reproducible_research_project_2

Anh

8/21/2021

Reproducible Research Project 2: NOAA Storm Data Analysis

Sypnosis

An analysis of NOAA Storm Events Data ranging from 1950 to 2011. We aggregate the data and look at the total number of injuries, fatalities, and amount of damage caused. Overall, floods are responsible for the most economic damage, but tornadoes cause the most injuries and fatalities. They are also the 3rd leading cause of damage.

Data Processing

Download data and unzip

```
# library("R.utils")
# download.file(
#     "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
#     "NOAA.csv.bz2", method = "curl")
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data = read.csv("NOAA.csv.bz2.csv")
```

Results

What causes the most injuries?

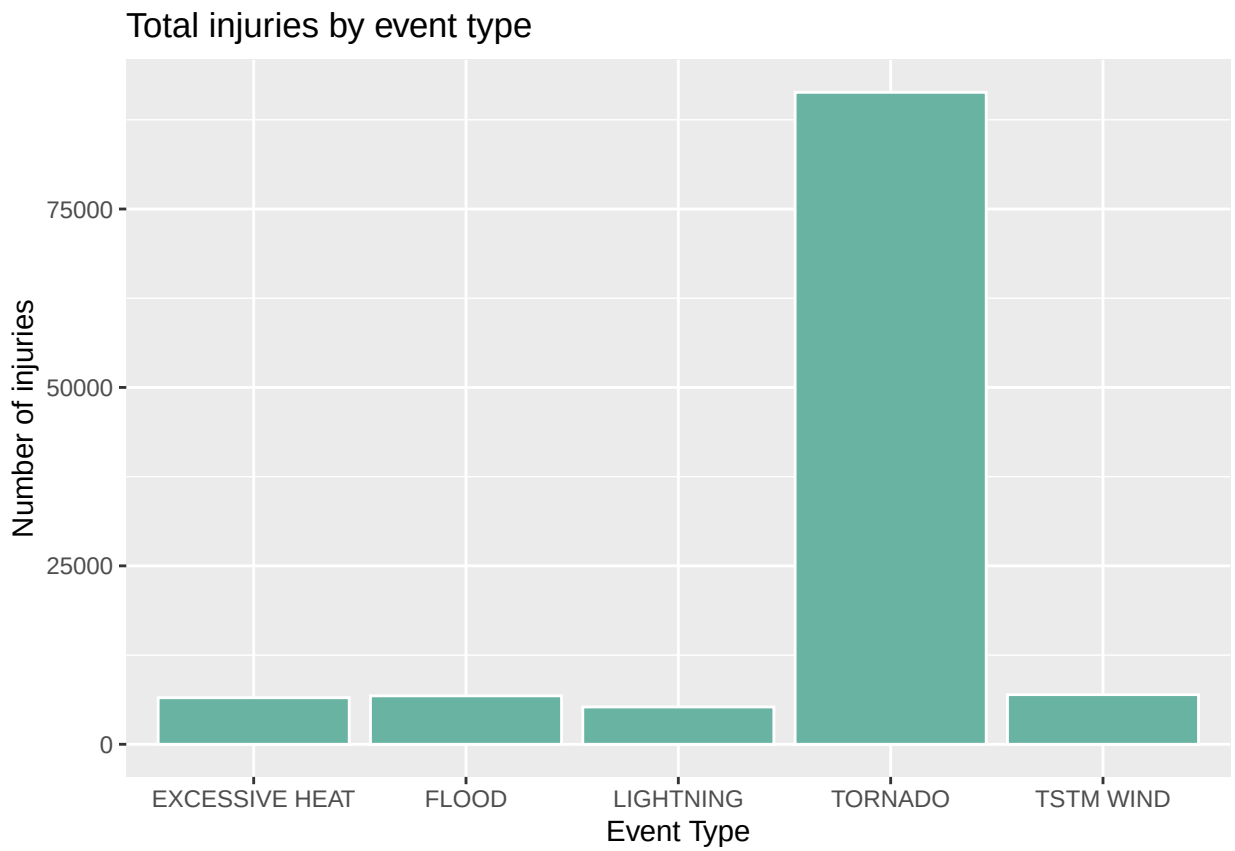
```
injuries = data %>%
            group_by(EVTYPE) %>%
            summarise(total_injuries_by_type = sum(INJURIES)) %>%
            arrange(desc(total_injuries_by_type))
head(injuries)
```

```
## # A tibble: 6 x 2
##   EVTYPE         total_injuries_by_type
##   <chr>                           <dbl>
## 1 TORNADO                         91346
## 2 TSTM WIND                        6957
## 3 FLOOD                            6789
## 4 EXCESSIVE HEAT                   6525
## 5 LIGHTNING                        5230
## 6 HEAT                             2100
```

We can see below that Tornadoes cause the most injuries.

