Storm Data Analysis for the United States from 1950 to 2011 based on the National Oceanic and Atmospheric Administration's (NOAA) storm database

AZATI YAKUFUJIANG

2023-07-01

Contents

1 Storm Data Analysis for the United States from 1950 to 2011 based on the National Oceanic and Atmospheric Administration's (NOAA) storm database 1
1.1 Synopsis . . . 1
1.2 Data Processing . . . 1
1.3 Results . . . 4
1.3.1 Which types of events are most harmful to population health? . . . 4
1.3.2 Which types of events have the greatest economic consequences? . . . 5

1 Storm Data Analysis for the United States from 1950 to 2011 based on the National Oceanic and Atmospheric Administration's (NOAA) storm database

1.1 Synopsis

This analysis explores the NOAA Storm Database and answers some basic questions about severe weather events. The database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete. The analysis addresses the question of which types of events are most harmful to population health and which types of events have the greatest economic consequences.

1.2 Data Processing

The data is processed with the following steps:

1. Read the data from the `repdata_data_StormData.csv.bz2` file into a data frame.

```r
# 1. Read the data from the `repdata_data_StormData.csv.bz2` file into a data frame.

storm_data <- read.csv("repdata_data_StormData.csv.bz2")
head(storm_data[, c("EVTYPE", "INJURIES", "FATALITIES", "PROPDMG", "CROPDMG")])
```

EVTYPE	INJURIES	FATALITIES	PROPDMG	CROPDMG
TORNADO	15	0	25.0	0

EVTYPE	INJURIES	FATALITIES	PROPDMG	CROPDMG
TORNADO	0	0	2.5	0
TORNADO	2	0	25.0	0
TORNADO	2	0	2.5	0
TORNADO	2	0	2.5	0
TORNADO	6	0	2.5	0

```r
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 ...
```

The data includes 902297 observations and 37 variables. The variables are described as follows:

- **STATE___**: State where the event occurred
- **BGN_DATE**: Date when the event began
- **BGN_TIME**: Time when the event began
- **TIME_ZONE**: Time zone (of the recorded time) for BGN_DATE

- **COUNTY**: County where the event occurred
- **COUNTYNAME**: County where the event occurred
- **STATE**: State where the event occurred
- **EVTYPE**: Type of event
- **BGN_RANGE**: Range of begin date
- **BGN_AZI**: 1-2 character directional indicator of the beginning of the range of azimuths
- **BGN_LOCATI**: Beginning location
- **END_DATE**: Date when the event ended
- **END_TIME**: Time when the event ended
- **COUNTY_END**: County where the event ended
- **COUNTYENDN**: County where the event ended
- **END_RANGE**: Range of end date
- **END_AZI**: 1-2 character directional indicator of the end of the range of azimuths
- **END_LOCATI**: Ending location
- **LENGTH**: Length of the event in hours
- **WIDTH**: Width of the event in miles
- **F**: F-scale (F0 to F5) of tornado
- **MAG**: Numeric value for magnitude of tornado
- **FATALITIES**: Number of fatalities
- **INJURIES**: Number of injuries
- **PROPDMG**: Property damage (in dollars)
- **PROPDMGEXP**: Exponent used for PROPDMG
- **CROPDMG**: Crop damage (in dollars)
- **CROPDMGEXP**: Exponent used for CROPDMG
- **WFO**: Weather Forecast Office
- **STATEOFFIC**: State Officer
- **ZONENAMES**: Zone name
- **LATITUDE**: Latitude (in decimal degrees)
- **LONGITUDE**: Longitude (in decimal degrees)
- **LATITUDE_E**: Latitude (in decimal degrees) for end location
- **LONGITUDE_**: Longitude (in decimal degrees) for end location
- **REMARKS**: Remarks
- **REFNUM**: Unique identifier assigned to each event

