Reproducible Research Assessment 2: Impact of severe weather events on US population health and economy

Rosalind

21 December, 2014

Synopsis

Aim: The overarching aim of this report is to understand the impact of severe weather events on the US population health and economy.

Data source: Findings from this report is based on the NOAA Storm Database (period: 1950 to 2011)

Settings

```
#Make code visible, set working directory and read library
echo = TRUE
setwd("~/RepData_PeerAssessment2")
library(ggplot2)
library(plyr)
library(utils)
require(gridExtra)
```

```
## Loading required package: gridExtra
## Loading required package: grid
```

Data exploration

```
data <- read.csv(file="data/StormData.csv", header=T, sep=",", quote="", as.is=TRUE)
summary(data)
```

```
##   X.STATE__.        X.BGN_DATE.        X.BGN_TIME.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.TIME_ZONE.        X.COUNTY.         X.COUNTYNAME.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##    X.STATE.          X.EVTYPE.         X.BGN_RANGE.
```

```
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   X.BGN_AZI.         X.BGN_LOCATI.      X.END_DATE.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.END_TIME.        X.COUNTY_END.      X.COUNTYENDN.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.END_RANGE.        X.END_AZI.        X.END_LOCATI.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   X.LENGTH.            X.WIDTH.            X.F.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##     X.MAG.          X.FATALITIES.      X.INJURIES.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   X.PROPDMG.         X.PROPDMGEXP.       X.CROPDMG.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.CROPDMGEXP.         X.WFO.          X.STATEOFFIC.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.ZONENAMES.       X.LATITUDE.        X.LONGITUDE.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  X.LATITUDE_E.      X.LONGITUDE_.       X.REMARKS.
##  Length:1773320     Length:1773320     Length:1773320
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   X.REFNUM.
##  Length:1773320
##  Class :character
##  Mode  :character
```

```
if (!"StormData" %in% ls()) {
    Data <- read.csv("data/StormData.csv", sep = ",")
}
dim(Data)
```

```
## [1] 902297     37
```

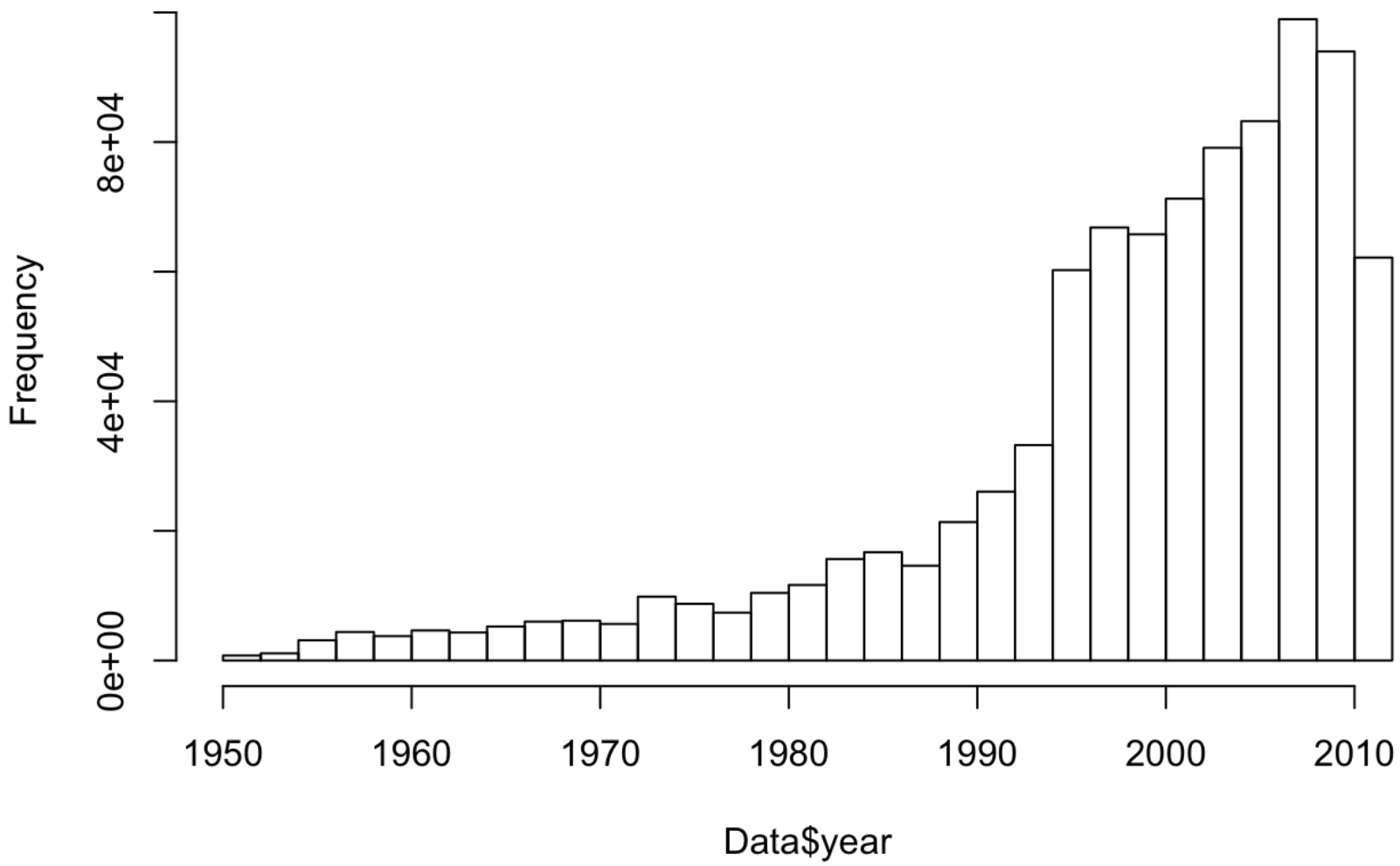
```
head(Data, n = 2)
```

```
##   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
```

