NOAA Storms (1950–2011): Health and Economic Impacts Across the U.S.

Synopsis

Based on the study "NOAA Storms (1950–2011): Health and Economic Impacts Across the U.S.", I have analysed, which types of events (as indicated in the EVTYPEEVTYPE) are most harmful with respect to population health across the U.S.

Based on the same study, I have also analysed, which types of events have the greatest economic consequences across the U.S.

The conclusion is that tornadoes are responsible for the most harm but the greatest economic consequences across the U.S. comes from floods.

Data Processing

Download the zip file, if needed

```
bz2_file <- "repdata_data_StormData.csv.bz2"

url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

 if (!file.exists(bz2_file)) {
   download.file(url, destfile = bz2_file, mode = "wb", quiet = TRUE)
 }
```

Load the data

```
storms <- read.csv(bz2_file, stringsAsFactors = FALSE)
```

Normalize the data, including,

1. Ensuring that EVTYPE is always in uppercase,
2. Convert exponent code
3. Calculate economic damage as PROP_DMG_USD + CROP_DMG_USD,
4. Calculate HEALTH_HARM as FATALITIES + INJURIES

```
library(dplyr)
library(stringr)

# Light EVTYPE normalization
storms <- storms %>%
  mutate(EVTYPE = toupper(str_trim(EVTYPE)))

# Helper to map exponent codes to multipliers
exp_to_mult <- function(e) {
  e <- toupper(str_trim(ifelse(is.na(e), "", e)))
  # Standard codes
  if (e %in% c("", NA)) return(1)
  if (e %in% c("H")) return(1e2)
  if (e %in% c("K")) return(1e3)
  if (e %in% c("M")) return(1e6)
  if (e %in% c("B")) return(1e9)
  # Digits interpreted as 10^digit (seen in legacy entries)
  if (grepl("^[0-8]$", e)) return(10^as.numeric(e))
  # Non-informative symbols → neutral multiplier
  if (e %in% c("+","-","?")) return(1)
  # Default fallback
  return(1)
}

# Vectorized mapping
prop_mult <- vapply(storms$PROPDMGEXP, exp_to_mult, numeric(1))
crop_mult <- vapply(storms$CROPDMGEXP, exp_to_mult, numeric(1))

storms <- storms %>%
  mutate(
    PROP_DMG_USD = as.numeric(PROPDMG) * prop_mult,
    CROP_DMG_USD = as.numeric(CROPDMG) * crop_mult,
    ECON_DMG_USD = PROP_DMG_USD + CROP_DMG_USD,
    HEALTH_HARM  = as.numeric(FATALITIES) + as.numeric(INJURIES)
  )
```

Generate health summary based on type of event (EVTYPE)

