

Tornadoes and floods cause most health and environmental damage in the United States

Yuki Joyama

2023-02-24

Data Processing

We imported csv file that includes data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. No changes were made to this cvs file prior to import.

```
# Loading and preprocessing the data
df <- read.csv("/Users/yukijoyama/Desktop/Documents/online course/coursera/Reproducible Research/wk4arc
```

```
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
```

```
library(ggplot2)
```

We aimed to investigate the type of weather events and its impact on population health and economic consequences. For health outcomes, a new variable “h_outcome” was created by adding the number of fatalities and injuries. For economic consequences, we created a numeric variable “e_outcome” that has information on the monetary damage amounts (property and/or crop) due to the weather events. We then extracted top 5 events most harmful to either health or environmental outcomes.

```
# create a dataset with health and economic outcomes
df1 <- df %>%
  mutate(
    h_outcome = FATALITIES + INJURIES,
    ep_outcome = case_when(
      PROPDMGEXP == "K" ~ PROPDMG * 1000,
      PROPDMGEXP == "M" ~ PROPDMG * 1000000,
      PROPDMGEXP == "B" ~ PROPDMG * 1000000000
```

```

    ), # property damage
    ec_outcome = case_when(
      CROPDMGEXP == "K" ~ CROPDMG * 1000,
      CROPDMGEXP == "M" ~ CROPDMG * 1000000,
      CROPDMGEXP == "B" ~ CROPDMG * 1000000000
    ) # crop damage
  ) %>%
  mutate(
    e_outcome = ep_outcome + ec_outcome # sum of property and crop damage
  ) %>%
  select(EVTYPE, h_outcome, e_outcome) %>% # select variables relevant to this report
  filter(!is.na(h_outcome)) %>% # remove rows with missing values
  filter(!is.na(e_outcome)) %>% # remove rows with missing values
  group_by(EVTYPE) %>%
  summarise(h_outcome = sum(h_outcome), e_outcome = sum(e_outcome)) # sum outcomes by each weather event

# top 10 events that are harmful to health
df1_h <- df1 %>%
  select(EVTYPE, h_outcome) %>%
  arrange(desc(h_outcome)) %>%
  head(10)

# Reorder the x-axis labels based on the sorted data frame
df1_h$EVTYPE <- factor(df1_h$EVTYPE, levels = df1_h$EVTYPE)

# top 10 events that are harmful to environment
df1_e <- df1 %>%
  select(EVTYPE, e_outcome) %>%
  arrange(desc(e_outcome)) %>%
  head(10)

# Reorder the x-axis labels based on the sorted data frame
df1_e$EVTYPE <- factor(df1_e$EVTYPE, levels = df1_e$EVTYPE)

```

Results

We plotted graphs that show top 10 weather events that resulted in the most harmful health and environmental consequences.

