Assignment: An anlysis with NOAA Storm Database on severe weather events

Nanyuchentw

2023-10-07

Overview

1. Synopsis:

This analysis used the NOAA Storm Database to find the event types that are either most harmful to population health or that caused the greatest economic damage. Data were first imported from the bz2 zipped file. The columns related to economic loss (PROPDMG, PROPDMGEXP, CROPDMG and CROPDMG-EXP) were transformed from characters to numbers using the exponent indicated in the PROPDMGEXP and CROPDMGEXP columns. Total economic loss was then calculated by sum of property and crop loss amount. The fatality and injuries were calculated directed by the original data and grouped by event types. Finally, Top 5 most harmful event types were plotted.

2. Data Processing

2.1 Load libraries

```r
library(tidyverse)
```

2.2 Read in data

```r
# read in data
df <- read.csv("repdata_data_StormData.csv.bz2")
```

2.3 Data transformation

```r
#Because the PROPDMGEXP and CROPDMGEXP are exponents,
#transform the PROPDMG and PROPDMGEXP columns to property loss amounts
df$PROPDMGEXP <- str_replace_all(df$PROPDMGEXP, "[Hh]", "2")
df$PROPDMGEXP <- str_replace_all(df$PROPDMGEXP, "[Kk]", "3")
df$PROPDMGEXP <- str_replace_all(df$PROPDMGEXP, "[Mm]", "6")
df$PROPDMGEXP <- str_replace_all(df$PROPDMGEXP, "[Bb]", "9")
df$PROPDMGEXP <- str_replace_all(df$PROPDMGEXP, "[?+-]", "0")
df$PROPDMGEXP[which(df$PROPDMGEXP=="")] <- "0"
# add a column for total property damage
df <- df %>% mutate(total_PROPDMG= as.numeric(PROPDMG)*(10^as.numeric(PROPDMGEXP)))

