assignment_2

Rakesh Patel

May 25, 2016

Effects of weather on states

Synopsis

In this data analysis we are going to see the effects of weather events across the United States on population health and economy . 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.Data can be found in this link and There is also some documentation of the database available in this link .

after setting working directory to the folder containing this file reading the data

Data Processing

Loading and Processing the Raw Data

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

1.Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

we can calculate the number of injuries to see which are most harmful events

```
dt1<-aggregate(INJURIES~EVTYPE,data=dt,FUN = sum)
head(dt1)
```

```
##                    EVTYPE INJURIES
## 1     HIGH SURF ADVISORY        0
## 2         COASTAL FLOOD        0
## 3          FLASH FLOOD        0
## 4             LIGHTNING        0
## 5             TSTM WIND        0
## 6        TSTM WIND (G45)        0

library(dplyr)
```

```
## 
## 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

harmful1<-arrange(dt1,desc(INJURIES))
head(harmful1)

##            EVTYPE INJURIES
## 1          TORNADO    91346
## 2        TSTM WIND     6957
## 3            FLOOD     6789
## 4  EXCESSIVE HEAT     6525
## 5        LIGHTNING     5230
## 6             HEAT     2100
```
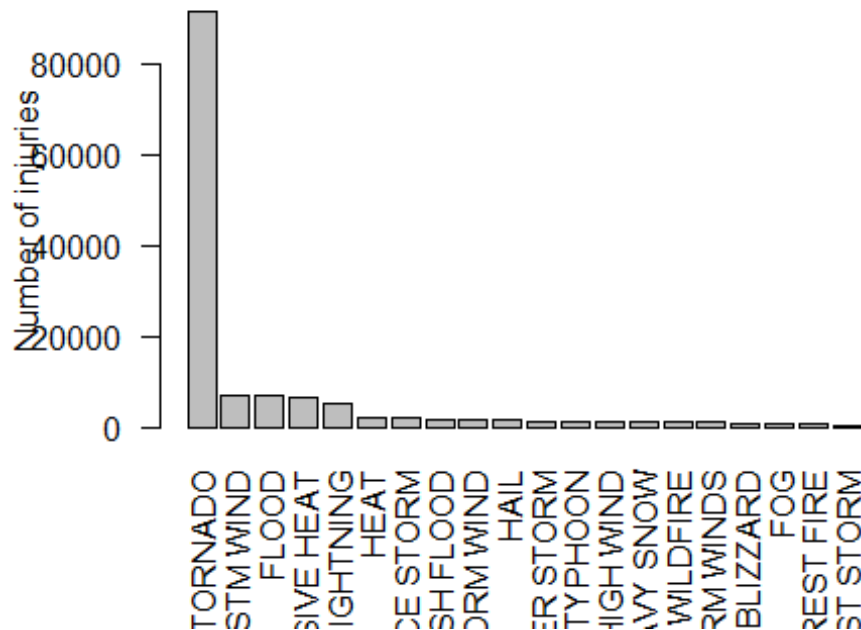
1. Tornado is the most harmful w.r.t. number of injuries

we can plot Total injuries vs top 20 events graph to see which events are more Harmful

```
barplot(height = harmful1$INJURIES[1:20],names.arg =
harmful1$EVTYPE[1:20],las=2 ,ylab = "Number of injuries",main = "Most Harmful
events w.r.t. Injuries" )
```

Most Harmful events w.r.t. Injuries

we can calculate the number of fatalities to see which are most harmful events

```
dt2<-aggregate(FATALITIES~EVTYPE,data=dt,FUN = sum)
head(dt2)
```

```
##                   EVTYPE FATALITIES
## 1     HIGH SURF ADVISORY          0
## 2          COASTAL FLOOD          0
## 3            FLASH FLOOD          0
## 4              LIGHTNING          0
## 5              TSTM WIND          0
## 6        TSTM WIND (G45)          0
```

```
library(dplyr)
```

```
harmful2<-arrange(dt2,desc(FATALITIES))
head(harmful2)
```

```
##            EVTYPE FATALITIES
## 1         TORNADO       5633
## 2  EXCESSIVE HEAT       1903
## 3     FLASH FLOOD        978
## 4            HEAT        937
## 5       LIGHTNING        816
## 6       TSTM WIND        504
```

2. Tornado is the most harmful w.r.t. number of fatalities

we can plot Total injuries vs top 20 events graph to see which events are more Harmful

```
barplot(height = harmful2$FATALITIES[1:20],names.arg =
harmful2$EVTYPE[1:20],las=2 ,ylab = "Number of fatalities",main = "Most
Harmful events w.r.t. fatalities" )
```

