Storm Data Analysis

Synopsys

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern. 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.

- Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

This analysis is based on 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 file can be downloaded from the course web site:

Storm Data (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) [47Mb]

Data Processing

```
#Required packages
library(ggplot2)
```

```
# Download and unzip the file:
if(!file.exists("./stormData")) {dir.create("./stormData")}
urlzip <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(urlzip, destfile = "./stormData/StormData.csv.bz2" )
# Load data into R
stormData <- read.csv("./stormData/StormData.csv.bz2")
# See the structure of tha data
str(stormData)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116
2224 2224 2260 383 3980 3980 ...
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186
242 1683 3186 3186 ...
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE
LT MI",..: 13513 1873 4598 10592 4372 10094 1973 23873 24418 4598 ...
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 83
4 834 834 834 834 834 ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","- 1 N Albion",..: 1 1 1 1 1 1 1 1 1 1
...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1
...
##  $ END_TIME  : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","- .5 NNW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 1
7 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","  CI","$AC",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1
...
##  $ ZONENAMES : Factor w/ 25112 levels "","
"| __truncated__,..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","-2 at Deer Park\n",..: 1 1 1 1 1 1 1 1
1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

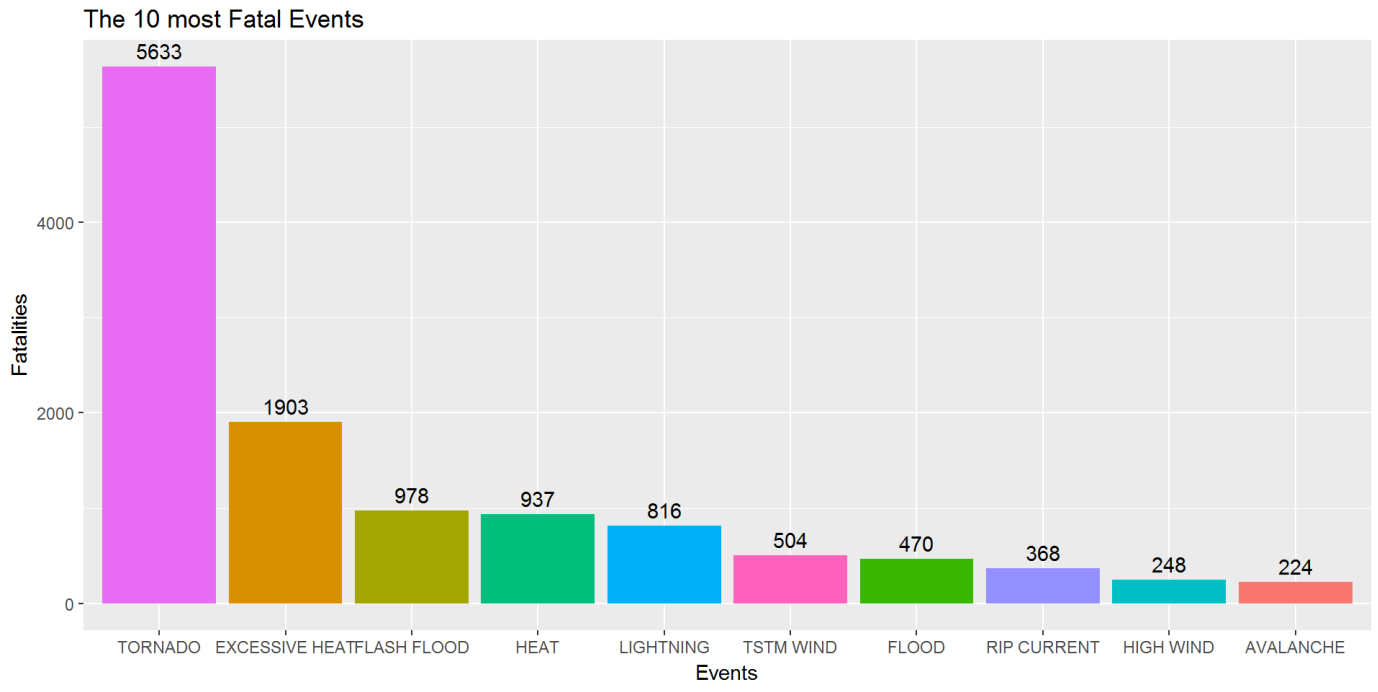
Fatalities calculated by type:

```
fatal <- aggregate(FATALITIES ~ EVTYPE, data = stormData, sum)
fatal <- fatal[fatal$FATALITIES > 0, ]
fatal <- fatal[order(fatal$FATALITIES, decreasing = T),]
head(fatal, 10)
```

```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

Result:

```
ggplot(fatal[1:10,], aes(reorder(EVTYPE, -FATALITIES), FATALITIES, fill = EVTYPE)) +
      geom_bar(stat = "identity") +
      geom_text(aes(label = FATALITIES), vjust = -0.5, colour = "black") +
      labs(title = "The 10 most Fatal Events", y = "Fatalities", x = "Events") +
      scale_fill_discrete(guide = FALSE)
```



