Course Project 2

Simon Coburg

02/03/2021

Analysing the NOAA storm database - A summary of health and economic impacts in the United States from 1950 - 2011.

Severe weather events cause major issues for public health and have acute economic effects. Such severe events result in fatalities, injuries, and generate massive property damages.

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 from the years 1950 to 2011, including when and where they took place with estimates of any fatalities, injuries, and property damage that occured.

The data analysis addresses the following questions:

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

Data processing

Loading libraries, importing dataset

The libraries used for this assessment are:

```r
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
```

```r
library(ggplot2)
```

The dataset can be downloaded as a comma-separated-value file 47Mb compressed by the *bzip2* algorithm.

Documentation of the dataset are available from:

- National Weather Service Storm Data Documentation
- National Climatic Data Center Storm Events FAQ

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

The following steps load the dataset:

```
#Load data
path <- getwd()
download.file(url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",    dest

dir() #Check if data present
```

```
## [1] "Courseproject2.R"               "README.md"
## [3] "RepData_PeerAssessment2.Rproj" "StormData.csv.bz2"
## [5] "StormData.html"                 "StormData.Rmd"
```

```
#Import dataset
StormData <- read.csv(bzfile("StormData.csv.bz2"), header = TRUE)

#Checking dataset
str(StormData)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ 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: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

The dataset consists of **37 columns** (variables) and **902,297 rows** (records).

Extracting variables of interest for analysis of weather impact on health and economy

From the list of variables in *StormData*, the following columns of interest were selected:

Health variables:

- FATALITIES: approx. number of deaths
- INJURIES: approx. number of injuries

Economic variables:

- PROPDMG: approx. property damags
- PROPDMGEXP: the units for property damage value
- CROPDMG: approx. crop damages
- CROPDMGEXP: the units for crop damage value

Events - target variable:

- EVTYPE: weather event (Tornados, Wind, Snow, Flood, etc..)

Health impact

To evaluate the impact on health, the total fatalities and the total injuries for each event type (EVTYPE) are calculated. The codes for this calculation are shown as follows.

```
#Fatalities
storm.fatalities <- StormData %>% select(EVTYPE, FATALITIES) %>%
                    group_by(EVTYPE) %>%
                    summarise(total = sum(FATALITIES)) %>%
                    arrange(-total)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(storm.fatalities, 10)
```

```
## # A tibble: 10 x 2
##    EVTYPE         total
##    <chr>          <dbl>
##  1 TORNADO         5633
##  2 EXCESSIVE HEAT  1903
##  3 FLASH FLOOD      978
##  4 HEAT             937
##  5 LIGHTNING        816
##  6 TSTM WIND        504
##  7 FLOOD            470
##  8 RIP CURRENT      368
##  9 HIGH WIND        248
## 10 AVALANCHE        224
```

```
#Injuries
storm.injuries <- StormData %>% select(EVTYPE, INJURIES) %>%
                  group_by(EVTYPE) %>%
```

```r
                   summarise(total = sum(INJURIES)) %>%
                   arrange(-total)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
head(storm.injuries, 10)
```

```
## # A tibble: 10 x 2
##    EVTYPE            total
##    <chr>             <dbl>
##  1 TORNADO           91346
##  2 TSTM WIND          6957
##  3 FLOOD              6789
##  4 EXCESSIVE HEAT     6525
##  5 LIGHTNING          5230
##  6 HEAT               2100
##  7 ICE STORM          1975
##  8 FLASH FLOOD        1777
##  9 THUNDERSTORM WIND  1488
## 10 HAIL               1361
```

Economic impact

To evaluate the impact on the economy, the property and the crop damages for each event type (EVTYPE) are calculated. The codes for this calculation are shown as follows:

```r
storm.damage <- StormData %>% select(EVTYPE, PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
head(storm.damage)
```

```
##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO    25.0          K       0
## 2 TORNADO     2.5          K       0
## 3 TORNADO    25.0          K       0
## 4 TORNADO     2.5          K       0
## 5 TORNADO     2.5          K       0
## 6 TORNADO     2.5          K       0
```

Assessing the parameters for PROPDMGEXP and CROPDMGEXP:

```r
sort(table(storm.damage$PROPDMGEXP), decreasing = TRUE)[1:10]
```

```
##
##             K      M      0      B      5      1      2      ?      m
## 465934 424665  11330    216     40     28     25     13      8      7
```

```r
sort(table(storm.damage$CROPDMGEXP), decreasing = TRUE)[1:10]
```

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

The dataset contains of two types of economic impact; the property damage (PROPDMG) and crop damage (CROPDMG). The actual damage in USD is indicated by PROPDMGEXP and CROPDMGEXP parameters. Based on information from this source, the parameters in the PROPDMGEXP and CROPDMGEXP columns are interpreted as follows:

- blank -> *0
- (-) -> *0

- (?) -> *0
- (+) -> *1
- H, h -> hundreds (*100)
- K, k -> kilos (*1,000)
- M, m -> millions (*1,000,000)
- B,b -> billions (*1,000,000,000)

The total damage caused by each event type is calculated by the following codes:

```
#Transfrom character values to numeric values in PROPDMGEXP & CROPDMGEXP
storm.damage <- StormData %>% select(EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

symbols <- sort(unique(as.character(storm.damage$PROPDMGEXP)))
symbols
```

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

```
#Set a multiplier per symbol in the order of the symbols shown
multiplier <- c(0, 0, 0, 1, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10^9, 10^2, 10^2, 10^3, 10^6, 10^6)

#Create dataset for symbol values
storm.convert <- data.frame(symbols, multiplier)
storm.convert
```

```
##    symbols multiplier
## 1               0e+00
## 2        -      0e+00
## 3        ?      0e+00
## 4        +      1e+00
## 5        0      1e+01
## 6        1      1e+01
## 7        2      1e+01
## 8        3      1e+01
## 9        4      1e+01
## 10       5      1e+01
## 11       6      1e+01
## 12       7      1e+01
## 13       8      1e+01
## 14       B      1e+09
## 15       h      1e+02
## 16       H      1e+02
## 17       K      1e+03
## 18       m      1e+06
## 19       M      1e+06
```

```
#Fuse symbols with numeric values
storm.damage$property.multiplier <- storm.convert$multiplier[match(storm.damage$PROPDMGEXP,
                                      storm.convert$symbols)]
storm.damage$crop.multiplier <- storm.convert$multiplier[match(storm.damage$CROPDMGEXP,
                                      storm.convert$symbols)]

#Property damages
storm.property <- storm.damage %>%
                  mutate(DMG = PROPDMG*property.multiplier) %>%
                  select(EVTYPE, DMG) %>%
                  arrange(-DMG)
```

```r
head(storm.property)
```

```
##               EVTYPE       DMG
## 1              FLOOD 1.150e+11
## 2        STORM SURGE 3.130e+10
## 3  HURRICANE/TYPHOON 1.693e+10
## 4        STORM SURGE 1.126e+10
## 5  HURRICANE/TYPHOON 1.000e+10
## 6  HURRICANE/TYPHOON 7.350e+09
```

```r
#Crop damages
storm.crop <- storm.damage %>%
              mutate(DMG = CROPDMG*crop.multiplier) %>%
              select(EVTYPE, DMG) %>%
              arrange(-DMG)