Most Harmful events w.r.t. fatalities
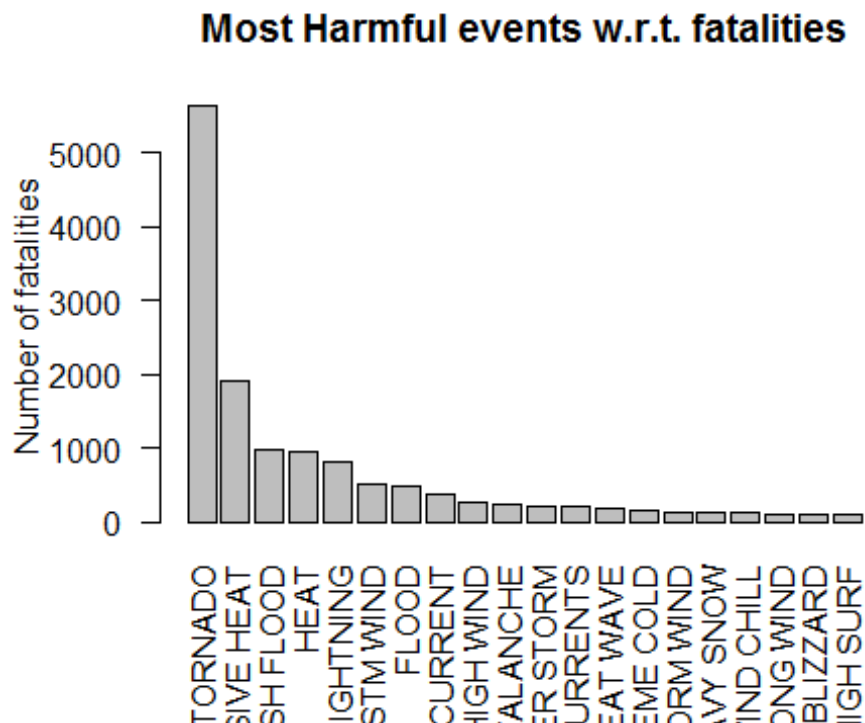
2.Across the United States, which types of events have the greatest economic consequences?

For this we can calculate property damage and crop damage for all types of events

In the PROPDMGEXP and CROPDMGEXP column damage is given in 10^power and power is

```
levels(dt$PROPDMGEXP)

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

levels(dt$CROPDMGEXP)

## [1] ""  "?" "0" "2" "B" "k" "K" "m" "M"
```

But we need to convert it in Dollers

```
index<-c(""  , "-", "?" ,"+" ,"0" ,"1" ,"2", "3", "4" ,"5" ,"6" ,"7" ,"8", "B"
,"h" ,"H" ,"k","K","m" ,"M")
Power<-c(0, 0, 0, 0, 0: 9, 2, 2, 3, 3, 6, 6)
df<-data.frame(index, Power)
```

Now we need to find character representation of 10th power and assign corresponding numerical 10th power

```
dt$prop_dmg<-dt$PROPDMG*10^(df[match(dt$PROPDMGEXP, df$index), 2])
dt$crop_dmg<-dt$CROPDMG*10^(df[match(dt$CROPDMGEXP, df$index), 2])
dt$total_dmg<-dt$prop_dmg+dt$crop_dmg
max(dt$total_dmg)

## [1] 115032500000
```

We can covert it in Billions of USD

```
dt$total_dmg<-dt$total_dmg/1E9
```

we can make plot to see top 20 events that are more Harmful w.r.t Economy

```
dt3<-aggregate(total_dmg~EVTYPE, dt, FUN=sum)

library(dplyr)

harmful3<-arrange(dt3, desc(total_dmg))
head(harmful3)

##                EVTYPE total_dmg
## 1               FLOOD 150.31968
## 2 HURRICANE/TYPHOON    71.91371
## 3             TORNADO  57.36233
## 4         STORM SURGE  43.32354
## 5                HAIL  18.76122
## 6         FLASH FLOOD  18.24399

barplot(height = harmful3$total_dmg[1:20], names.arg =
harmful3$EVTYPE[1:20], las=2 ,ylab = "Loss in Billions of USD", main = "Most
Harmful events w.r.t. Economy" )
```
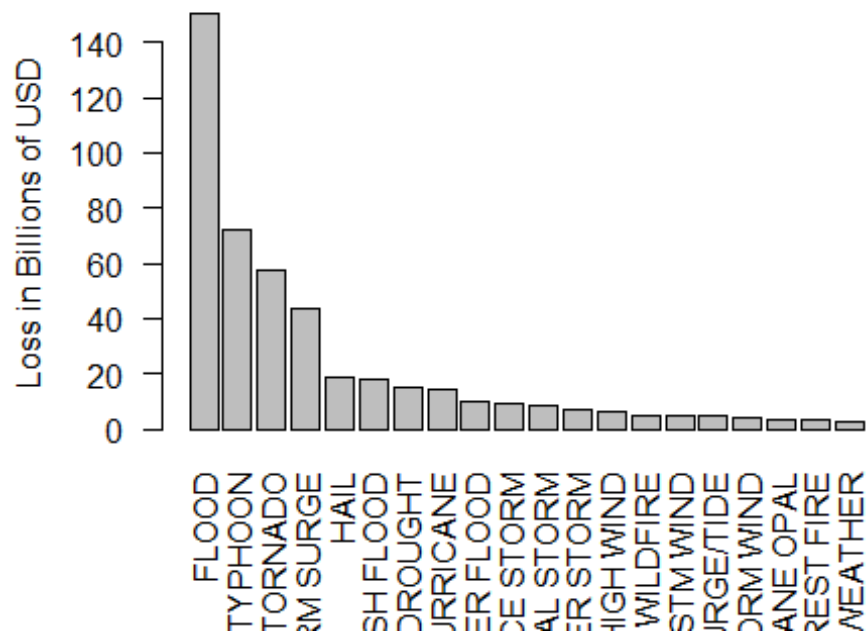
Most Harmful events w.r.t. Economy

Results

1.Across the United StatesTORNADO is most harmful with respect to population health.

2.Across the United States Flood has the greatest economic consequences.