Peer-graded Assignment: Course Project 2

Álvaro Ritoré

2024-02-11

Data analysis of the NOAA Storm Database

Synopsis

Storms and other severe weather events can cause problems for both public health and the economy of communities and municipalities. In many cases, these events result in fatalities, injuries, and property damage, making it a key concern to prevent such consequences as much as possible.

This project involves exploring the database of the National Oceanic and Atmospheric Administration (NOAA) of the United States. This database records the characteristics of major storms and weather events in the country, including when and where they occur, as well as estimates of associated fatalities, injuries, and property damage. It is an investigation that evaluates the most harmful weather events to public health, as well as those with the greatest economic consequences.

To conduct this analysis, RStudio has been used to compile this report in PDF format. The results are presented in a summarized manner through tables and summaries, providing a clear overview of the findings.

The aim of this analysis is to provide insights into the impact of severe weather events on both public health and the economy. Such insights can be valuable for policy-making and disaster preparedness efforts at the local, state, and national levels.

Environment

```
library(magrittr)
library(dplyr)
library(ggplot2)
library(scales)
library(ggplot2)
library(knitr)
```

Data Processing

Code for extracting the zip file and load the csv file into the environment

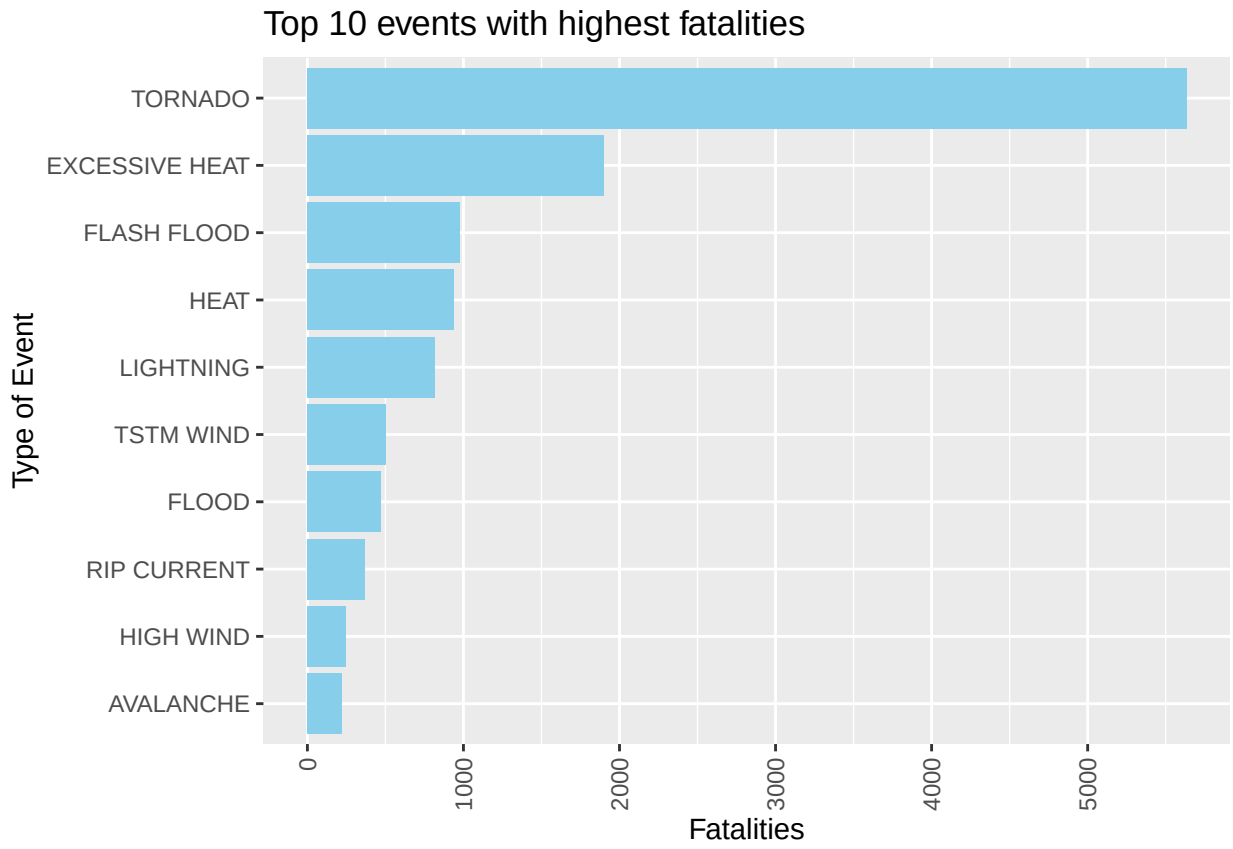
```
# Define the URL of the ZIP file
zip_link <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

# Name of the zip file
```

```
zip_file <- "repdata_data_StormData.csv"

# Name of the csv file
csv_file <- "repdata_data_StormData.csv"

# Check if the CSV file already exists in the environment
if (!exists("storm_data")) {
  # Download the ZIP file
  download.file(zip_link, zip_file)

  # Unzip the ZIP file
  unzip(zip_file)

  # Read the CSV file
  storm_data <- read.csv(csv_file)

  # Remove the ZIP file
  file.remove(zip_file)
}

# View the first few rows of the dataframe
head(storm_data)
```

