Reproducible Research - Course Project 2

Eilder Jorge

2020/7/4

Introduction

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This 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 analysis below will analyze the major storm events causing injuries and fatalities. Similarly, we will also examine the major Storm Event causing highest property damage.

Synopsis

The analysis on the storm event database revealed that tornadoes are the most dangerous weather event to the populations health. The second most dangerous event type is excessive heat. The economic impact of weather events was also analyzed. Flash floods and thunderstorm winds caused billions of dollars in property damages between 1950 and 2011. The largest damage to crops were caused by droughts, followed by floods and hailing.

Load libraries used

```
library(ggplot2)
library(R.utils)
library(dplyr)
```

Data load

```
if (!file.exists("StormData.csv")) {
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url, "StormData.csv.bz2")
bunzip2("StormData.csv.bz2", "StormData.csv")
}
df <- read.csv("StormData.csv")
```

Data Processing

Health Impact

To evaluate the health impact, the total fatalities and the total injuries for each event type (EVTYPE) are calculated. The codes for this calculation are shown as follows.

```
df.fatalities <- df %>% select(EVTYPE, FATALITIES) %>% group_by(EVTYPE) %>% summarise(total.fatalities =
head(df.fatalities, 10)
```

```
## # A tibble: 10 x 2
##    EVTYPE         total.fatalities
##    <fct>                     <dbl>
##  1 TORNADO                    5633
##  2 EXCESSIVE HEAT             1903
##  3 FLASH FLOOD                 978
##  4 HEAT                        937
##  5 LIGHTNING                   816
##  6 TSTM WIND                   504
##  7 FLOOD                       470
##  8 RIP CURRENT                 368
##  9 HIGH WIND                   248
## 10 AVALANCHE                   224
```

```
df.injuries <- df %>% select(EVTYPE, INJURIES) %>% group_by(EVTYPE) %>% summarise(total.injuries = sum(I
head(df.injuries, 10)
```

```
## # A tibble: 10 x 2
##    EVTYPE            total.injuries
##    <fct>                      <dbl>
##  1 TORNADO                    91346
##  2 TSTM WIND                   6957
##  3 FLOOD                       6789
##  4 EXCESSIVE HEAT              6525
##  5 LIGHTNING                   5230
##  6 HEAT                        2100
##  7 ICE STORM                   1975
##  8 FLASH FLOOD                 1777
##  9 THUNDERSTORM WIND           1488
## 10 HAIL                        1361
```

At this point we got the amount of fatalities and injuries per event type.

Economic Impact

The data provides two types of economic impact, namely property damage (PROPDMG) and crop damage (CROPDMG). The actual damage in $USD is indicated by PROPDMGEXP and CROPDMGEXP parameters.

The indexes in the PROPDMGEXP and CROPDMGEXP have the following multipliers:

H, h -> hundreds = x100 K, K -> kilos = x1,000 M, m -> millions = x1,000,000 B,b -> billions = x1,000,000,000 (+) -> x1 (-) -> x0 (?) -> x0 blank -> x0

So we need to make some math and conversions to get the actual damage values.

```
df.damage <- df %>% select(EVTYPE, PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)

Symbol <- sort(unique(as.character(df.damage$PROPDMGEXP)))
```

```
Multiplier <- c(0,0,0,1,10,10,10,10,10,10,10,10,10,10^9,10^2,10^2,10^3,10^6,10^6)
convert.Multiplier <- data.frame(Symbol, Multiplier)

df.damage$Prop.Multiplier <- convert.Multiplier$Multiplier[match(df.damage$PROPDMGEXP, convert.Multiplie
df.damage$Crop.Multiplier <- convert.Multiplier$Multiplier[match(df.damage$CROPDMGEXP, convert.Multiplie

df.damage <- df.damage %>% mutate(PROPDMG = PROPDMG*Prop.Multiplier) %>% mutate(CROPDMG = CROPDMG*Crop.M

df.damage.total <- df.damage %>% group_by(EVTYPE) %>% summarize(TOTAL.DMG.EVTYPE = sum(TOTAL.DMG))%>% ar

head(df.damage.total,10)
```

```
## # A tibble: 10 x 2
##    EVTYPE            TOTAL.DMG.EVTYPE
##    <fct>                        <dbl>
##  1 FLOOD                 150319678250
##  2 HURRICANE/TYPHOON      71913712800
##  3 TORNADO                57352117607
##  4 STORM SURGE            43323541000
##  5 FLASH FLOOD            17562132111
##  6 DROUGHT                15018672000
##  7 HURRICANE              14610229010
##  8 RIVER FLOOD            10148404500
##  9 ICE STORM               8967041810
## 10 TROPICAL STORM          8382236550
```

