Synopsis

Based on the data analysis - across the United States, tornado are most harmful with respect to population health. - across the United States floods have the greatest economic consequences

Data Processing

This section provides the data processing activities done to the data set.

```
library(ggplot2)

# read the dat file for storm data
data <- read.csv("repdata-data-StormData.csv")

# get familiar with the data
names(data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

```
head(data,n=2)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1 4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14   100 3   0          0
## 2         NA         0                       2   150 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
```

```
summary(data)
```

```
##     STATE__                  BGN_DATE             BGN_TIME
##  Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163
##  1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350
##  Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261
##  Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891
##  3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703
##  Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700
##                 (Other)          :895866   (Other)    :857229
##    TIME_ZONE          COUNTY           COUNTYNAME         STATE
##  CST    :547493   Min.   :  0.0   JEFFERSON :  7840   TX     : 83728
##  EST    :245558   1st Qu.: 31.0   WASHINGTON:  7603   KS     : 53440
##  MST    : 68390   Median : 75.0   JACKSON   :  6660   OK     : 46802
##  PST    : 28302   Mean   :100.6   FRANKLIN  :  6256   MO     : 35648
##  AST    :  6360   3rd Qu.:131.0   LINCOLN   :  5937   IA     : 31069
##  HST    :  2563   Max.   :873.0   MADISON   :  5632   NE     : 30271
##  (Other):  3631                   (Other)   :862369   (Other):621339
##                EVTYPE         BGN_RANGE           BGN_AZI
##  HAIL             :288661   Min.   :   0.000          :547332
##  TSTM WIND        :219940   1st Qu.:   0.000   N      : 86752
##  THUNDERSTORM WIND: 82563   Median :   0.000   W      : 38446
##  TORNADO          : 60652   Mean   :   1.484   S      : 37558
##  FLASH FLOOD      : 54277   3rd Qu.:   1.000   E      : 33178
##  FLOOD            : 25326   Max.   :3749.000   NW     : 24041
##  (Other)          :170878                      (Other):134990
##         BGN_LOCATI                  END_DATE             END_TIME
##              :287743                     :243411             :238978
##  COUNTYWIDE   : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802
##  Countywide   :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314
##  SPRINGFIELD  :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104
##  SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483
##  NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184
##  (Other)      :591444   (Other)          :653450   (Other)    :622432
##    COUNTY_END COUNTYENDN       END_RANGE           END_AZI
##  Min.   :0    Mode:logical   Min.   :  0.0000          :724837
##  1st Qu.:0    NA's:902297    1st Qu.:  0.0000   N      : 28082
##  Median :0                   Median :  0.0000   S      : 22510
##  Mean   :0                   Mean   :  0.9862   W      : 20119
##  3rd Qu.:0                   3rd Qu.:  0.0000   E      : 20047
##  Max.   :0                   Max.   :925.0000   NE     : 14606
##                                                 (Other): 72096
##            END_LOCATI         LENGTH              WIDTH
##                 :499225   Min.   :   0.0000   Min.   :   0.000
##  COUNTYWIDE     : 19731   1st Qu.:   0.0000   1st Qu.:   0.000
##  SOUTH PORTION  :   833   Median :   0.0000   Median :   0.000
##  NORTH PORTION  :   780   Mean   :   0.2301   Mean   :   7.503
##  CENTRAL PORTION:   617   3rd Qu.:   0.0000   3rd Qu.:   0.000
##  SPRINGFIELD    :   575   Max.   :2315.0000   Max.   :4400.000
```