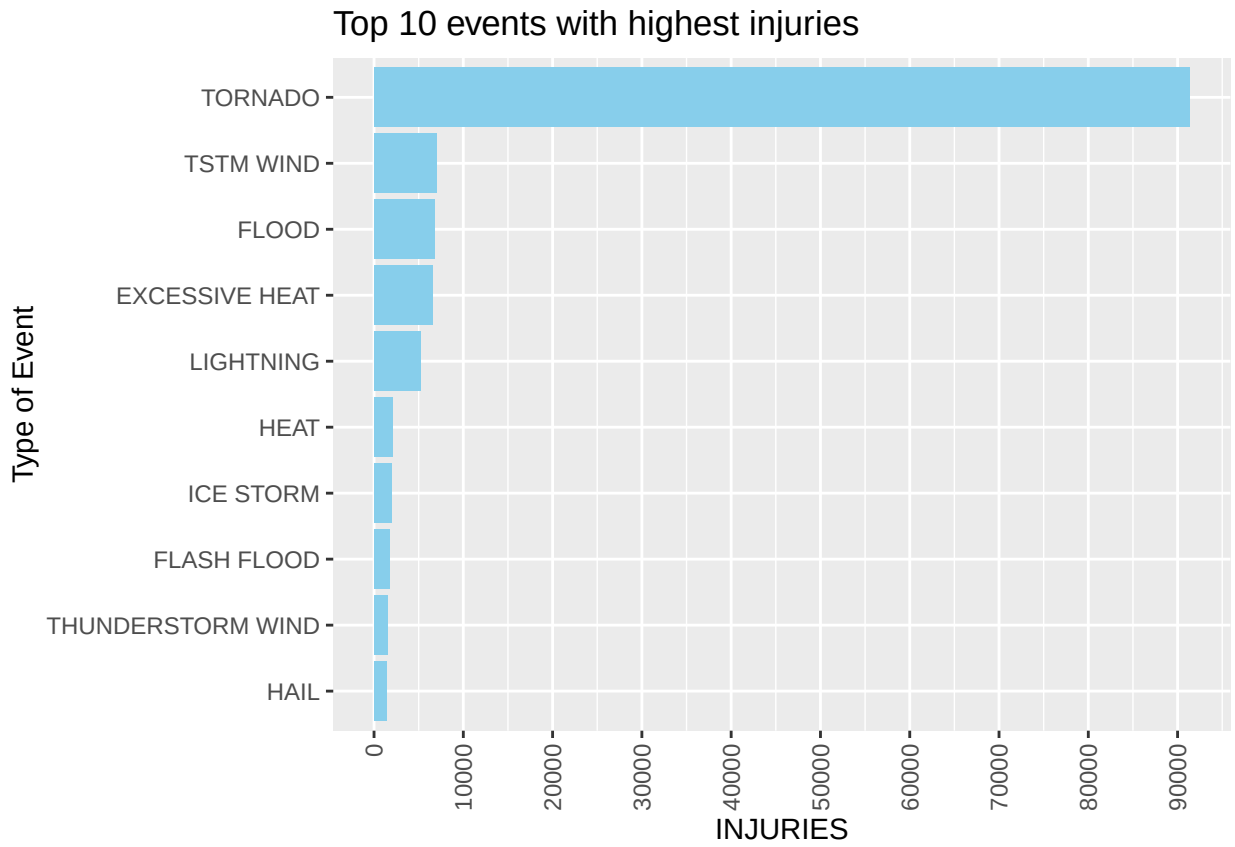
Data Analysis and Results

Effect in population Health

```
# Group event types by nº of fatalities and injuries
storm_data_health <- storm_data %>%
  mutate(EVTYPE=as.factor(EVTYPE)) %>%
  select(EVTYPE, FATALITIES, INJURIES) %>%
  group_by(EVTYPE) %>%
  summarize(
    FATALITIES=sum(FATALITIES),
    INJURIES=sum(INJURIES)
  ) %>%
  arrange(desc(FATALITIES), desc(INJURIES))

# Save top 10 event fatalities
top_events_fatalities <- head(arrange(storm_data_health, desc(FATALITIES)), 10)

# Plot top 10 event fatalities
ggplot(top_events_fatalities, aes(x=reorder(EVTYPE, FATALITIES), y = FATALITIES)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Type of Event", y = "Fatalities",