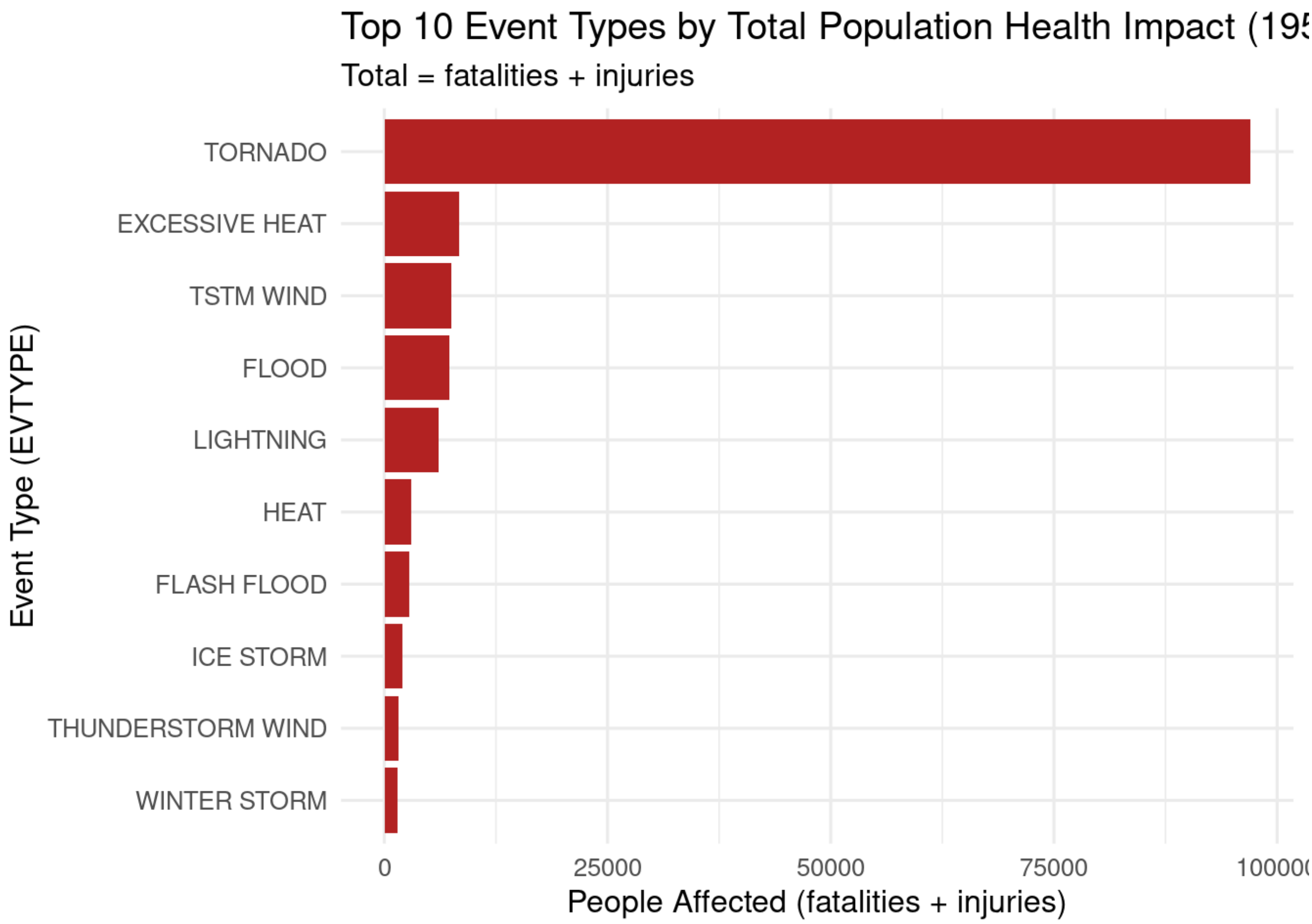
```
health_summary <- storms %>%
  group_by(EVTYPE) %>%
  summarize(
    fatalities = sum(FATALITIES, na.rm = TRUE),
    injuries   = sum(INJURIES,   na.rm = TRUE),
    total_harm = sum(HEALTH_HARM, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  arrange(desc(total_harm)) %>%
  slice_head(n = 10)

health_summary
```

```
## # A tibble: 10 × 4
##    EVTYPE            fatalities injuries total_harm
##    <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
```

```r
library(ggplot2)

ggplot(health_summary,
       aes(x = reorder(EVTYPE, total_harm), y = total_harm)) +
  geom_col(fill = "firebrick") +
  coord_flip() +
  labs(title = "Top 10 Event Types by Total Population Health Impact (1950-2011)",
       subtitle = "Total = fatalities + injuries",
       x = "Event Type (EVTYPE)",
       y = "People Affected (fatalities + injuries)") +
  theme_minimal(base_size = 12)
```