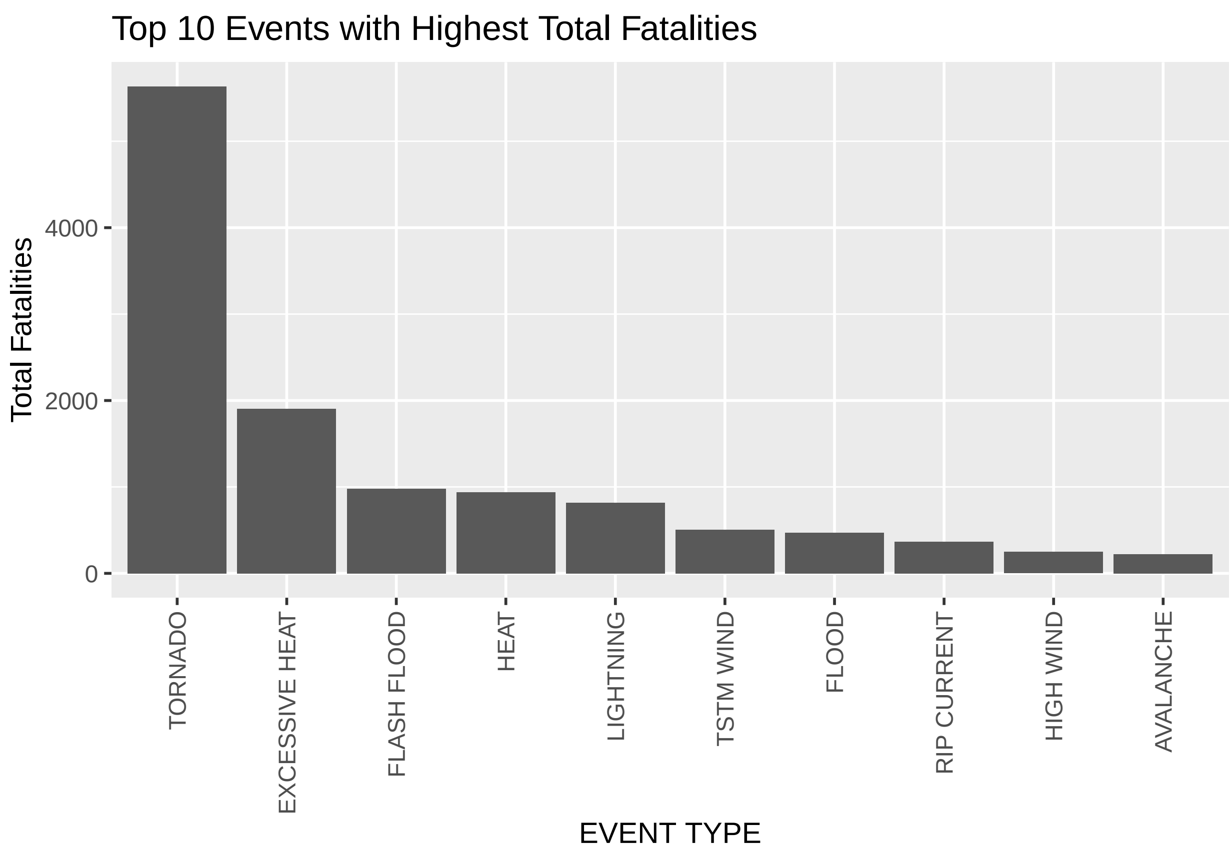
At this point we got the amount of economic damage per event type.