2. Converting data and time variables to the appropriate format.

```r
# 2. Converting Date variables to the appropriate format.

storm_data$BGN_DATE <- as.Date(storm_data$BGN_DATE, "%m/%d/%Y")
storm_data$END_DATE <- as.Date(storm_data$END_DATE, "%m/%d/%Y")
```

3. Transforming data in corresponding datatypes for the analysis.

```r
# 3. Transforming data in corresponding datatypes for the analysis.
storm_data$INJURIES <- as.numeric(storm_data$INJURIES)
storm_data$PROPDMG <- as.numeric(storm_data$PROPDMG)
storm_data$CROPDMG <- as.numeric(storm_data$CROPDMG)
```

4. Agregate new variables for the analysis.

```r
# 4. Agregate new variables for the analysis.

storm_data$TOTALDMG <- storm_data$PROPDMG + storm_data$CROPDMG
```

1.3 Results

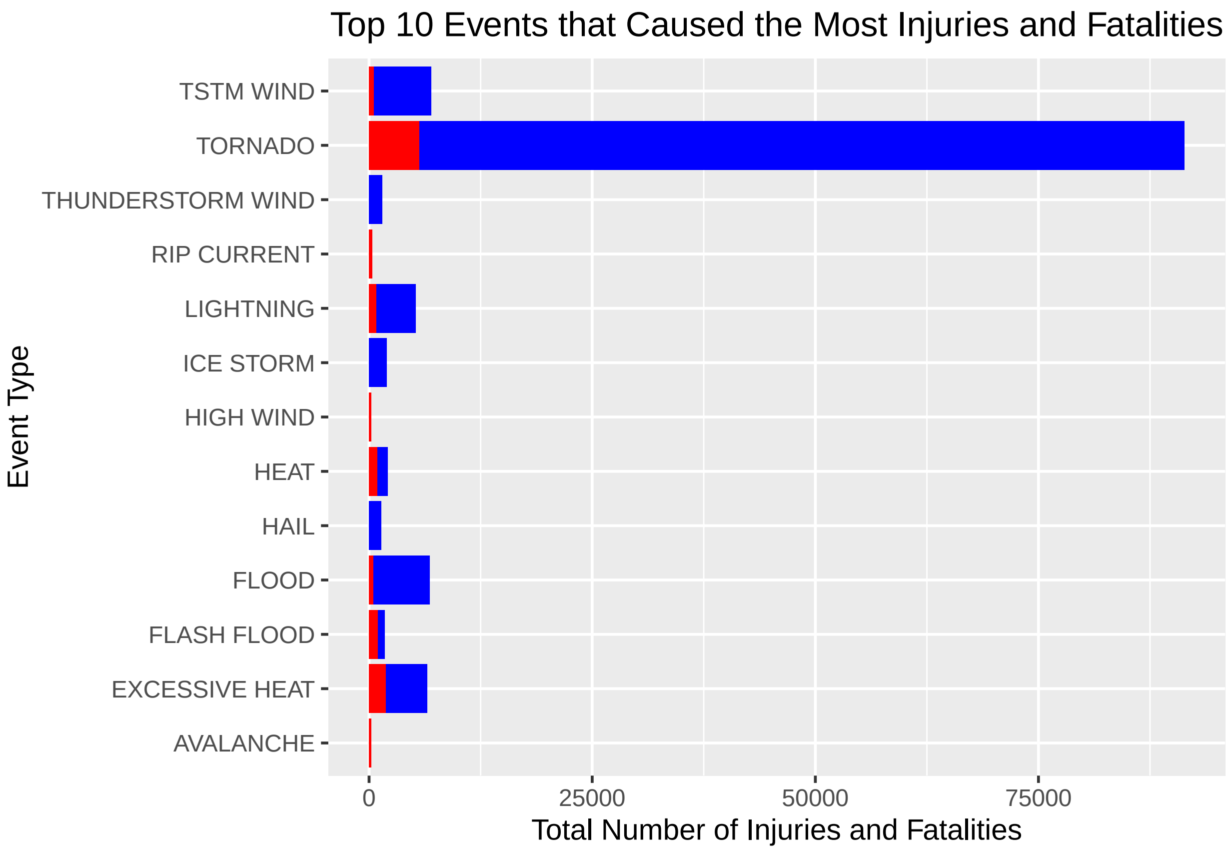
1.3.1 Which types of events are most harmful to population health?

