Reproducible Research: Course Project 2

Sima N May 21, 2020

Synopsis

This report explores the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database to identify when and where major storms and weather events occur in the United States, and what is the estimate of any fatalities, injuries, and property damage.

1 - which types of events are most harmful to population health?

2 - which types of events have the greatest economic consequences?

Results show that tornados are the most damaging weather type to population health; and floods are the most damaging weather event.

knitr::opts_chunk$set(warning=FALSE)

Data Processing

Loading the data into R

```
#add libraries
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.5.3
```

```
#download data
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url, destfile ='repdata-data-StormData.csv.bz2')

#read data
stormdata <- read.csv("repdata-data-StormData.csv.bz2")
head(stormdata)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                      14.0   100 3   0          0       15    25.0
## 2         0                       2.0   150 2   0          0        0     2.5
## 3         0                       0.1   123 2   0          0        2    25.0
## 4         0                       0.0   100 2   0          0        2     2.5
## 5         0                       0.0   150 2   0          0        2     2.5
## 6         0                       1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

variables used in the analysis:

EVTYPE: Event Type (Tornados, Flood, ..)

FATALITIES: Number of Fatalities

INJURIES: Number of Injuries

PROGDMG: Property Damage

PROPDMGEXP: Units for Property Damage (magnitudes - K,B,M)

CROPDMG: Crop Damage

CROPDMGEXP: Units for Crop Damage (magnitudes - K,BM,B)

```
varsNedeed <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDM
G", "CROPDMGEXP")
storm <- stormdata[varsNedeed]
dim(storm)
```