head(storm.crop)
```

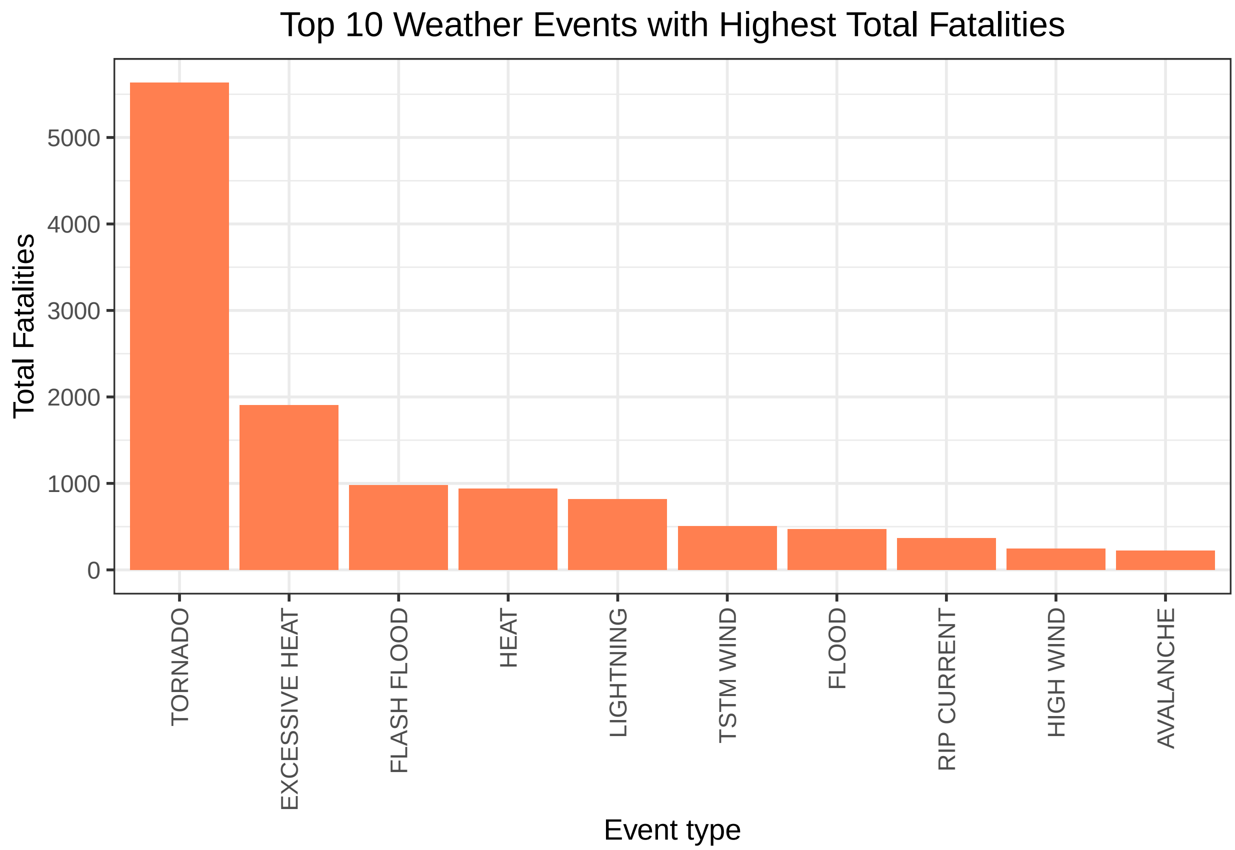
```
##               EVTYPE        DMG
## 1        RIVER FLOOD 5000000000
## 2          ICE STORM 5000000000
## 3  HURRICANE/TYPHOON 1510000000
## 4            DROUGHT 1000000000
## 5       