```
##  (Other)        :380536
##         F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG           PROPDMGEXP        CROPDMG           CROPDMGEXP
##  Min.   :   0.00          :465934   Min.   :  0.000          :618413
##  1st Qu.:   0.00   K      :424665   1st Qu.:  0.000   K      :281832
##  Median :   0.00   M      : 11330   Median :  0.000   M      :  1994
##  Mean   :  12.06   0      :   216   Mean   :  1.527   k      :    21
##  3rd Qu.:   0.50   B      :    40   3rd Qu.:  0.000   0      :    19
##  Max.   :5000.00   5      :    28   Max.   :990.000   B      :     9
##                    (Other):    84                     (Other):     9
##       WFO                                         STATEOFFIC
##         :142069                                        :248769
##  OUN    : 17393   TEXAS, North                          : 12193
##  JAN    : 13889   ARKANSAS, Central and North Central: 11738
##  LWX    : 13174   IOWA, Central                         : 11345
##  PHI    : 12551   KANSAS, Southwest                     : 11212
##  TSA    : 12483   GEORGIA, North and Central            : 11120
##  (Other):690738   (Other)                               :595920
##
ZONENAMES
##
:594029
##
:205988
##  GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
:   639
##  GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
:   592
##  JEFFERSON - JEFFERSON
:   303
##  MADISON - MADISON
:   302
##  (Other)
:100444
##     LATITUDE      LONGITUDE        LATITUDE_E     LONGITUDE_
##  Min.   :   0   Min.   :-14451   Min.   :   0   Min.   :-14455
##  1st Qu.:2802   1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0
##  Median :3540   Median :  8707   Median :   0   Median :     0
##  Mean   :2875   Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:4019   3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   :9706   Max.   : 17124   Max.   :9706   Max.   :106220
##  NA's   :47                      NA's   :40
##                                          REMARKS           REFNUM
```

```
##                                                 :287433   Min.   :     1
##                                                 : 24013   1st Qu.:225575
##  Trees down.\n                                  :  1110   Median :451149
##  Several trees were blown down.\n               :   569   Mean   :451149
##  Trees were downed.\n                           :   446   3rd Qu.:676723
##  Large trees and power lines were blown down.\n:   432   Max.   :902297
##  (Other)                                        :588294
```

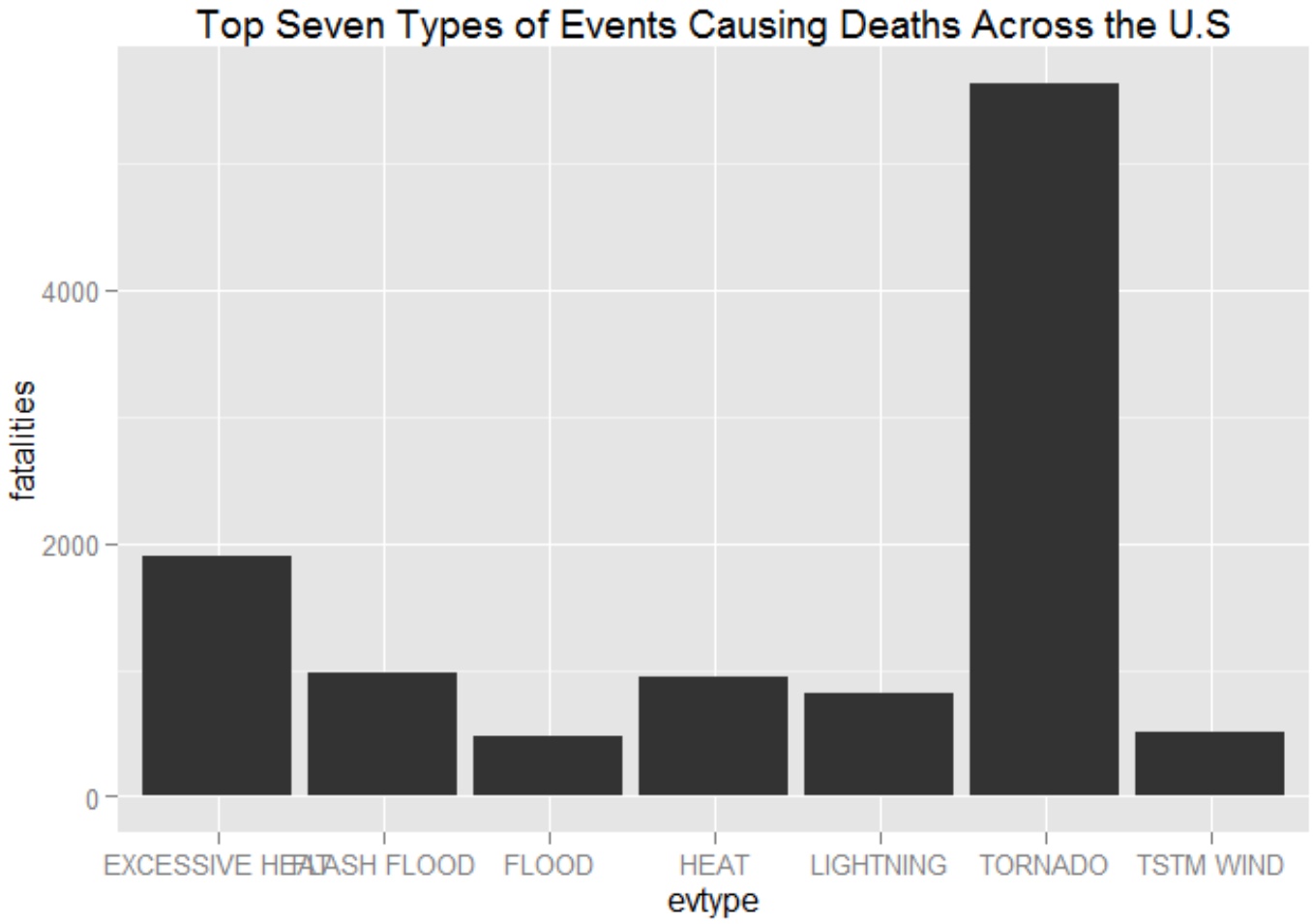
```
table(complete.cases(data))
```