```
## [1] 902297      7
```

Results

Which events are most harmful to population Health?

Fatalities Data

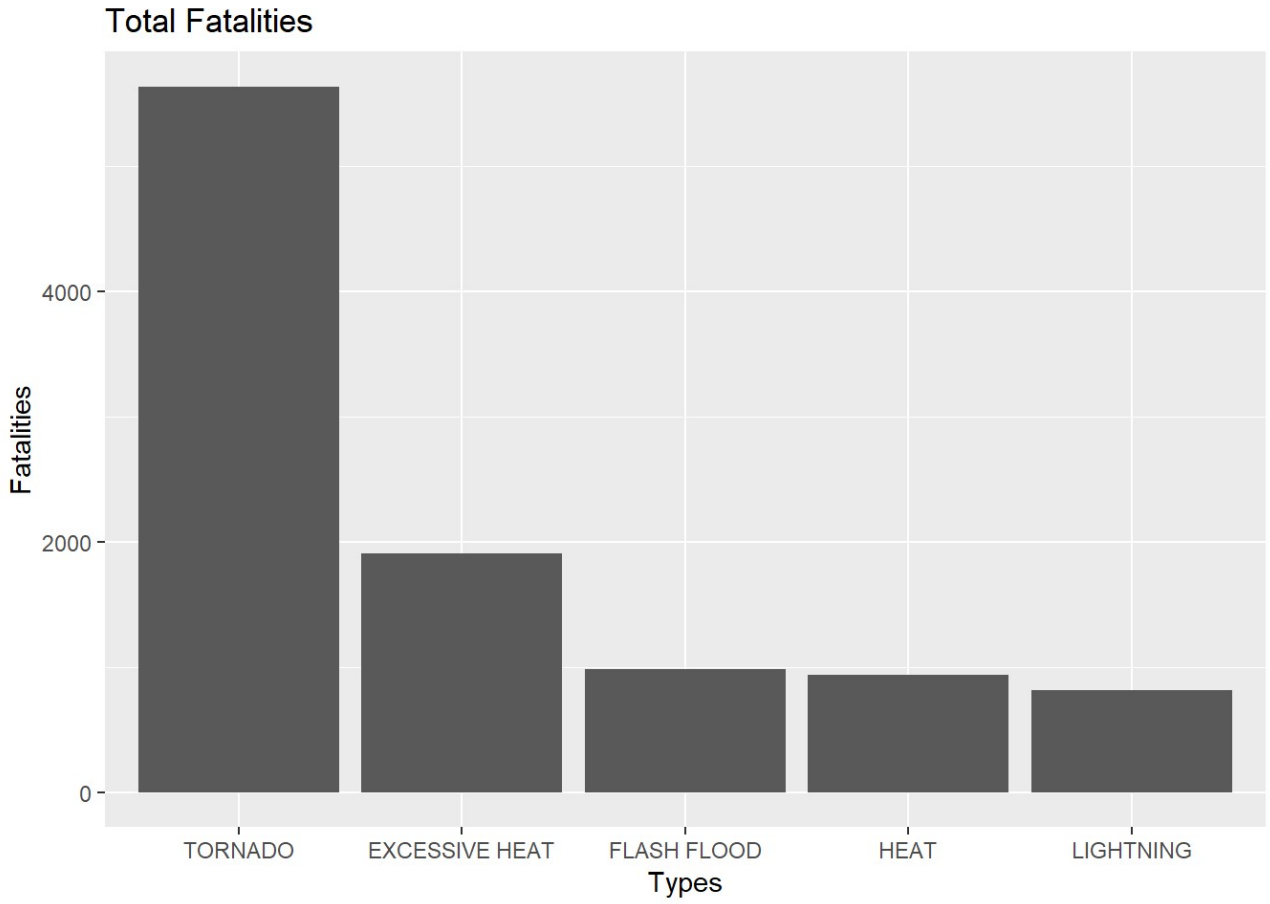
```
fatalityData <- (subset(stormdata, FATALITIES > 0))[c(8, 23)]
fatalityGroupedData <- aggregate(FATALITIES ~ EVTYPE, data = fatalityData, FUN = "su
m", na.rm = TRUE)
fatalityGroupedData <- fatalityGroupedData[order(fatalityGroupedData$FATALITIES, decre
asing=TRUE), ]
fatalityGroupedData <- fatalityGroupedData[1:5, ]
fatalityGroupedData$EVTYPE <- factor(fatalityGroupedData$EVTYPE, levels=fatalityGroupe
dData$EVTYPE)
```

Injuries Data

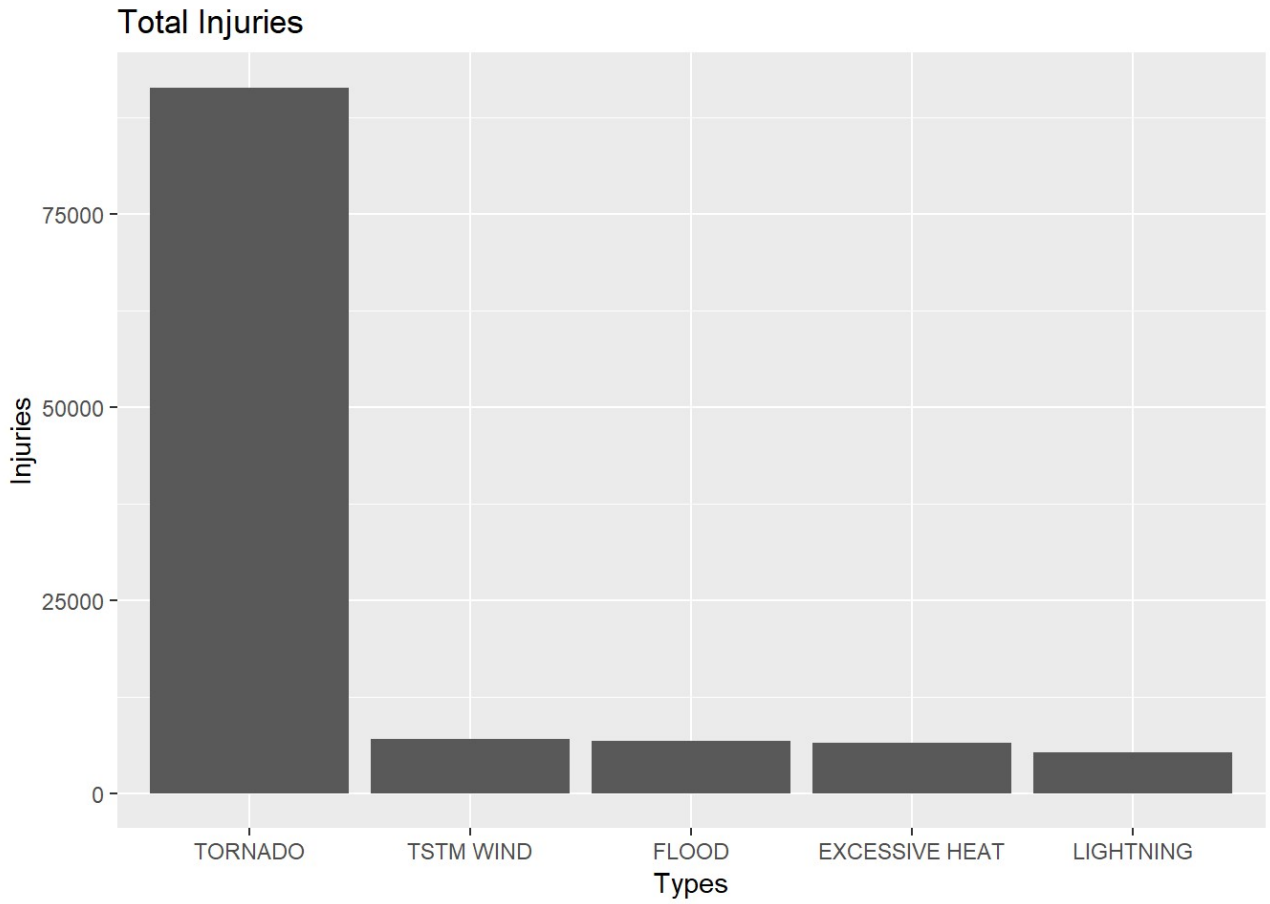
```
injuryData <- (subset(stormdata, INJURIES > 0))[c(8, 24)]
injuryGroupedData <- aggregate(INJURIES ~ EVTYPE, data = injuryData, FUN = "sum", na.r
m = TRUE)
injuryGroupedData <- injuryGroupedData[order(injuryGroupedData$INJURIES, decreasing=TR
UE), ]
injuryGroupedData <- injuryGroupedData[1:5, ]
injuryGroupedData$EVTYPE <- factor(injuryGroupedData$EVTYPE, levels=injuryGroupedData
$EVTYPE)
```

ploting the results