The following plot shows the top 10 events that caused the most injuries and fatalities.

```
# Which types of events are most harmful to population health? (injuries and fatalities)

library(dplyr)
```

```
##
##          : 'dplyr'
##                  'package:stats':
##
##     filter, lag
##                  'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

# plot a histogram of the total number of injuries and fatalities by event type, two seperate bar one f

top10_injuries <- storm_data %>%
  group_by(EVTYPE) %>%
  summarise(total_injuries = sum(INJURIES)) %>%
  arrange(desc(total_injuries)) %>%
  head(10)

top10_fatalities <- storm_data %>%
    group_by(EVTYPE) %>%
    summarise(total_fatalities = sum(FATALITIES)) %>%
    arrange(desc(total_fatalities)) %>%
    head(10)

top10_injuries$EVTYPE <- factor(top10_injuries$EVTYPE, levels = top10_injuries$EVTYPE)
top10_fatalities$EVTYPE <- factor(top10_fatalities$EVTYPE, levels = top10_fatalities$EVTYPE)

ggplot() +
  geom_bar(data = top10_injuries, aes(x = EVTYPE, y = total_injuries), stat = "identity", fill = "blue")
  geom_bar(data = top10_fatalities, aes(x = EVTYPE, y = total_fatalities), stat = "identity", fill = "re
  coord_flip() +
  labs(title = "Top 10 Events that Caused the Most Injuries and Fatalities",
       x = "Event Type",
       y = "Total Number of Injuries and Fatalities") +
  theme(plot.title = element_text(hjust = 0.5))
```