The total number of rows and columns are 902297 and 37 respectively.

The events started from 1950 and end in 2011. There were few recorded events in the earlier years due to the lack of good records. While the records were more complete in the recent years.

```
if (dim(Data)[2] == 37) {
    Data$year <- as.numeric(format(as.Date(Data$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"),
"%Y"))
}
hist(Data$year, breaks = 30)
```



Based on the graph, the number of events tracked start to increase significantly from 1995. Hence, 1990 to 2011 will be use to make most of the data.

```
storm <- Data[Data$year >= 1995, ]
dim(storm)
```

```
## [1] 681500     38
```

With this, the total number of rows and column 681500 and 38 respectively.

Impact on Population Health

To identify the no. of **fatalities** and **injuries** that are caused by the severe weather events, we list the top 10 most severe types of weather events.

```
sortHelper <- function(fieldName, top = 10, dataset = Data) {
    index <- which(colnames(dataset) == fieldName)
    field <- aggregate(dataset[, index], by = list(dataset$EVTYPE), FUN = "sum")
    names(field) <- c("EVTYPE", fieldName)
    field <- arrange(field, field[, 2], decreasing = T)
    field <- head(field, n = top)
    field <- within(field, EVTYPE <- factor(x = EVTYPE, levels = field$EVTYPE))
    return(field)
}

fatalities <- sortHelper("FATALITIES", dataset = storm)
injuries <- sortHelper("INJURIES", dataset = storm)
```

Impact on Economy

We will convert the **property damage** and **crop damage** data into comparable numerical forms. Both `PROPDMGEXP` and `CROPDMGEXP` columns record a multiplier for each observation where we have Hundred (H), Thousand (K), Million (M) and Billion (B).

```
convertHelper <- function(dataset = storm, fieldName, newFieldName) {
    totalLen <- dim(dataset)[2]
    index <- which(colnames(dataset) == fieldName)
    dataset[, index] <- as.character(dataset[, index])
    logic <- !is.na(toupper(dataset[, index]))
    dataset[logic & toupper(dataset[, index]) == "B", index] <- "9"
    dataset[logic & toupper(dataset[, index]) == "M", index] <- "6"
    dataset[logic & toupper(dataset[, index]) == "K", index] <- "3"
    dataset[logic & toupper(dataset[, index]) == "H", index] <- "2"
    dataset[logic & toupper(dataset[, index]) == "", index] <- "0"
    dataset[, index] <- as.numeric(dataset[, index])
    dataset[is.na(dataset[, index]), index] <- 0
    dataset <- cbind(dataset, dataset[, index - 1] * 10^dataset[, index])
    names(dataset)[totalLen + 1] <- newFieldName
    return(dataset)
}

storm <- convertHelper(storm, "PROPDMGEXP", "propertyDamage")
```

```
## Warning in convertHelper(storm, "PROPDMGEXP", "propertyDamage"): NAs
## introduced by coercion
```

```
storm <- convertHelper(storm, "CROPDMGEXP", "cropDamage")
```

```
## Warning in convertHelper(storm, "CROPDMGEXP", "cropDamage"): NAs
## introduced by coercion
```

```
names(storm)
```

```
##  [1] "STATE__"        "BGN_DATE"       "BGN_TIME"       "TIME_ZONE"
##  [5] "COUNTY"         "COUNTYNAME"     "STATE"          "EVTYPE"
##  [9] "BGN_RANGE"      "BGN_AZI"        "BGN_LOCATI"     "END_DATE"
## [13] "END_TIME"       "COUNTY_END"     "COUNTYENDN"     "END_RANGE"
## [17] "END_AZI"        "END_LOCATI"     "LENGTH"         "WIDTH"
## [21] "F"              "MAG"            "FATALITIES"     "INJURIES"
## [25] "PROPDMG"        "PROPDMGEXP"     "CROPDMG"        "CROPDMGEXP"
## [29] "WFO"            "STATEOFFIC"     "ZONENAMES"      "LATITUDE"
## [33] "LONGITUDE"      "LATITUDE_E"     "LONGITUDE_"     "REMARKS"
## [37] "REFNUM"         "year"           "propertyDamage" "cropDamage"
```

```
options(scipen=999)
property <- sortHelper("propertyDamage", dataset = storm)
crop <- sortHelper("cropDamage", dataset = storm)
```