Plot top 5 events by total injuries

```
ggplot(injuries[1:5, ], aes(EVTYPE, total_injuries_by_type)) +
  geom_bar(stat = "identity", fill = "#69b3a2", color = "White") +
  xlab("Event Type") +
  ylab("Number of injuries") +
  ggtitle("Total injuries by event type")
```



What causes the most fatalities?

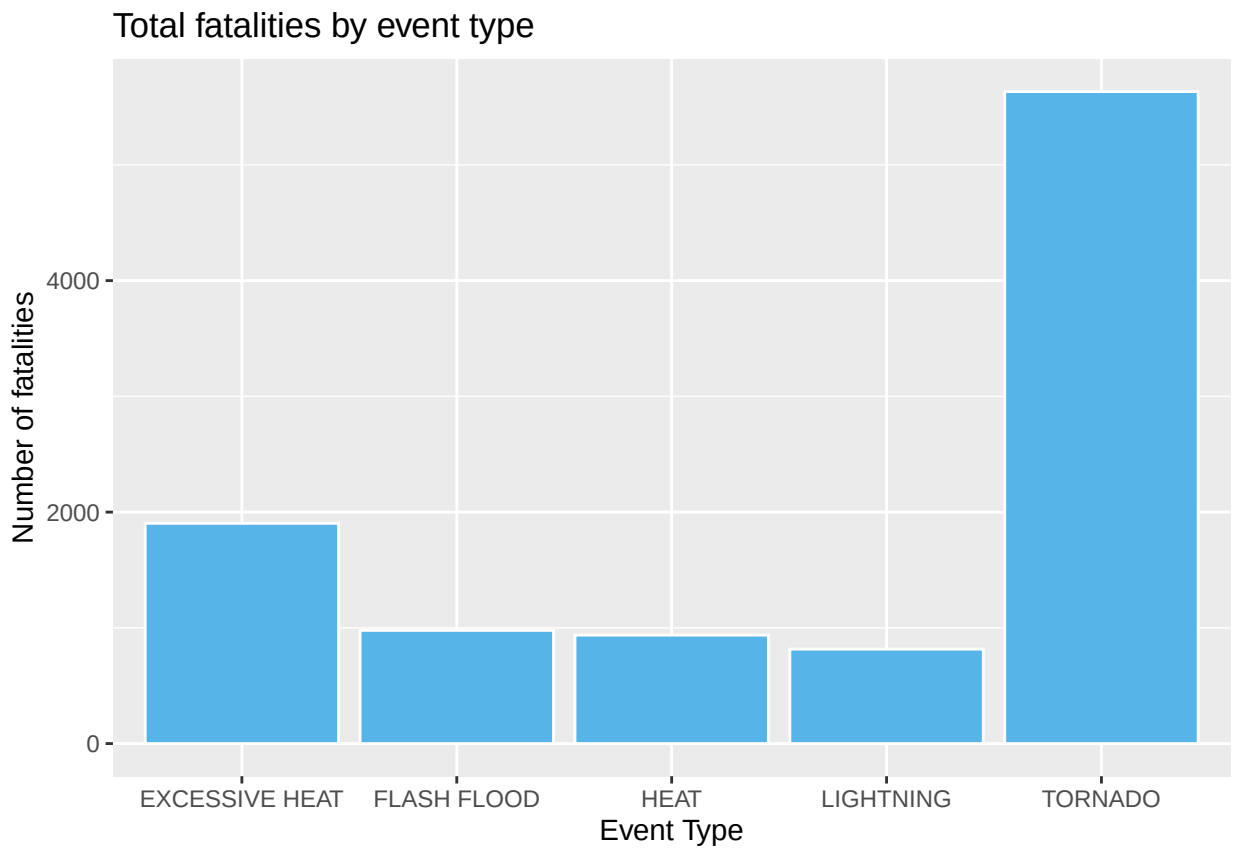
```
fatalities = data %>%
            group_by(EVTYPE) %>%
            summarise(total_fatalities_by_type = sum(FATALITIES)) %>%
            arrange(desc(total_fatalities_by_type))
head(fatalities)
```

```
## # A tibble: 6 x 2
##   EVTYPE         total_fatalities_by_type
##   <chr>                             <dbl>
## 1 TORNADO                            5633
## 2 EXCESSIVE HEAT                     1903
## 3 FLASH FLOOD                         978
## 4 HEAT                                937
## 5 LIGHTNING                           816
## 6 TSTM WIND                           504
```