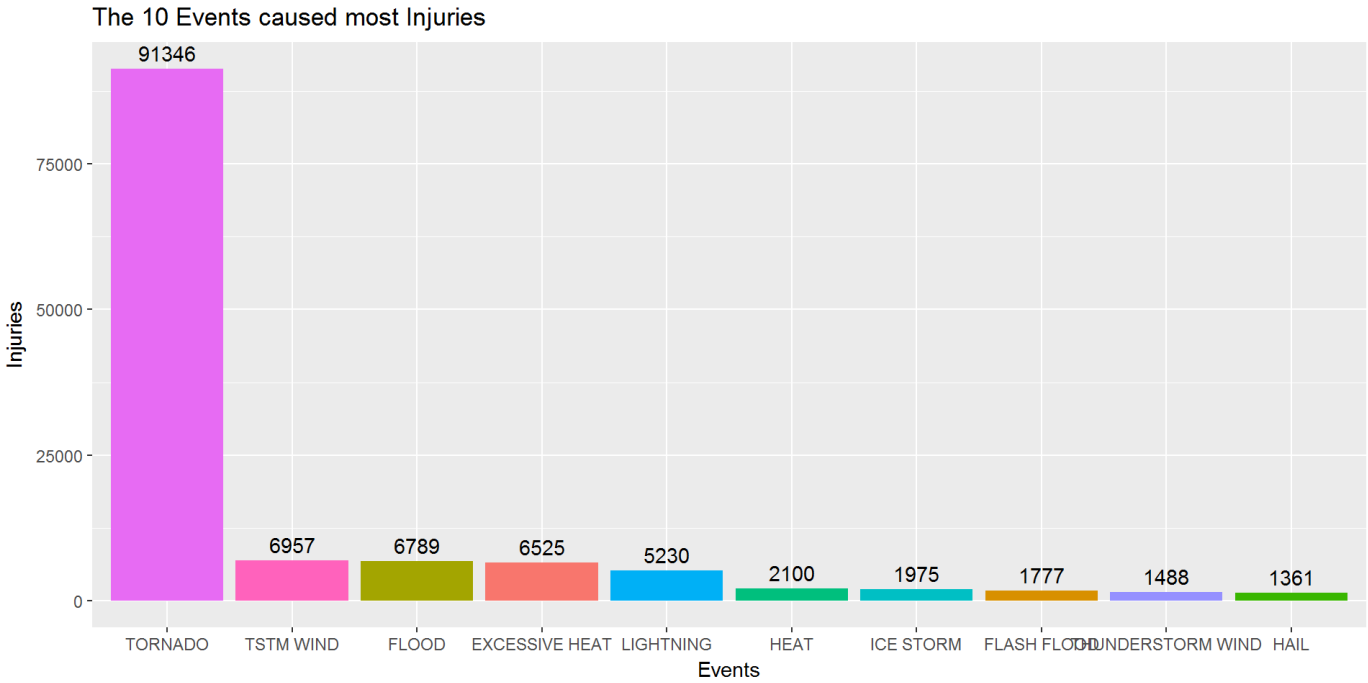
Injuries calculated by type:

```
inj <- aggregate(INJURIES ~ EVTYPE, data = stormData, sum)
inj <- inj[inj$INJURIES > 0, ]
inj <- inj[order(inj$INJURIES, decreasing = T), ]
head(inj, 10)
```

```
##                EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

Result:

```
ggplot(inj[1:10,], aes(reorder(EVTYPE, -INJURIES), INJURIES, fill = EVTYPE)) +
      geom_bar(stat = "identity") +
      geom_text(aes(label = INJURIES), vjust = -0.5, colour = "black") +
      labs(title = "The 10 Events caused most Injuries", y = "Injuries", x = "Events"
) +
      scale_fill_discrete(guide = FALSE)
```



The events caused both major fatalities and injuries

```
intersect(fatal[1:10, 1], inj[1:10,1 ])
```

```
## [1] "TORNADO"         "EXCESSIVE HEAT" "FLASH FLOOD"     "HEAT"
## [5] "LIGHTNING"       "TSTM WIND"      "FLOOD"
```

For the second qestion we need to calculate the economic cost of the storm events:

```
economicDamage <- aggregate(CROPDMG + PROPDMG ~ EVTYPE, data = stormData, sum)
economicDamage <- economicDamage[order(economicDamage$`CROPDMG + PROPDMG`, decreasing
= T),]
head(economicDamage, 10)
```

```
##                 EVTYPE CROPDMG + PROPDMG
## 834            TORNADO          3312276.7
## 153        FLASH FLOOD          1599325.1
## 856          TSTM WIND          1445168.2
## 244               HAIL          1268289.7
## 170              FLOOD          1067976.4
## 760  THUNDERSTORM WIND           943635.6
## 464          LIGHTNING           606932.4
## 786 THUNDERSTORM WINDS           464978.1
## 359          HIGH WIND           342014.8
## 972       WINTER STORM           134699.6
```

Result:

```
ggplot(economicDamage[1:10,], aes(reorder(EVTYPE, -`CROPDMG + PROPDMG`), `CROPDMG + P
ROPDMG`, fill = EVTYPE)) +
      geom_bar(stat = "identity") +
      geom_text(aes(label = `CROPDMG + PROPDMG`), vjust = 1.2, colour = "white", size
= 3.5) +
      labs(title = "The 10 Events caused most economic damage in $", y = "Cost", x =
"Events") +
      scale_fill_discrete(guide = FALSE)
```