       title = "Top 10 events with highest fatalities") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  coord_flip() +
  scale_y_continuous(breaks = seq(0, max(top_events_fatalities$FATALITIES), by = 1000))
```



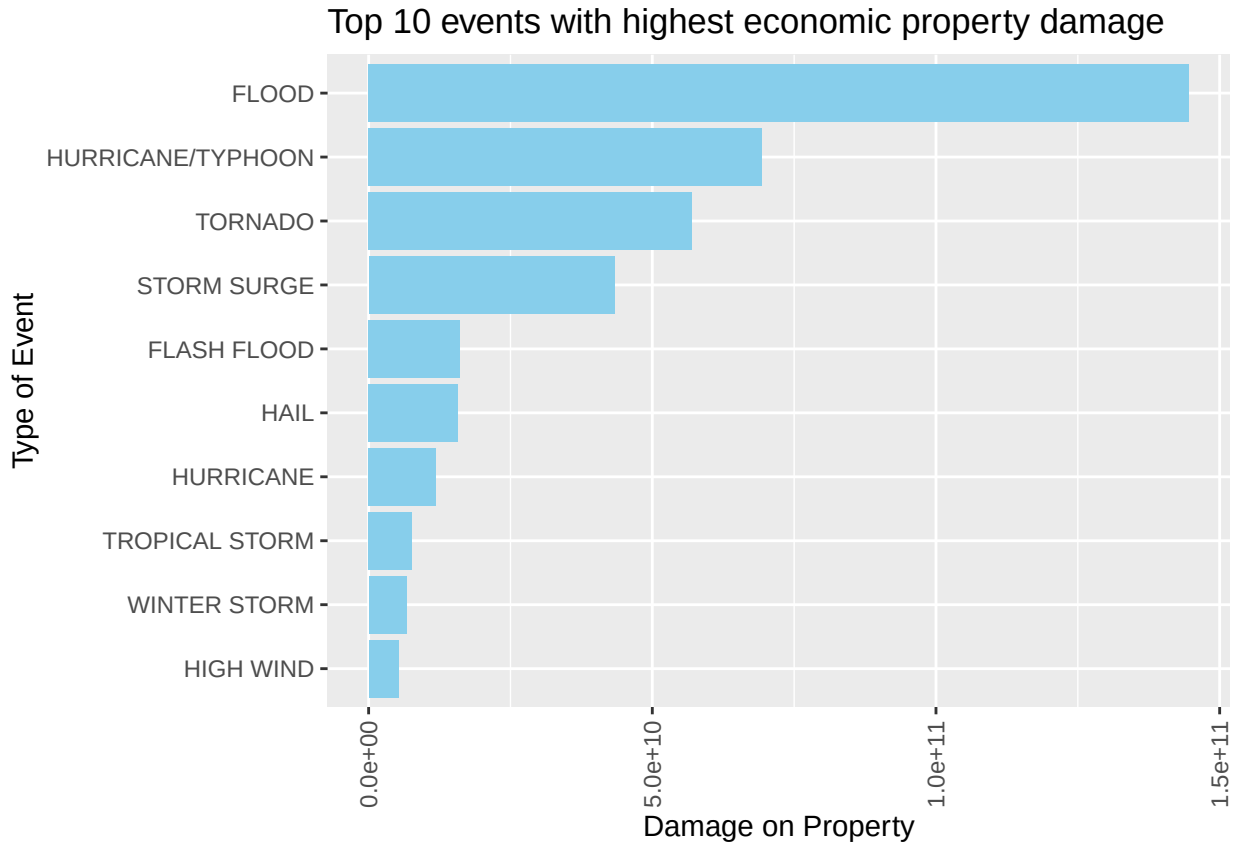

```
# Save top 10 event fatalities
top_events_injuries <- head(arrange(storm_data_health, desc(INJURIES)), 10)

# Plot top 10 event injuries
ggplot(top_events_injuries, aes(x=reorder(EVTYPE, INJURIES), y = INJURIES)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Type of Event", y = "INJURIES",
       title = "Top 10 events with highest injuries") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  coord_flip() +
  scale_y_continuous(breaks = seq(0, max(top_events_injuries$INJURIES), by = 10000))
```