# transform the CROPDMG and CROPDMGEXP columns to crops loss amounts
df$CROPDMGEXP <- str_replace_all(df$CROPDMGEXP, "[Kk]", "3")
df$CROPDMGEXP <- str_replace_all(df$CROPDMGEXP, "[Mm]", "6")
df$CROPDMGEXP <- str_replace_all(df$CROPDMGEXP, "[Bb]", "9")
df$CROPDMGEXP <- str_replace_all(df$CROPDMGEXP, "[?]", "0")
df$CROPDMGEXP[which(df$CROPDMGEXP=="")] <- "0"
df <- df %>% mutate(total_CROPDMG= as.numeric(CROPDMG)*(10^as.numeric(CROPDMGEXP)))

#add a column for total economic loss
df <- df %>% mutate(eco_loss= total_CROPDMG+ total_PROPDMG)

#find the type of events that cost the top 5 economic loss and then make a plot
p <- df %>% group_by(EVTYPE) %>%
        summarise(Total_loss= sum(eco_loss)) %>%
        arrange(desc(Total_loss)) %>% slice_head(n=5)
ggplot(p, aes(x=EVTYPE, y=Total_loss))+ geom_col()+
        labs(x= "Event type", y="Economic loss (dollars)" ,
            title = "Top 5 event types with the greatest economic consequences")
```
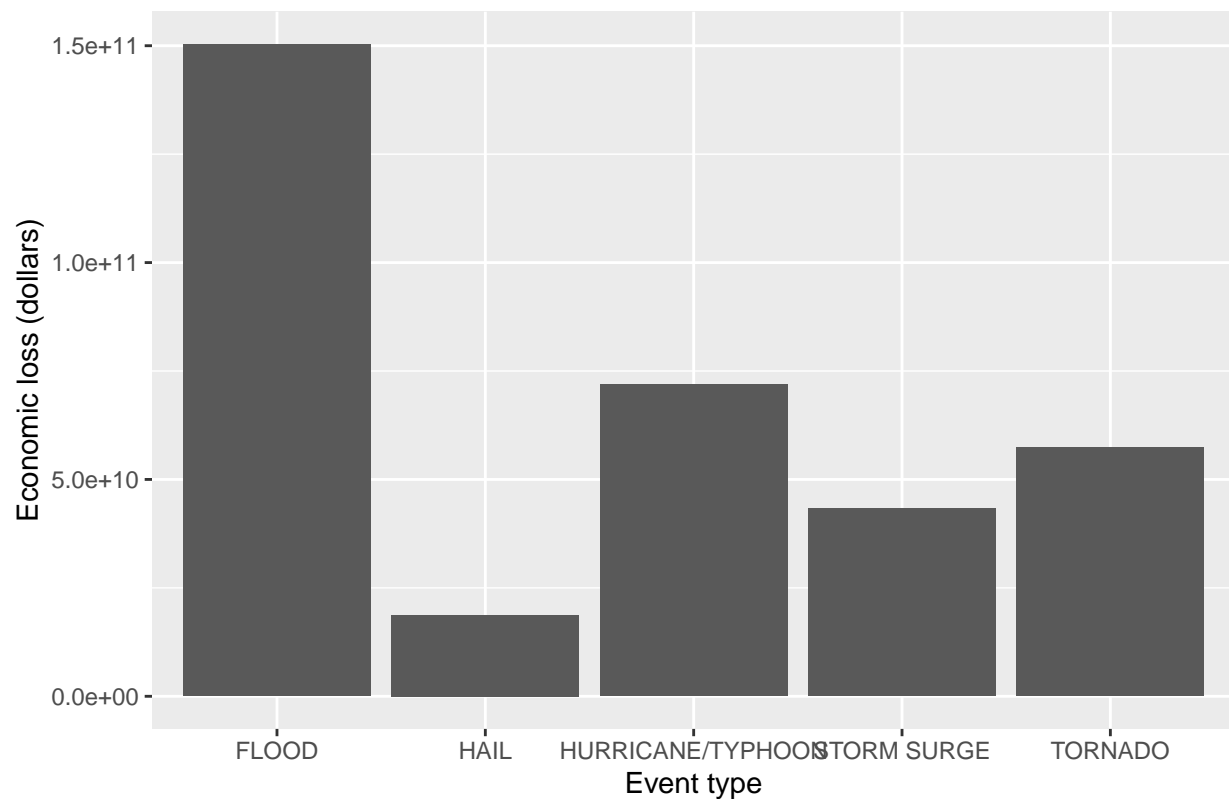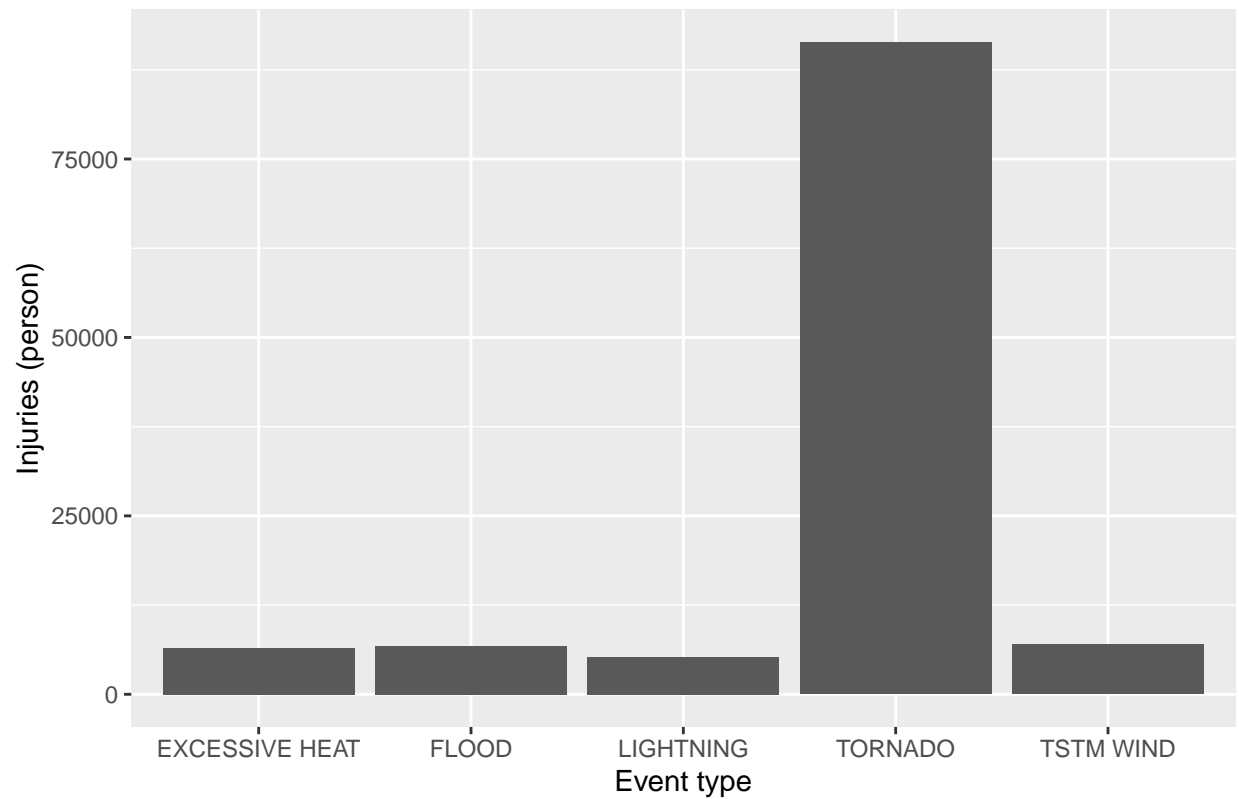
## Top 5 event types with the greatest economic consequences


```r
#Evaluate event types that are most harmful to population health
#Find the event types that caused greatest injuries
health_cost <- df %>% group_by(EVTYPE) %>%
        summarise(inj= sum(INJURIES), death=sum(FATALITIES))
p <- health_cost %>% arrange(desc(inj)) %>% slice(1:5)
ggplot(p, aes(x= EVTYPE, y=inj))+ geom_col()+
        labs(x="Event type", y="Injuries (person)",
            title= "Top 5 event types causing the greatest number of injuries")
```