We can see below that Tornadoes cause the most fatalities.

Plot top 5 events by total fatalities

```
ggplot(fatalities[1:5, ], aes(EVTYPE, total_fatalities_by_type)) +
  geom_bar(stat = "identity", fill = "#56B4E9", color = "White") +
  xlab("Event Type") +
  ylab("Number of fatalities") +
  ggtitle("Total fatalities by event type")
```



Calcultaing damage

The dataset provides an exponential multiplier in the form of 10x. The following code cleans up the inconsistencies in numeric values. Since we are most interested in total damage, we combine the values for Crop and Property damage.

```
# First, make everything upper case
data$PROPDMGEXP <- toupper(data$PROPDMGEXP)
data$CROPDMGEXP <- toupper(data$CROPDMGEXP)
unique(c(data$PROPDMGEXP, data$CROPDMGEXP))
```

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

```
# Slice necessary columns
damage <- data[, c("EVTYPE", "PROPDMG", "PROPDMGEXP", "CROPDMG",
                   "CROPDMGEXP")]
damage[damage$PROPDMGEXP %in% c("", "+", "-", "?"), "PROPDMGEXP"] <- "0"
```

```
damage[damage$CROPDMGEXP %in% c("", "+", "-", "?"), "CROPDMGEXP"] <- "0"
unique(c(damage$PROPDMGEXP, damage$CROPDMGEXP))
```

```
##  [1] "K" "M" "0" "B" "5" "6" "4" "2" "3" "H" "7" "1" "8"
```

```
# Create 10^x substitutions for Billion, Hundred, Kilo, and Million
damage[damage$PROPDMGEXP == "B", "PROPDMGEXP"] <- 9
damage[damage$CROPDMGEXP == "B", "CROPDMGEXP"] <- 9
damage[damage$PROPDMGEXP == "M", "PROPDMGEXP"] <- 6
damage[damage$CROPDMGEXP == "M", "CROPDMGEXP"] <- 6
damage[damage$PROPDMGEXP == "K", "PROPDMGEXP"] <- 3
damage[damage$CROPDMGEXP == "K", "CROPDMGEXP"] <- 3
damage[damage$PROPDMGEXP == "H", "PROPDMGEXP"] <- 2
damage[damage$CROPDMGEXP == "H", "CROPDMGEXP"] <- 2
unique(c(damage$PROPDMGEXP, damage$CROPDMGEXP))
```

```
##  [1] "3" "6" "0" "9" "5" "4" "2" "7" "1" "8"
```

```
# Now combine the exponent with the value
damage$PROPDMGEXP <- 10^(as.numeric(damage$PROPDMGEXP))
damage$CROPDMGEXP <- 10^(as.numeric(damage$CROPDMGEXP))
damage[is.na(damage$PROPDMG), "PROPDMG"] <- 0
damage[is.na(damage$CROPDMG), "CROPDMG"] <- 0
```