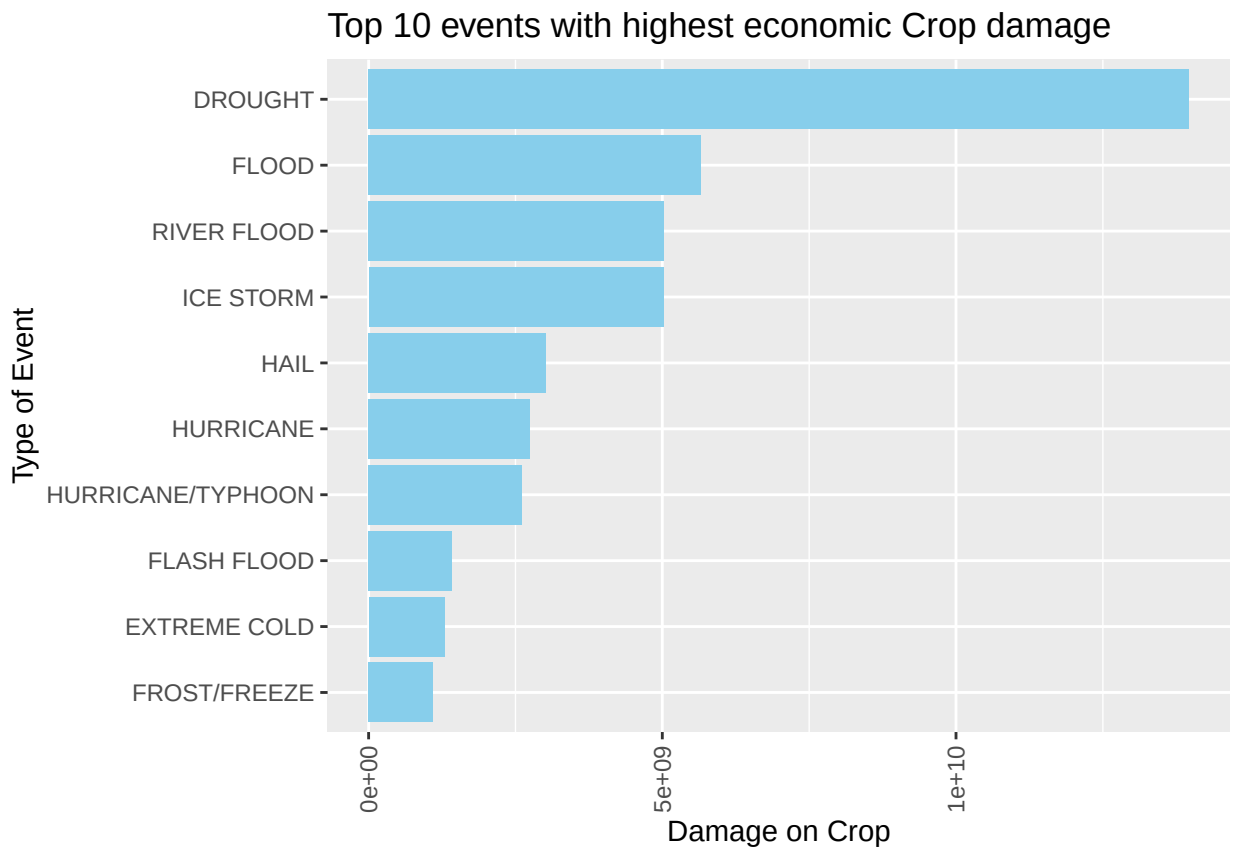
Economic effect of type of events

```
# Transform units of dollars for analysis
storm_data_economic <- storm_data %>%
  select(EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP) %>%
  mutate(PROPDMGEXP=case_when(PROPDMGEXP %in% c("H", "h")~2,
                              PROPDMGEXP=="K"~3,
                              PROPDMGEXP %in% c("M", "m")~6,
                              PROPDMGEXP=="B"~9,
                              TRUE ~ 0),
         PROPDMGEXP=as.numeric(PROPDMGEXP),
         CROPDMGEXP=case_when(CROPDMGEXP %in% c("K", "k")~3,
                              CROPDMGEXP %in% c("M", "m")~6,
                              CROPDMGEXP=="B"~9,
                              TRUE ~ 0),
         CROPDMGEXP=as.numeric(CROPDMGEXP)) %>%
  mutate(PROPDMG = PROPDMG * (10^PROPDMGEXP),
         CROPDMG = CROPDMG * (10^CROPDMGEXP)) %>%
  group_by(EVTYPE) %>%
  summarize(
    PROPDMG=sum(PROPDMG),
    CROPDMG=sum(CROPDMG)
  )
```

```r
# Saved top 10 event with the highest property damage
top_events_prop <- head(arrange(storm_data_economic, desc(PROPDMG)), 10)

# Plot top 10 event with the highest property damage
ggplot(top_events_prop, aes(x=reorder(EVTYPE, PROPDMG), y = PROPDMG)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Type of Event", y = "Damage on Property",
       title = "Top 10 events with highest economic property damage") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  coord_flip()
```




```r
# Saved top 10 event with the highest crop damage
top_events_crop <- head(arrange(storm_data_economic, desc(CROPDMG)), 10)

# Plot top 10 event with the highest crop damage
ggplot(top_events_crop, aes(x=reorder(EVTYPE, CROPDMG), y = CROPDMG)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Type of Event", y = "Damage on Crop",
       title = "Top 10 events with highest economic Crop damage") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  coord_flip()
```



Conclusions

```
cat(sprintf("The event with the highest fatalities is %s\n", top_events_fatalities$EVTYPE[which.max(top
```

```
## The event with the highest fatalities is TORNADO
```

```
cat(sprintf("The event with the highest injuries is %s\n", top_events_injuries$EVTYPE[which.max(top_eve
```

```
## The event with the highest injuries is TORNADO
```

```
cat(sprintf("The event with the highest economic property damage is %s\n", top_events_prop$EVTYPE[which
```

```
## The event with the highest economic property damage is FLOOD
```

```
cat(sprintf("The event with the highest economic crop damage is %s\n", top_events_crop$EVTYPE[which.max
```

```
## The event with the highest economic crop damage is DROUGHT
```