```
ggplot(fatalityGroupedData, aes(x=EVTYPE, y=FATALITIES)) +
  geom_bar(stat="identity") +
  xlab("Types") +
  ylab("Fatalities") +
  ggtitle("Total Fatalities")
```



```
ggplot(injuryGroupedData, aes(x=EVTYPE, y=INJURIES)) +
  geom_bar(stat="identity") +
  xlab("Types") +
  ylab("Injuries") +
  ggtitle("Total Injuries")
```



Which type of Events have the greatest Economic consequences?

```
#Convert economic impact to monitary value.
damageAmount <- function(amount, magnitude)
{
  returnAmount <- 0
  if (toupper(magnitude)[1]=="K")
  {
    returnAmount <- (amount * 1000)
  }
  if (toupper(magnitude)[1]=="M")
  {
    returnAmount <- (amount * 1000000)
  }
  if (toupper(magnitude)[1]=="B")
  {
    returnAmount <- (amount * 1000000000)
  }
  return(returnAmount)
}

damageData <- (subset(stormdata, PROPDMG > 0 | CROPDMG > 0))[c(8, 25, 26, 27, 28)]
damageData$DamageAmount <- ((mapply(damageAmount, damageData$PROPDMG, damageData$PROPD
MGEXP)) +
                                  (mapply(damageAmount, damageData$CROPDMG, damageData
$CROPDMGEXP)))

damageGroupedData <- aggregate(DamageAmount ~ EVTYPE, data = damageData, FUN = "sum",
na.rm = TRUE)
damageGroupedData <- damageGroupedData[order(damageGroupedData$DamageAmount, decreasin
g=TRUE), ]
damageGroupedData <- damageGroupedData[1:5, ]
damageGroupedData$EVTYPE <- factor(damageGroupedData$EVTYPE, levels=damageGroupedData
$EVTYPE)

#Most harmful events
head(damageGroupedData, 5)
```

```
##                EVTYPE DamageAmount
## 72              FLOOD 150319678250
## 197 HURRICANE/TYPHOON  71913712800
## 354           TORNADO  57352113590
## 299       STORM SURGE  43323541000
## 116              HAIL  18758221170
```