Findings

As for the impact on public health, we have the lists of severe weather events by the no. of people badly injured.

```
fatalities
```

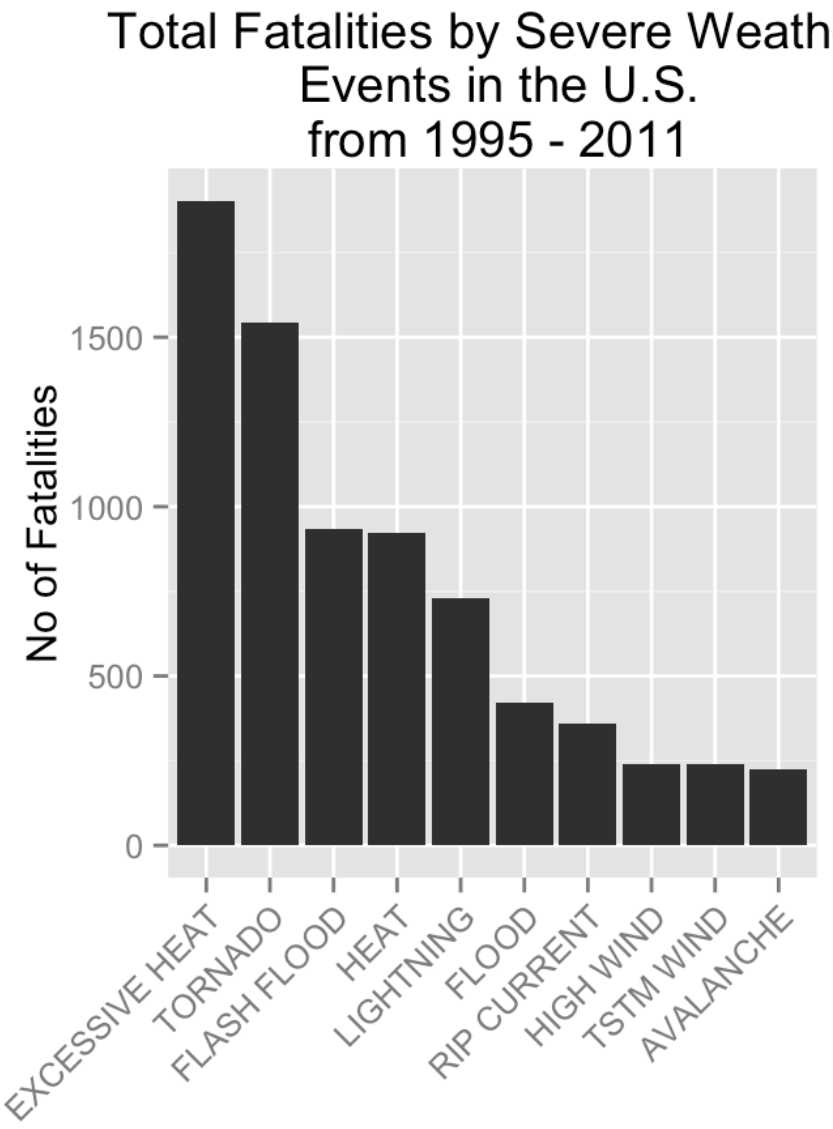
```
##            EVTYPE FATALITIES
## 1  EXCESSIVE HEAT       1903
## 2         TORNADO       1545
## 3     FLASH FLOOD        934
## 4            HEAT        924
## 5       LIGHTNING        729
## 6           FLOOD        423
## 7     RIP CURRENT        360
## 8       HIGH WIND        241
## 9       TSTM WIND        241
## 10      AVALANCHE        223
```