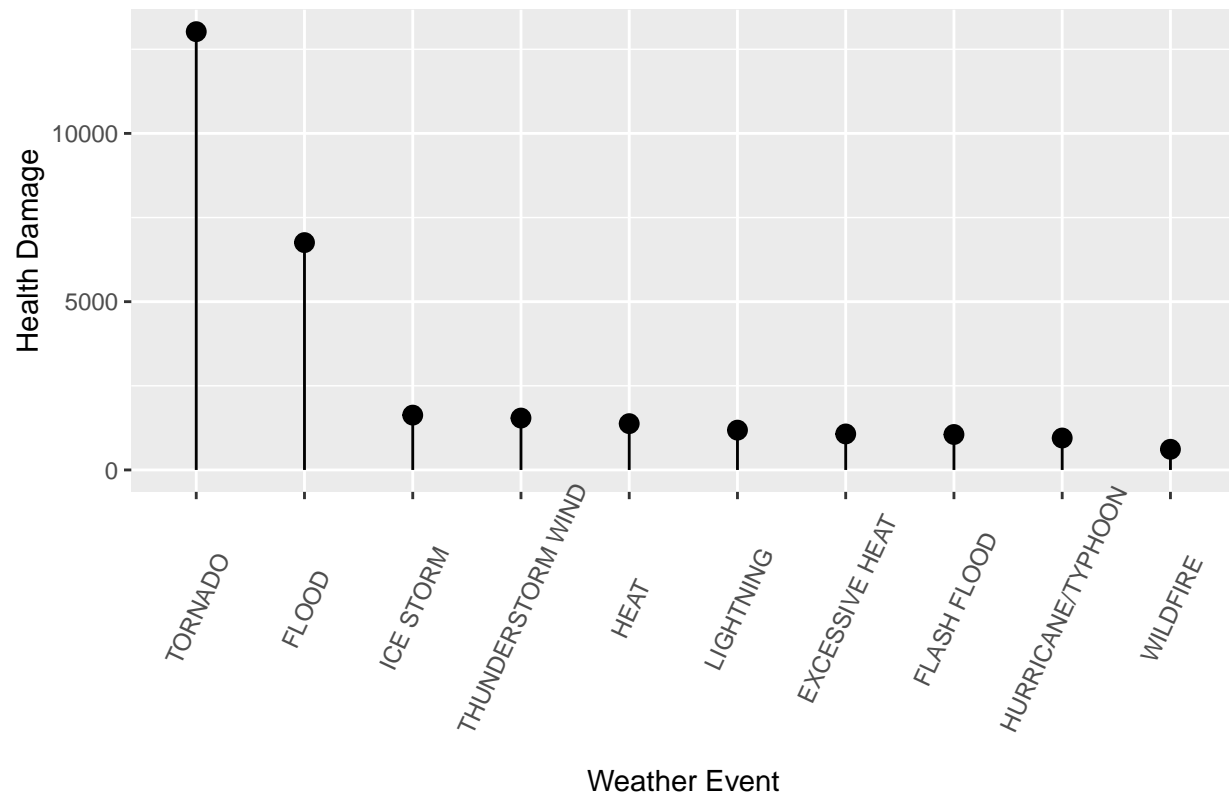
```

# library
library(ggplot2)

# Create the lollipop chart using the top 5 data
# health
ggplot(df1_h, aes(x=EVTYPE, y=h_outcome)) +
  geom_point(size=3) +
  geom_segment(aes(x=EVTYPE,
                  xend=EVTYPE,
                  y=0,
                  yend=h_outcome)) +
  labs(title="Top 10 Most Health Hazardous Weather Events", x="Weather Event", y="Health Damage") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

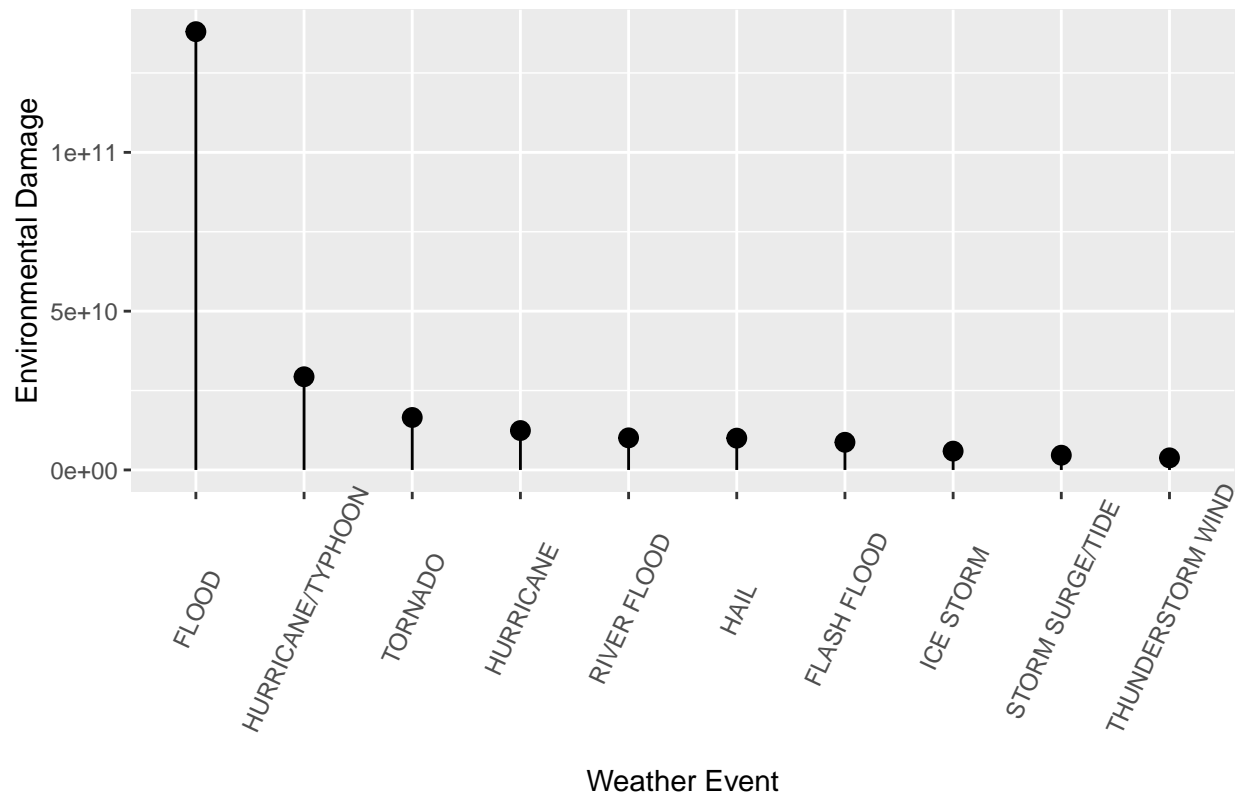
```

Top 10 Most Health Hazardous Weather Events



```
# environment
# Create the lollipop chart using the top 5 makes only
ggplot(df1_e, aes(x=EVTTYPE, y=e_outcome)) +
  geom_point(size=3) +
  geom_segment(aes(x=EVTTYPE,
                  xend=EVTTYPE,
                  y=0,
                  yend=e_outcome)) +
  labs(title="Top 10 Weather Events With the Most Environmental Damage", x="Weather Event", y="Environmental Damage")
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```

Top 10 Weather Events With the Most Environmental Damage



From these figures, we can conclude that tornadoes pose the greatest health hazard and floods the greatest environmental hazard.