Results

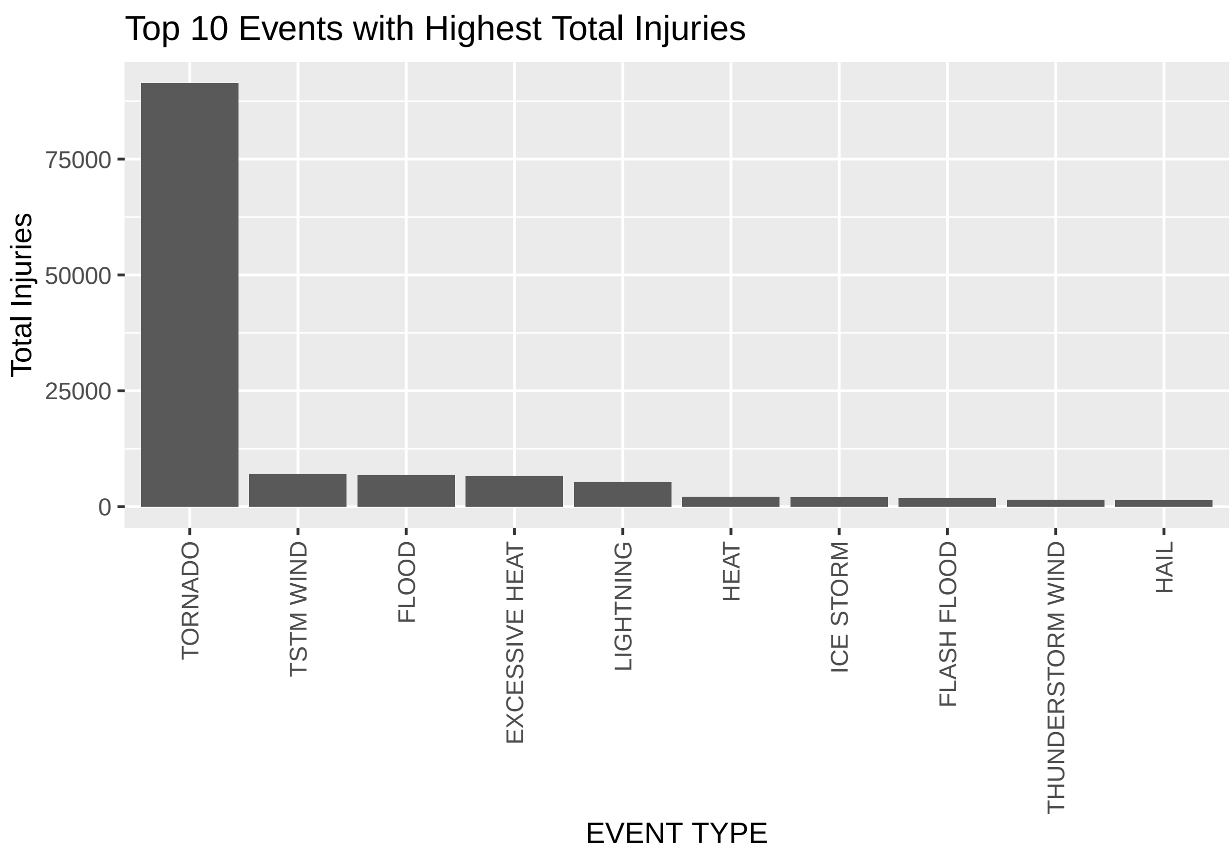
Health Impact

The top 10 events with the highest total fatalities and injuries are shown in the graphic.

```
ggplot(df.fatalities[1:10,], aes(x=reorder(EVTYPE, -total.fatalities), y=total.fatalities))+geom_bar(sta
```




```
ggplot(df.injuries[1:10,], aes(x=reorder(EVTYPE, -total.injuries), y=total.injuries))+geom_bar(stat="id
```