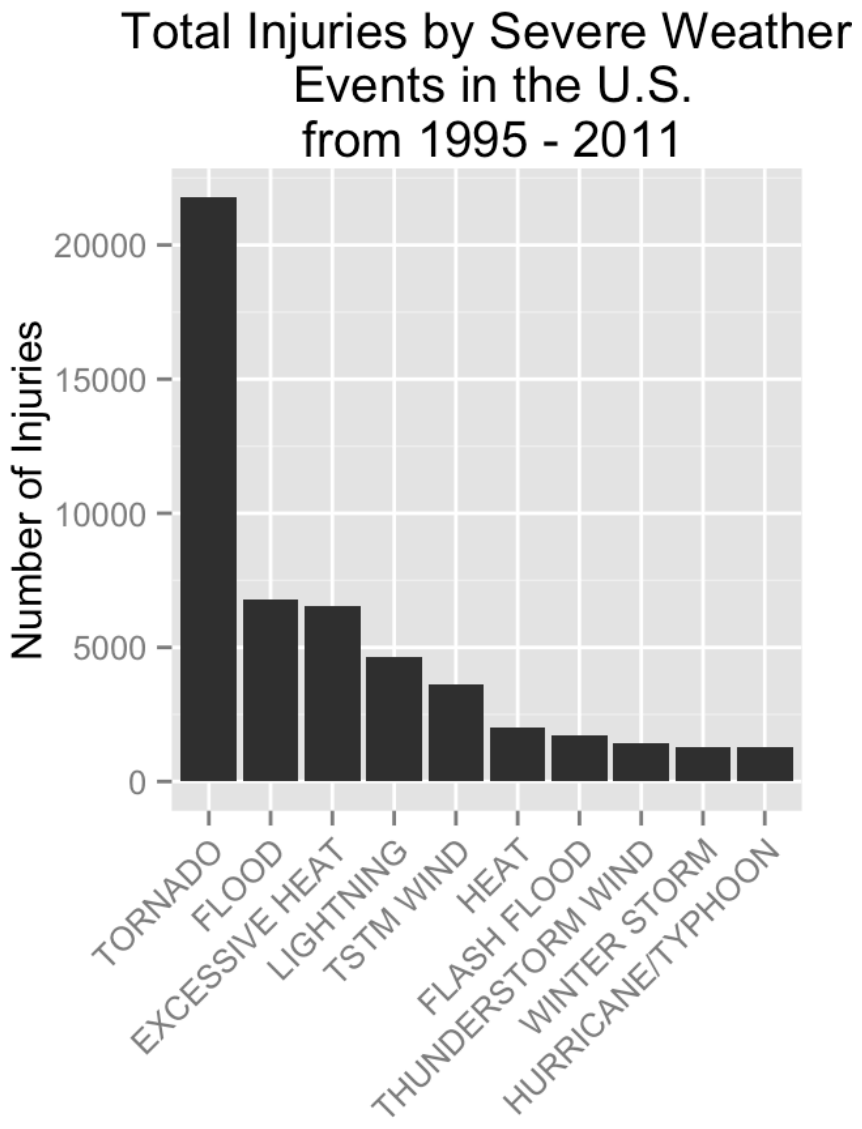
```
injuries
```

```
##                EVTYPE INJURIES
## 1             TORNADO    21765
## 2               FLOOD     6769
## 3      EXCESSIVE HEAT     6525
## 4           LIGHTNING     4631
## 5           TSTM WIND     3630
## 6                HEAT     2030
## 7         FLASH FLOOD     1734
## 8   THUNDERSTORM WIND     1426
## 9        WINTER STORM     1298
## 10 HURRICANE/TYPHOON     1275
```

Below illustrates the total fatalities and injuries caused by the severe weather events.

```
fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", bi
nwidth = 1) +
    scale_y_continuous("No of Fatalities") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Fatalities by Severe Weather\n Events in the U.S.\n from 1995 - 2011"
)
injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, geom = "bar", binwidth
= 1) +
    scale_y_continuous("Number of Injuries") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Injuries by Severe Weather\n Events in the U.S.\n from 1995 - 2011")
grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)
```





Excessive heat and **tornado** caused the highest fatalities, whereas **tornado** caused the most injuries in the US from 1995 to 2011.