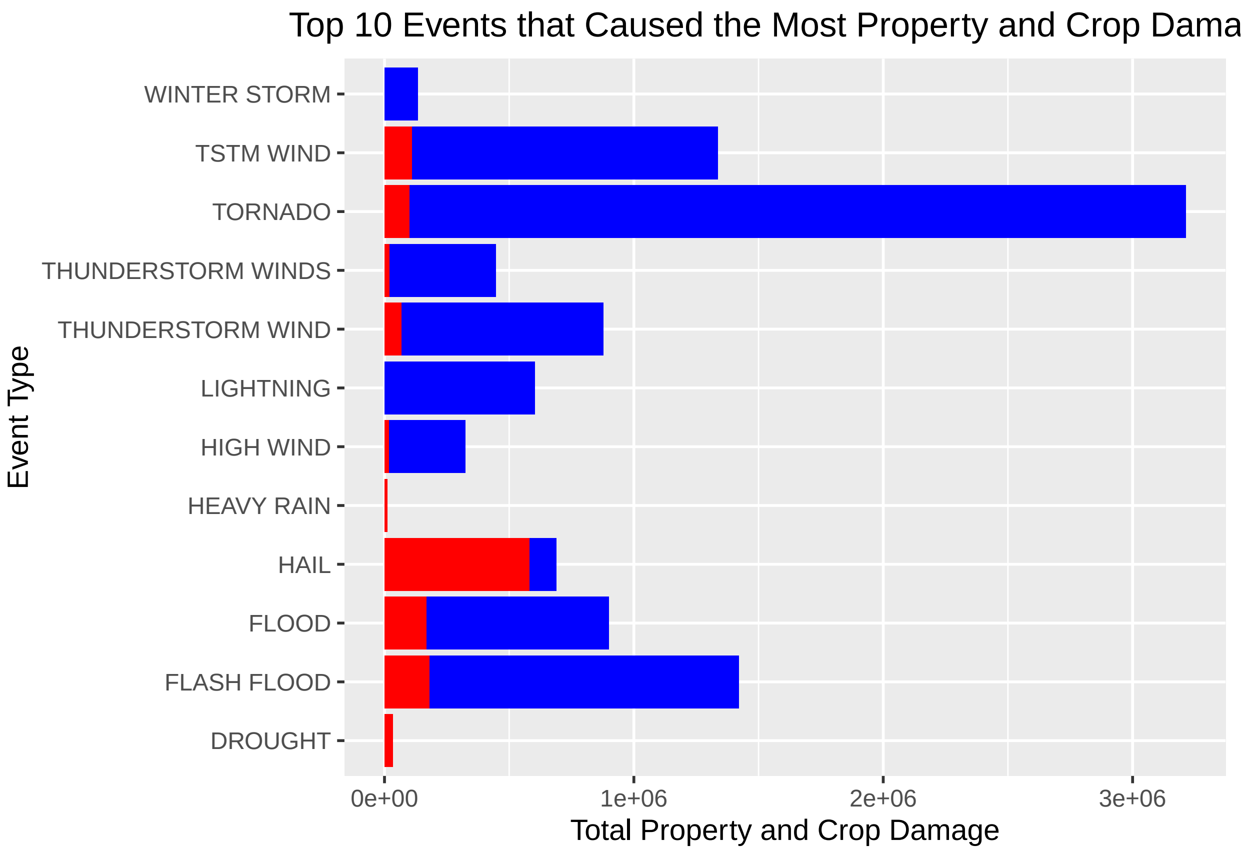
From the plot above, we can see that the top 10 events that caused the most injuries are Tornado, TSTM Wind, Hail, Flood, Thunderstorm Wind, Lightning, Heat, Ice Storm, Winter Storm, and Heavy Snow. The top 10 events that caused the most fatalities are Tornado, Excessive Heat, Flash Flood, Lightning, Heat, Flood, Rip Current, High Surf, TSTM Wind, and Astronomical Low Tide.

1.3.2 Which types of events have the greatest economic consequences?

The following plot shows the top 10 events that caused the most property and crop damage.

```
# Which types of events have the greatest economic consequences? (property and crop damage)

# plot a histogram of the total property and crop damage by event type, two seperate bar one for proper

top10_property <- storm_data %>%
  group_by(EVTYPE) %>%
  summarise(total_property = sum(PROPDMG)) %>%
  arrange(desc(total_property)) %>%
  head(10)

top10_crop <- storm_data %>%
    group_by(EVTYPE) %>%
    summarise(total_crop = sum(CROPDMG)) %>%
    arrange(desc(total_crop)) %>%
    head(10)

top10_property$EVTYPE <- factor(top10_property$EVTYPE, levels = top10_property$EVTYPE)
top10_crop$EVTYPE <- factor(top10_crop$EVTYPE, levels = top10_crop$EVTYPE)
```

```r
ggplot() +
  geom_bar(data = top10_property, aes(x = EVTYPE, y = total_property), stat = "identity", fill = "blue")
  geom_bar(data = top10_crop, aes(x = EVTYPE, y = total_crop), stat = "identity", fill = "red") +
  coord_flip() +
  labs(title = "Top 10 Events that Caused the Most Property and Crop Damage",
       x = "Event Type",
       y = "Total Property and Crop Damage") +
  theme(plot.title = element_text(hjust = 0.5))
```



From the plot above, we can see that the top 10 events that caused the most property damage are Flood, Hurricane/Typhoon, Tornado, Storm Surge/Tide, Hail, Flash Flood, Thunderstorm Wind, Ice Storm, Winter Storm, and Heavy Snow. The top 10 events that caused the most crop damage are Drought, Flood, Hail, Heat, Hurricane/Typhoon, Tornado, Storm Surge/Tide, Flash Flood, Thunderstorm Wind, and Winter Storm.