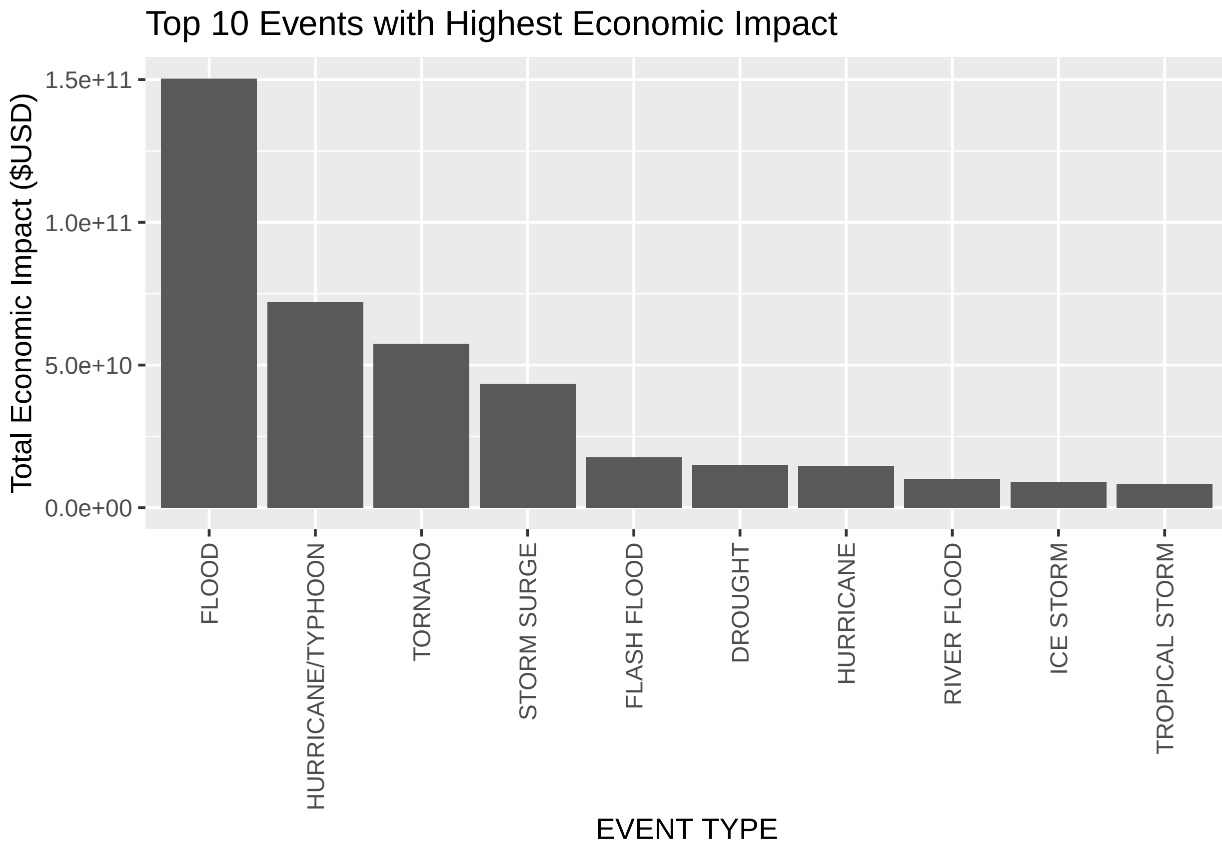


As explained in the synopsis, tornadoes have the highest amount of fatalities and injuries by a long margin.

Economic Impact

The top 10 events with the highest total economic damages are shown in the graphic.

```
ggplot(df.damage.total[1:10,], aes(x=reorder(EVTYPE, -TOTAL.DMG.EVTYPE), y=TOTAL.DMG.EVTYPE))+geom_bar(
```



We can observe that floods cause the highest economic impact.