EXTREME COLD  596000000
## 6            DROUGHT  578850000
```

Results

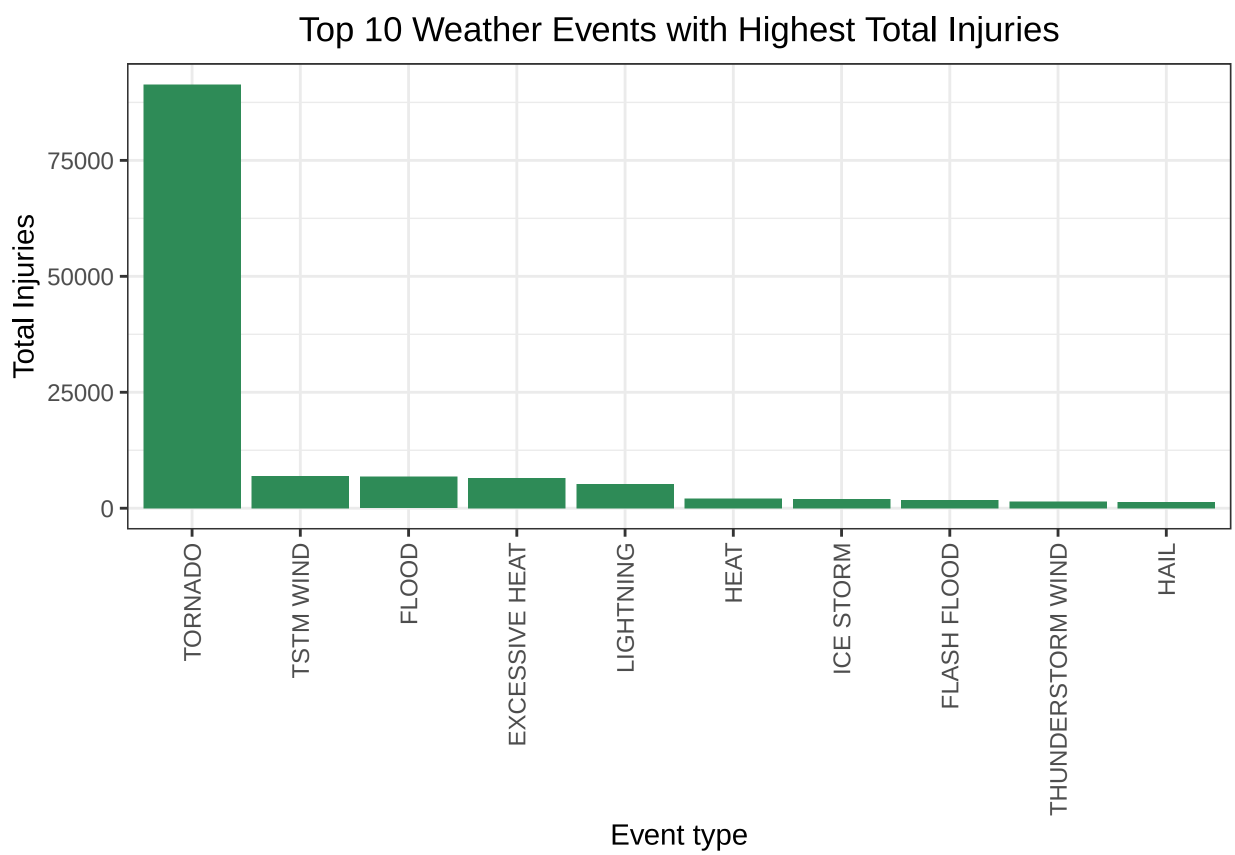
Health impact

The top 10 weather events that caused the highest fatalities and injuries are presented individually and as a combined dataset:

```r
#Fatalities
ggplot(storm.fatalities[1:10,], aes(x = reorder(EVTYPE, -total), y = total)) +
    geom_bar(stat="identity", fill = "coral") +
    scale_y_continuous(breaks=seq(0,6000,1000)) +
    ggtitle("Top 10 Weather Events with Highest Total Fatalities")+
    labs(x="Event type", y="Total Fatalities")+
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5)) +
    theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```




```
#Injuries
ggplot(storm.injuries[1:10,], aes(x = reorder(EVTYPE, -total), y = total)) +
    geom_bar(stat="identity", fill = "seagreen") +
    scale_y_continuous(breaks=seq(0,100000,25000)) +
    ggtitle("Top 10 Weather Events with Highest Total Injuries")+
    labs(x="Event type", y="Total Injuries") +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5)) +
    theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```

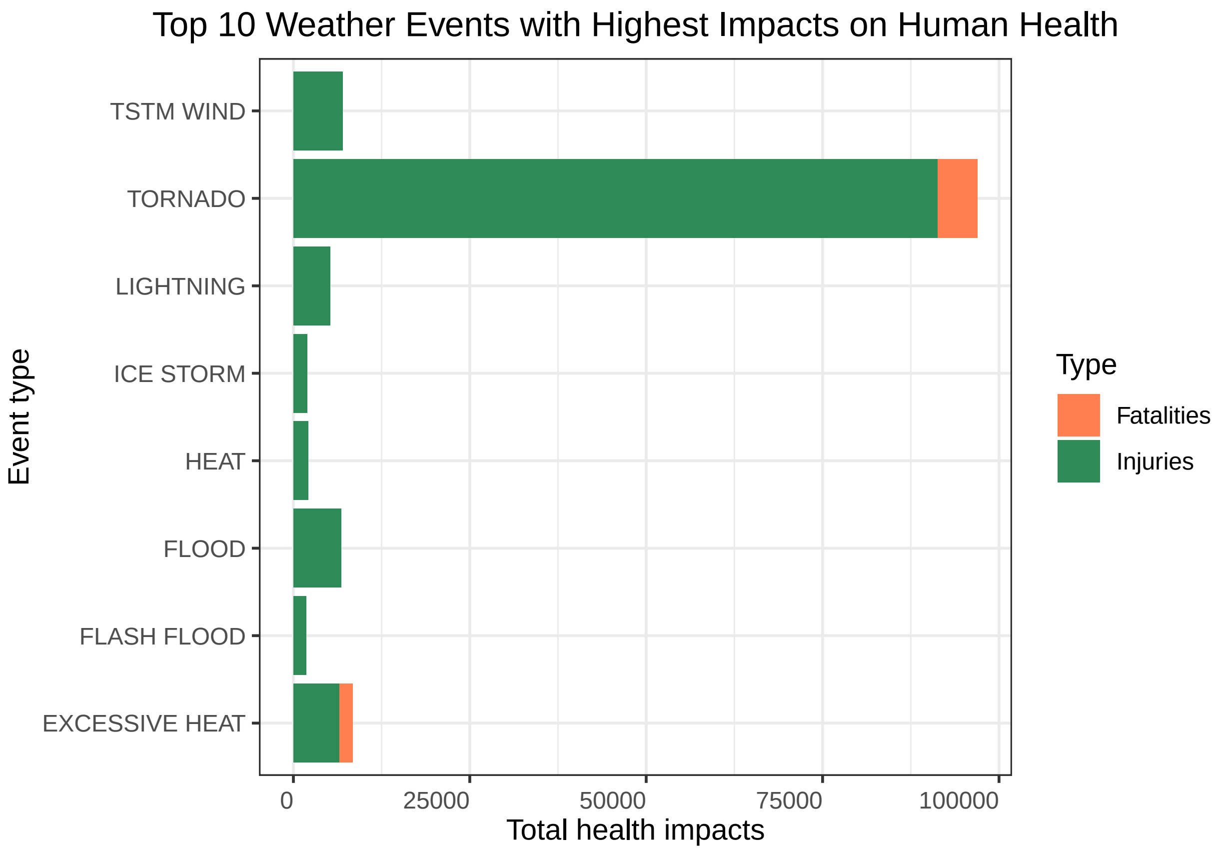