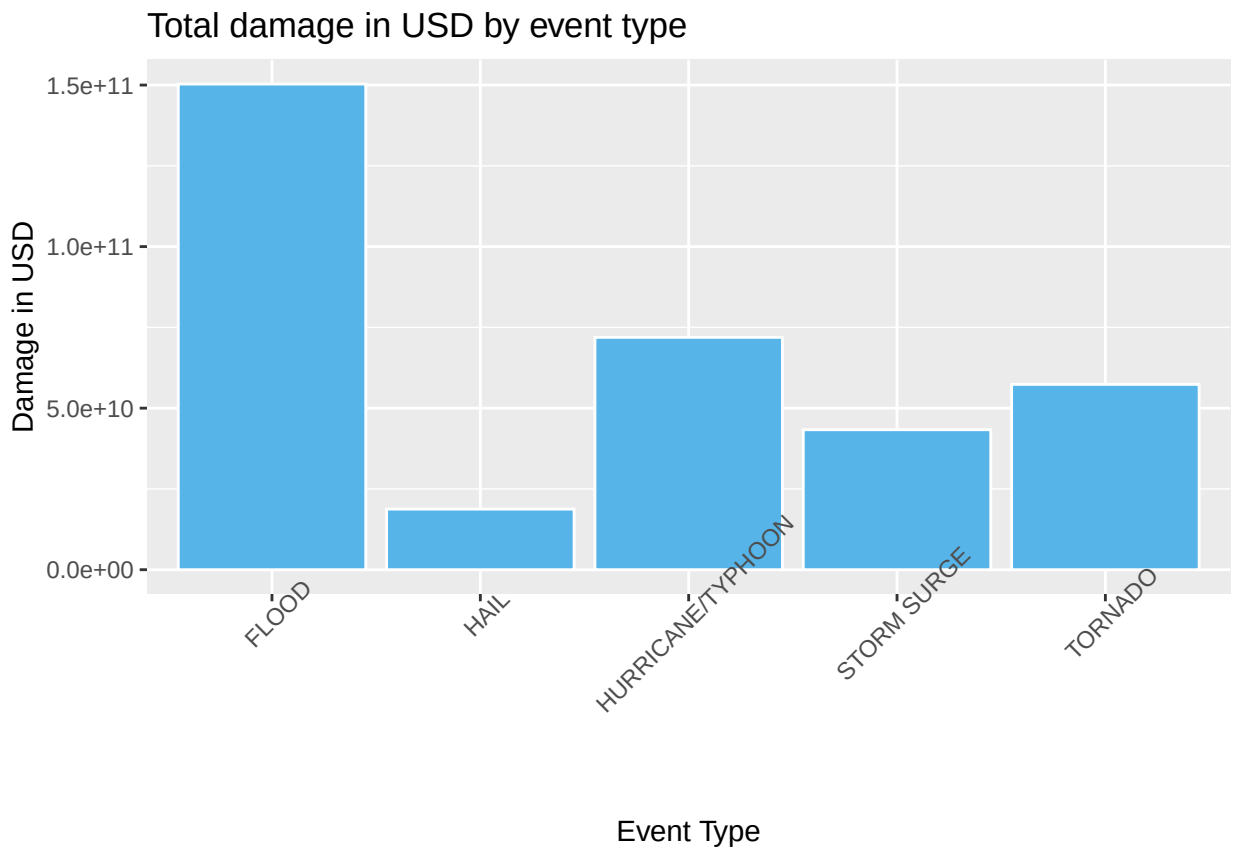
Calculate total damage by event type

```
total_damage = damage %>%
                group_by(EVTYPE) %>%
                summarise(total_damage_by_event_type = sum(PROPDMG * PROPDMGEXP
                                                           + CROPDMG * CROPDMGEXP)
                          ) %>%
                arrange(desc(total_damage_by_event_type))
head(total_damage)
```

```
## # A tibble: 6 x 2
##   EVTYPE            total_damage_by_event_type
##   <chr>                                  <dbl>
## 1 FLOOD                          150319678257
## 2 HURRICANE/TYPHOON               71913712800
## 3 TORNADO                         57362333946.
## 4 STORM SURGE                     43323541000
## 5 HAIL                            18761221986.
## 6 FLASH FLOOD                     18243991078.
```

We can see that Floods cause the most damage. Plot total damage by event type

```
ggplot(total_damage[1:5, ], aes(EVTYPE, total_damage_by_event_type)) +
  geom_bar(stat = "identity", fill = "#56B4E9", color = "White") +
  xlab("Event Type") +
  ylab("Damage in USD") +
  ggtitle("Total damage in USD by event type") +
  theme(axis.text.x = element_text(angle = 45))
```



Conclusion

Floods are responsible for the most economic damage, but tornadoes cause the most injuries and fatalities. Floods are also the 3rd leading cause of damage.