```
##
##  FALSE
## 902297
```

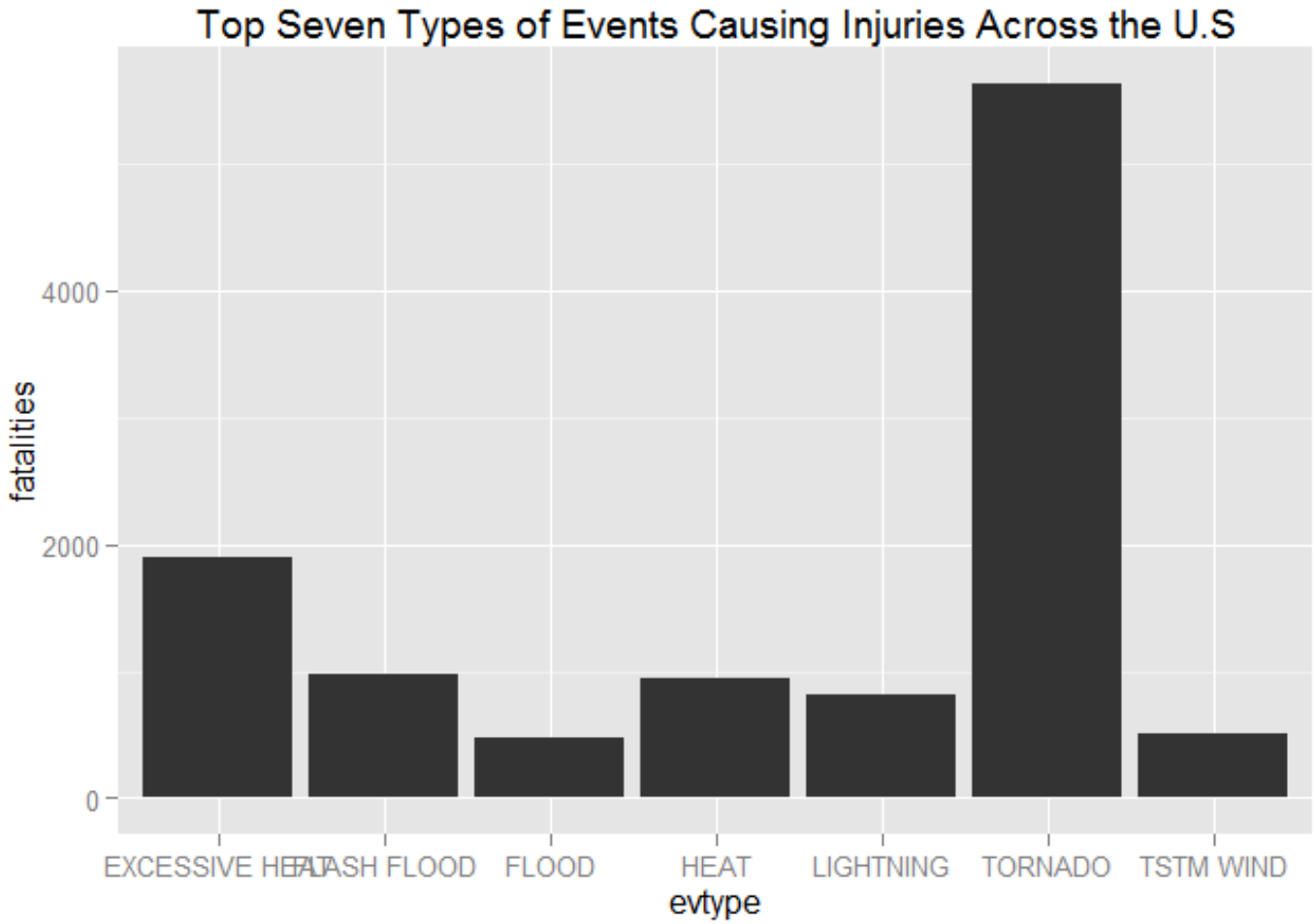
```
# get colum names to lowercase
names(data) <- tolower(names(data))
```

Now lets find the population harm for events

