) +
  theme_minimal() +
  theme(legend.position = "right")
```



Conclusion

Based on this analysis of the NOAA Storm Database from 1950 to 2011, we can draw the following conclusions:

1. **Population Health Impact**: Tornadoes are by far the most harmful weather events in terms of human casualties, causing significantly more fatalities and injuries than any other event type. Excessive heat and flash floods also have substantial impacts on public health.

2. **Economic Consequences**: Floods cause the greatest overall economic damage, with hurricanes/typhoons and tornadoes also resulting in massive property destruction. Drought stands out as

particularly damaging to agriculture, leading to the highest crop damage figures.

3. **Resource Allocation Implications**: Emergency managers and municipal authorities would be well-advised to allocate significant resources to tornado warning systems, flood prevention infrastructure, and drought management programs, as these events represent the greatest combined threat to both public health and economic stability.

This analysis demonstrates the importance of targeted preparedness efforts that address the specific patterns of damage associated with different types of severe weather events.