Calculate and display events based on greatest economic consequences

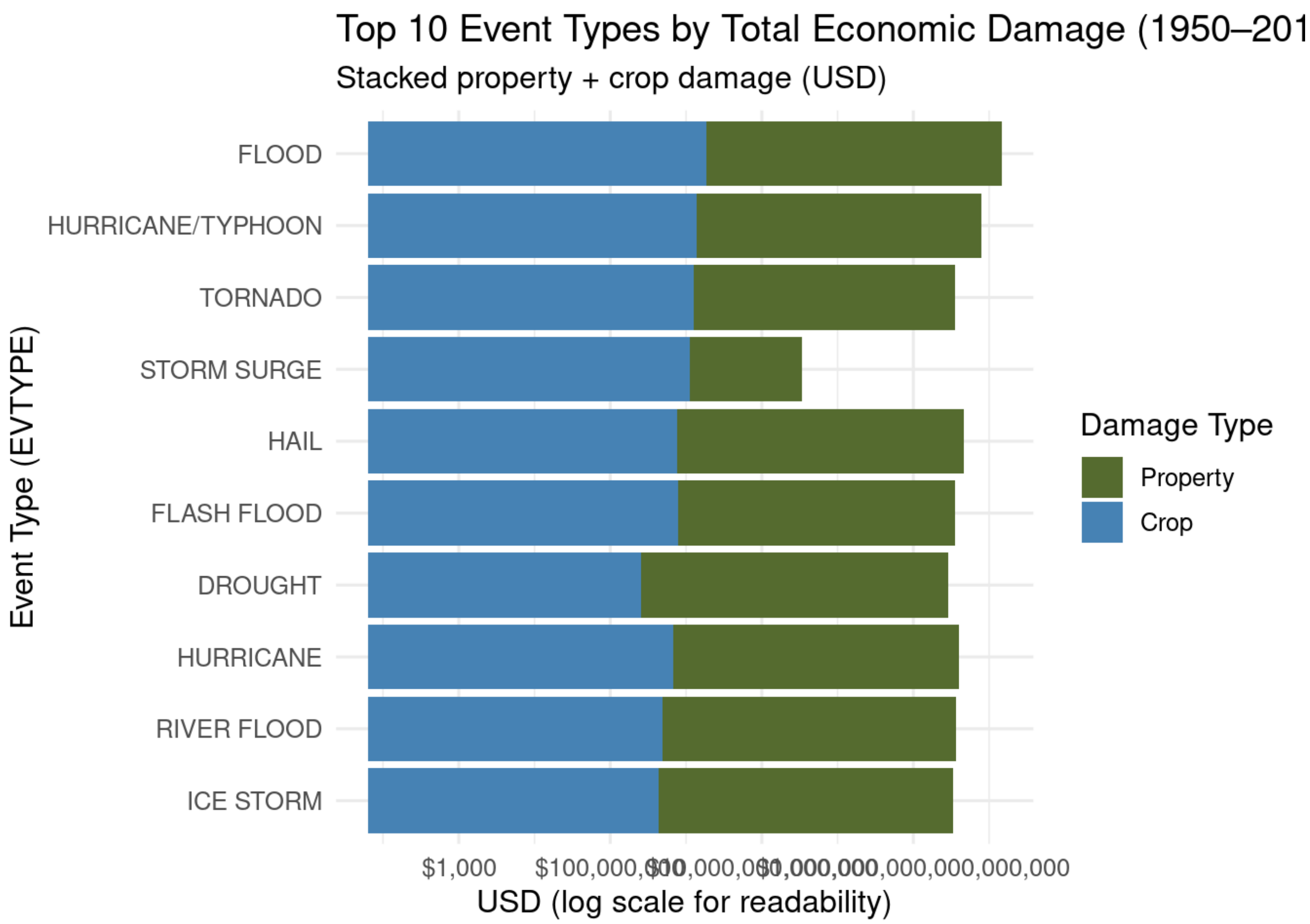
```r
econ_summary <- storms %>%
  group_by(EVTYPE) %>%
  summarize(
    prop_usd = sum(PROP_DMG_USD, na.rm = TRUE),
    crop_usd = sum(CROP_DMG_USD, na.rm = TRUE),
    total_usd = sum(ECON_DMG_USD, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  arrange(desc(total_usd)) %>%
  slice_head(n = 10)

# Present both total and component columns for transparency
econ_summary
```

```
## # A tibble: 10 × 4
##    EVTYPE                 prop_usd    crop_usd     total_usd
##    <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        16822723978.  1421317100  18244041078.
##  7 DROUGHT             1046106000  13972566000  15018672000
##  8 HURRICANE          11868319010   2741910000  14610229010
##  9 RIVER FLOOD         5118945500   5029459000  10148404500
## 10 ICE STORM           3944927860   5022113500   8967041360
```

```r
econ_long <- econ_summary %>%
  select(EVTYPE, prop_usd, crop_usd) %>%
  tidyr::pivot_longer(cols = c(prop_usd, crop_usd),
                      names_to = "component", values_to = "usd")

ggplot(econ_long,
       aes(x = reorder(EVTYPE, usd, FUN = sum), y = usd, fill = component)) +
  geom_col() +
  coord_flip() +
  scale_fill_manual(values = c("prop_usd" = "steelblue", "crop_usd" = "darkolivegre
en"),
                    labels = c("Property", "Crop"), name = "Damage Type") +
  labs(title = "Top 10 Event Types by Total Economic Damage (1950-2011)",
       subtitle = "Stacked property + crop damage (USD)",
       x = "Event Type (EVTYPE)",
       y = "USD (log scale for readability)") +
  scale_y_continuous(labels = scales::label_dollar(scale = 1),
                     trans = "log10") +
  theme_minimal(base_size = 12)
```



Results

The conclusion is that tornadoes are responsible for the most harm but the greatest economic consequences across the U.S. comes from floods.