```
# harmful for population is fatalaties and injuries ... aggregate & order them
fatal <- aggregate(fatalities ~ evtype, data = data, FUN = sum)
fatal <- fatal[order(fatal$fatalities, decreasing = T), ]
ggplot(fatal[1:7, ], aes(evtype, fatalities)) + geom_bar(stat = "identity") + ggtitle("To
p Seven Types of Events Causing Deaths Across the U.S")
```




```
injury <- aggregate(injuries ~ evtype, data = data, FUN = sum)
injury <- injury[order(injury$injuries, decreasing = T), ]
ggplot(fatal[1:7, ], aes(evtype, fatalities)) + geom_bar(stat = "identity") + ggtitle("To
p Seven Types of Events Causing Injuries Across the U.S")
```



This shows that top event for population harm is Tornado

Now lets move to the economic damage

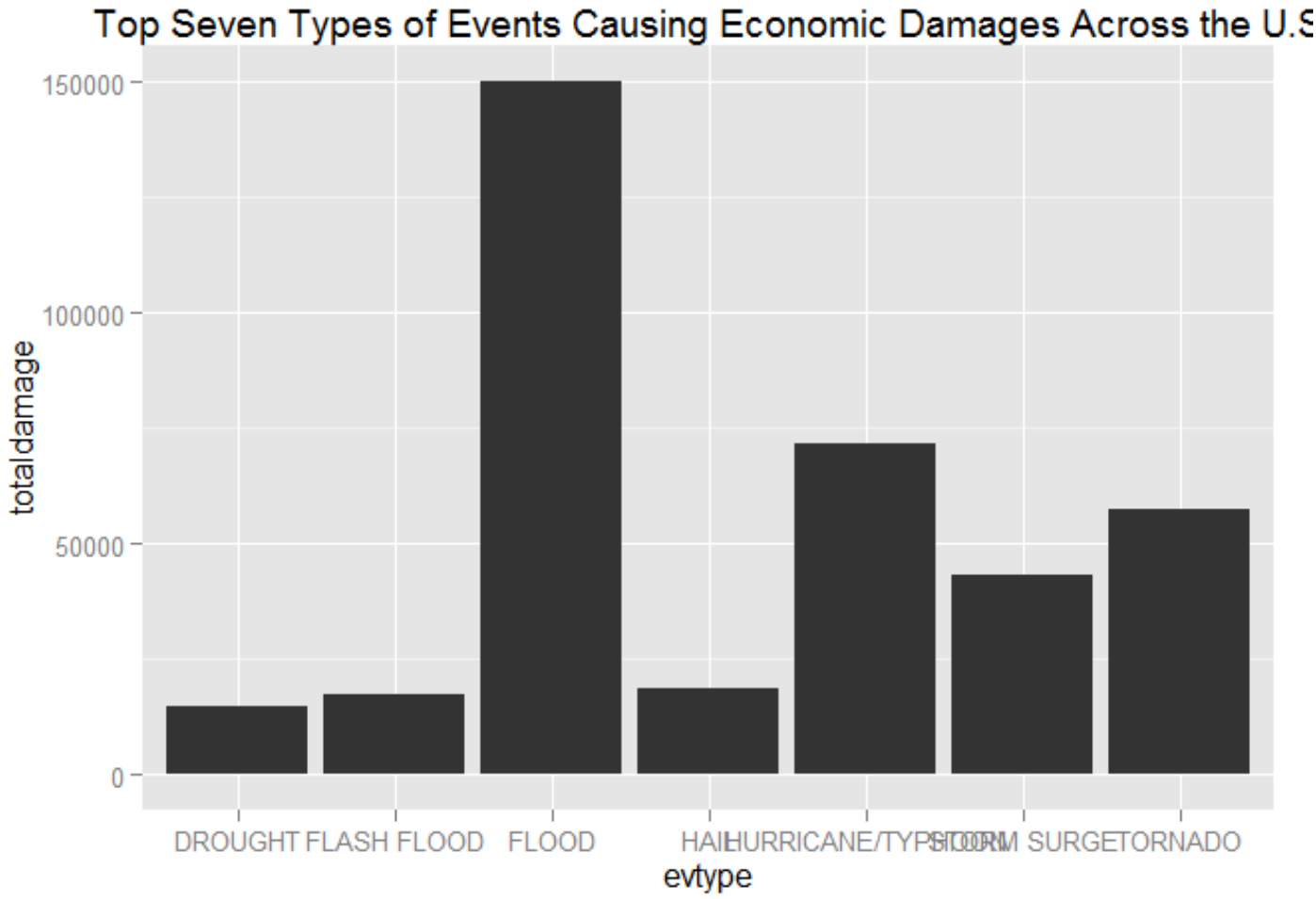
```
# check units
table(data$propdmgexp)
```