Top 5 event types causing the greatest number of injuries


```r
#Find the event types that caused greatest fatalities
p <- health_cost %>% arrange(desc(death)) %>% slice(1:5)
ggplot(p, aes(x= EVTYPE, y=death))+ geom_col()+
        labs(x="Event type", y="Fatality (person)",
             title= "Top 5 event types causing the greatest number of life loss")
```

Top 5 event types causing the greatest number of life loss
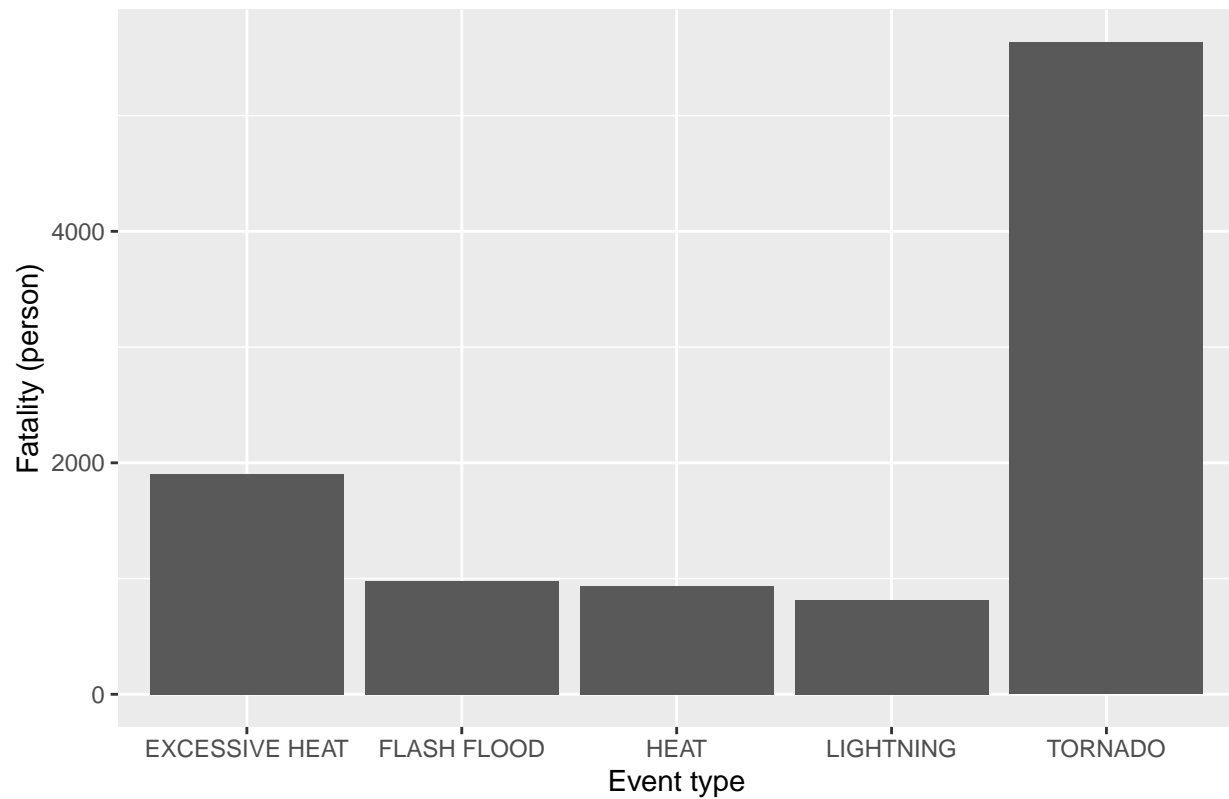

```r
#Find the greatest numbers
max_inj <- health_cost[which(health_cost$inj== max(health_cost$inj)),1:2]
max_death <- health_cost[which(health_cost$death== max(health_cost$death)),c(1,3)]
```

3. Result:

With respect to economic loss, Flood has the greatest economic consequences.

To population health, Tornados are the most harmful event.