As for the impact on economy, we have got two sorted lists below by the amount of money cost by damages.

```
property
```

```
##               EVTYPE propertyDamage
## 1              FLOOD   144022037057
## 2  HURRICANE/TYPHOON    69305840000
## 3        STORM SURGE    43193536000
## 4            TORNADO    24935939545
## 5        FLASH FLOOD    16047794571
## 6               HAIL    15048722103
## 7          HURRICANE    11812819010
## 8     TROPICAL STORM     7653335550
## 9          HIGH WIND     5259785375
## 10          WILDFIRE     4759064000
```

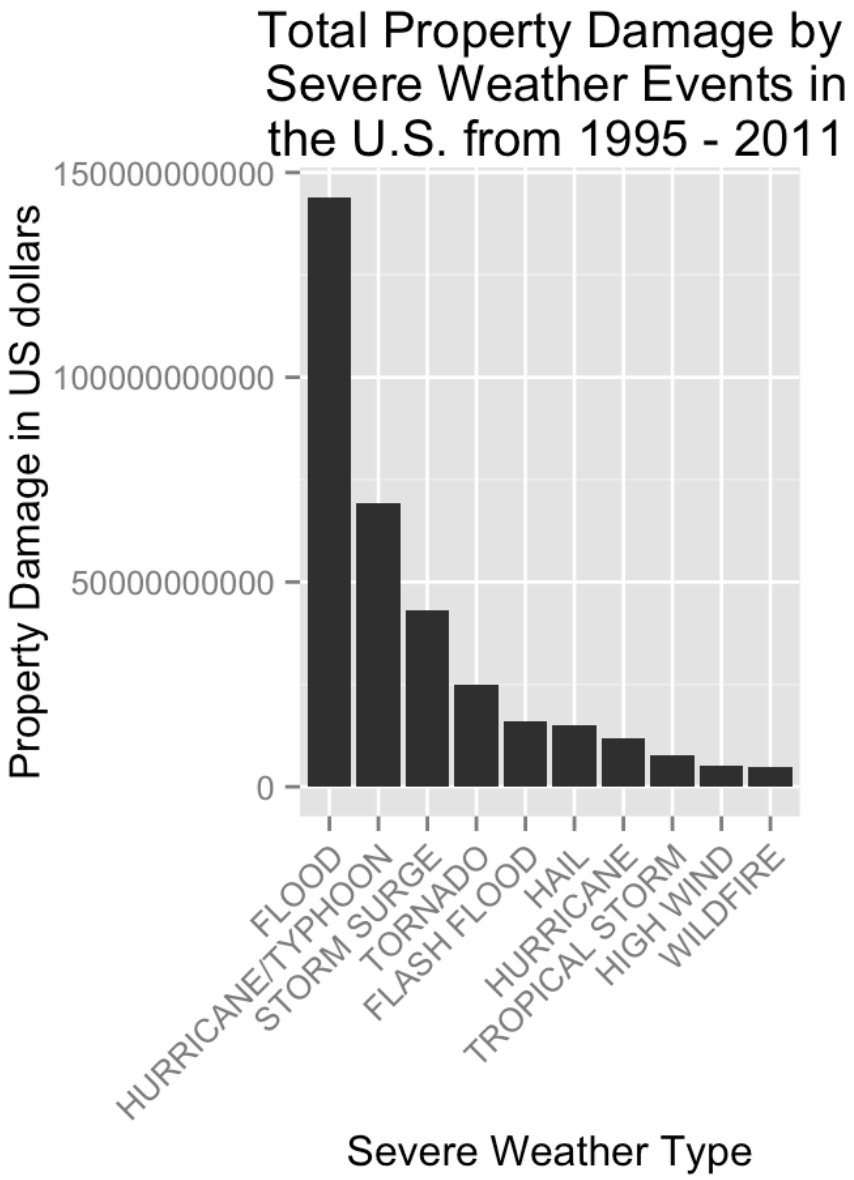
```
crop
```

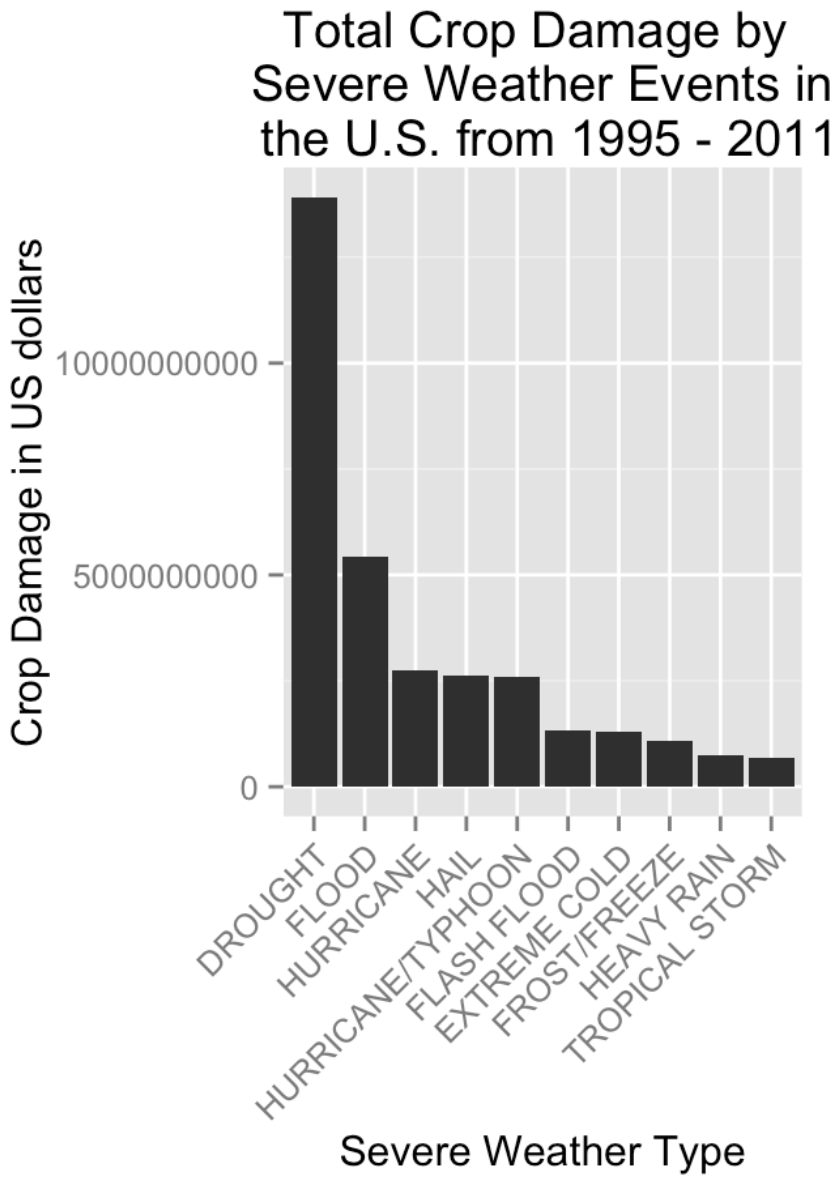
```
##               EVTYPE  cropDamage
## 1            DROUGHT 13922066000
## 2              FLOOD  5422810400
## 3          HURRICANE  2741410000
## 4               HAIL  2614127070
## 5  HURRICANE/TYPHOON  2607872800
## 6        FLASH FLOOD  1343915000
## 7       EXTREME COLD  1292473000
## 8       FROST/FREEZE  1094086000
## 9         HEAVY RAIN   728399800
## 10    TROPICAL STORM   677836000
```

Below illustrates the total property and crop damage affected by the severe weather events.

```
propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, geom = "bar", bi
nwidth = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Prope
rty Damage in US dollars")+
    xlab("Severe Weather Type") + ggtitle("Total Property Damage by\n Severe Weather Eve
nts in\n the U.S. from 1995 - 2011")

cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, geom = "bar", binwidth = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Crop
Damage in US dollars") +
    xlab("Severe Weather Type") + ggtitle("Total Crop Damage by \nSevere Weather Events
in\n the U.S. from 1995 - 2011")
grid.arrange(propertyPlot, cropPlot, ncol = 2)
```





Both **flood** and **hurricane/typhoon** caused the most property damage. While**drought** and **flood** caused the most crop damage from 1995 to 2011.

Conclusion

Based on the findings, **excessive heat** and **tornado** had the greatest impact on the population health, whereas **flood**, **drought**, and **hurricane/typhoon** had the greatest economic impact.