```
##
##             -      ?      +      0      1      2      3      4      5
## 465934      1      8      5    216     25     13      4      4     28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665      7  11330
```

```
table(data$cropdmgexp)
```

```
##
##             ?      0      2      B      k      K      m      M
## 618413      7     19      1      9     21 281832      1   1994
```

```
# normalise

pd <- data$propdmg
pde <- data$propdmgexp
cd <- data$cropdmg
cde <- data$cropdmgexp

pd[pde %in% "B"] <- pd[pde %in% "B"] * 1000
pd[pde %in% c("M", "m")] <- pd[pde %in% c("M", "m")] * 1
pd[pde %in% c("K")] <- pd[pde %in% c("K")] * 0.001
pd[pde %in% c("H", "h")] <- pd[pde %in% c("H", "h")] * 1e-04
pd[!(pde %in% c("B", "M", "m", "K", "H", "h"))] <- pd[!(pde %in% c("B", "M",
    "m", "K", "H", "h"))] * 1e-06

cd[cde %in% "B"] <- cd[cde %in% "B"] * 1000
cd[cde %in% c("M", "m")] <- cd[cde %in% c("M", "m")] * 1
cd[cde %in% c("K", "k")] <- cd[cde %in% c("K", "k")] * 0.001
cd[!(cde %in% c("B", "M", "m", "K", "k"))] <- cd[!(cde %in% c("B", "M", "m",
    "K", "k"))] * 1e-06

totaldamage <- cd + pd
edt <- aggregate(totaldamage ~ data$evtype, FUN = sum)
oedt <- edt[order(edt$totaldamage, decreasing = T), ]
names(oedt)[1] <- "evtype"

ggplot(oedt[1:7, ], aes(evtype, totaldamage)) + geom_bar(stat = "identity") + ggtitle("To
p Seven Types of Events Causing Economic Damages Across the U.S")
```



Results

Based on the data analysis - across the United States, tornado are most harmful with respect to population health (both injury & fatal. - across the United States floods have the greatest economic consequences for property & agriculture