```r
#Combine fatalities & injuries
storm.fatalities$Type <- "Fatalities"
storm.injuries$Type <- "Injuries"

storm.health <- rbind(storm.injuries, storm.fatalities)

storm.health <- storm.health %>%
                arrange(-total, EVTYPE)


ggplot(storm.health[1:10,], aes(x = EVTYPE, y = total, fill = Type)) +
  coord_flip() +
  geom_bar(stat="identity") +
  scale_fill_manual(values = c("coral", "seagreen")) +
  ggtitle("Top 10 Weather Events with Highest Impacts on Human Health") +
  labs(x="Event type", y="Total health impacts") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle=0, vjust=0.5, hjust=1))
```

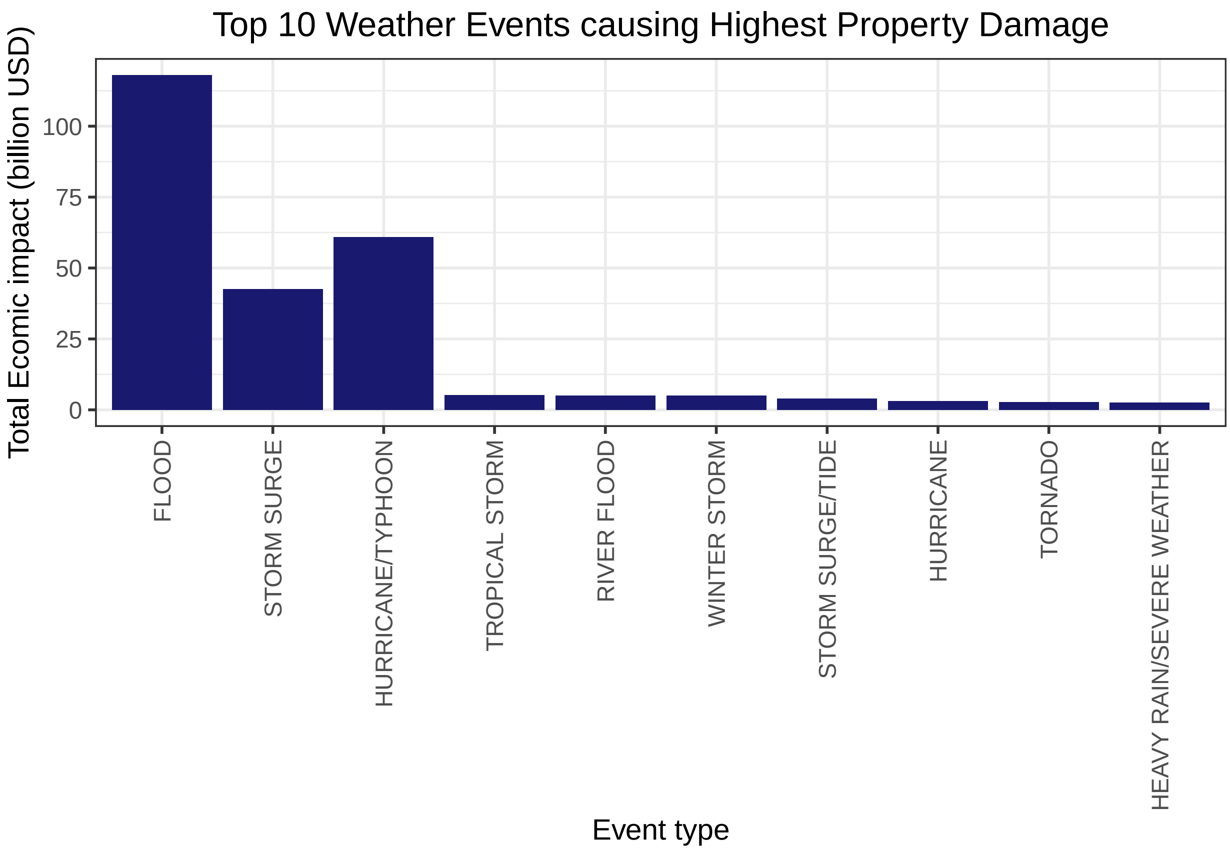


The figures show that most **Fatalities** and **Injuries** were caused by **Tornadoes**.

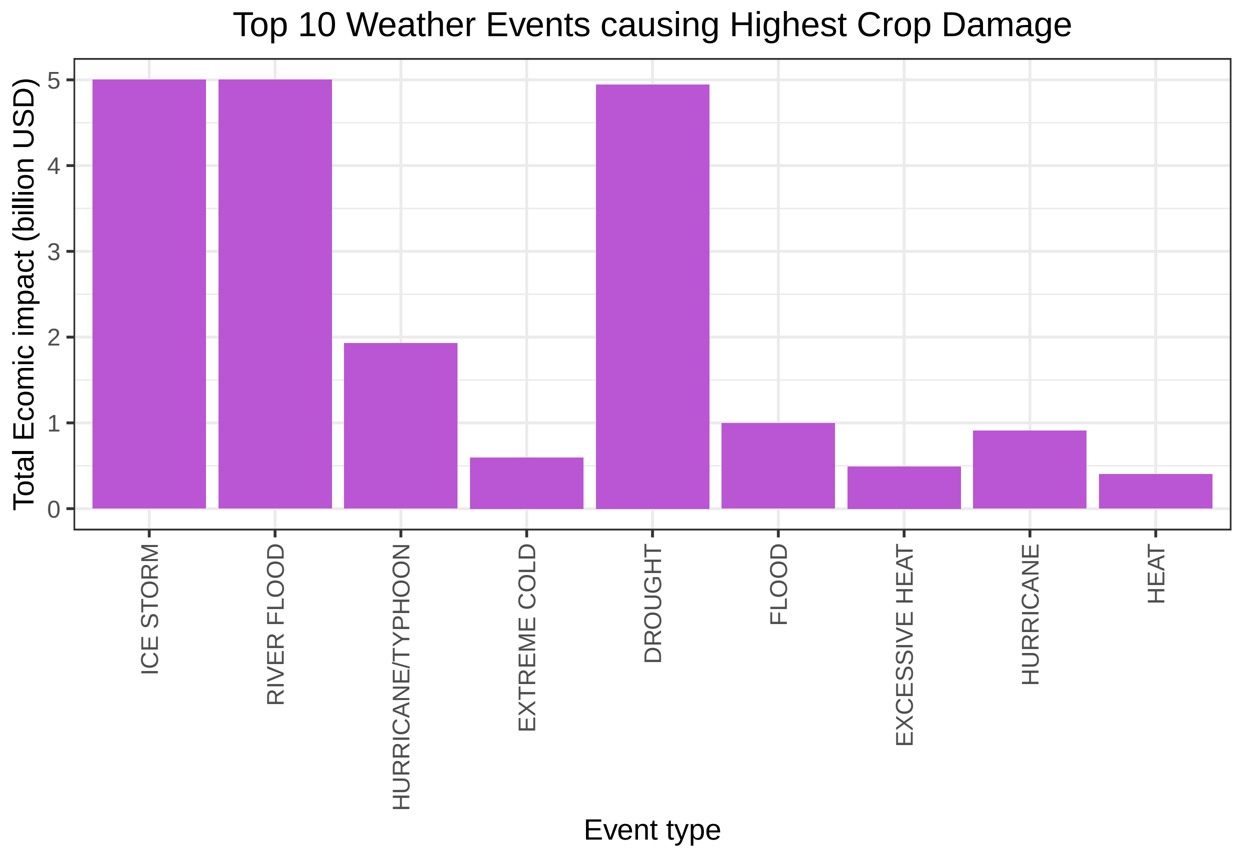
Economic impact

The top 10 weather events that caused the highest economic damages on properties and crops are presented individually and as a combined dataset:

```
#Property
ggplot(storm.property[1:20,], aes(x=reorder(EVTYPE, -DMG), y=DMG/10^9)) +
  geom_bar(stat="identity", fill = "midnightblue") +
  scale_y_continuous(breaks=seq(0,150,25)) +
  ggtitle("Top 10 Weather Events causing Highest Property Damage") +
  labs(x="Event type", y="Total Ecomic impact (billion USD)") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```




```r
#Crops
ggplot(storm.crop[1:20,], aes(x=reorder(EVTYPE, -DMG), y=DMG/10^9)) +
  geom_bar(stat="identity", fill = "mediumorchid") +
  ggtitle("Top 10 Weather Events causing Highest Crop Damage") +
  labs(x="Event type", y="Total Ecomic impact (billion USD)") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```

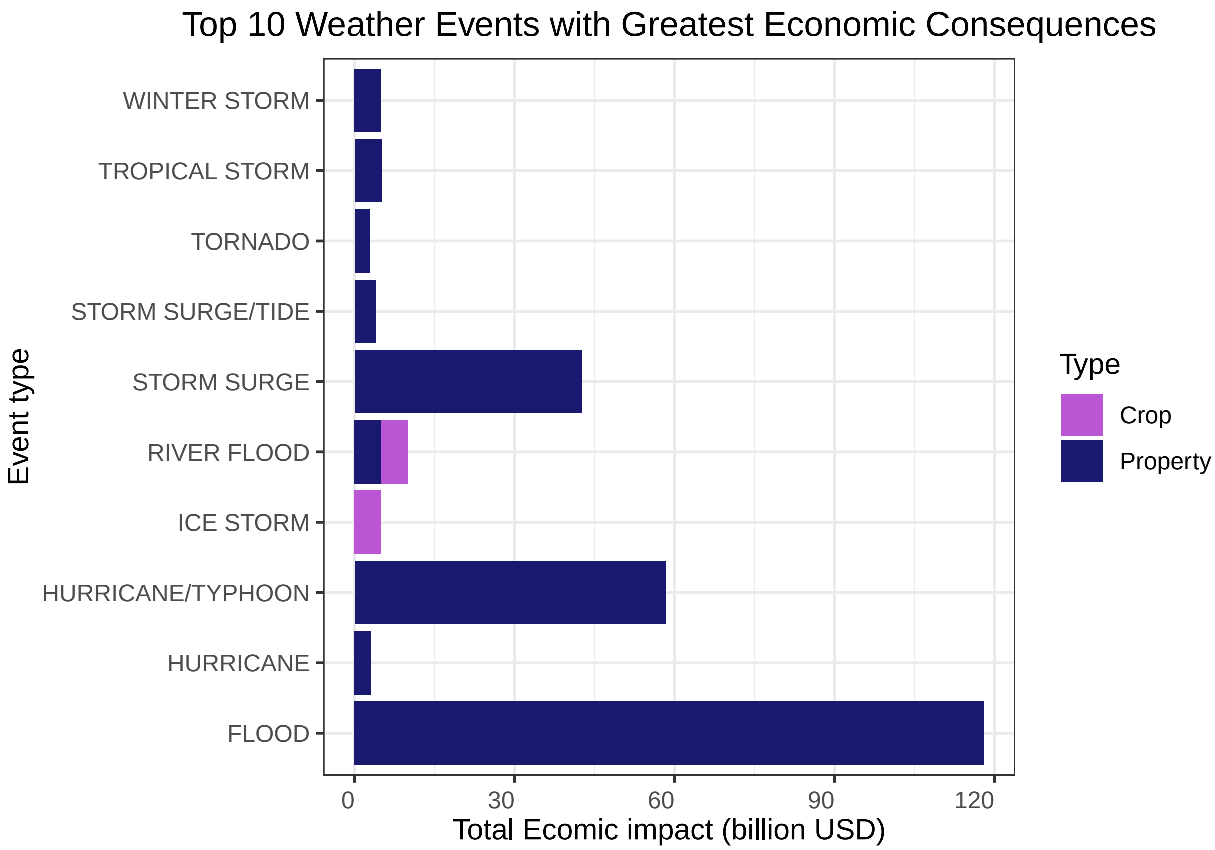



```r
#Combine property & crop damages
storm.property$Type <- "Property"
storm.crop$Type <- "Crop"

storm.damage <- rbind(storm.property, storm.crop)

storm.damage <- storm.damage %>%
                arrange(-DMG, EVTYPE)

ggplot(storm.damage[1:20,], aes(x = EVTYPE, y = DMG/10^9, fill = Type)) +
  coord_flip() +
  geom_bar(stat="identity") +
  scale_fill_manual(values = c("mediumorchid", "midnightblue")) +
  ggtitle("Top 10 Weather Events with Greatest Economic Consequences") +
  labs(x="Event type", y="Total Ecomic impact (billion USD)") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle=0, vjust=0.5, hjust=1))
```



The figures show that most **Property damages** were caused by **Floods** and highest **Crop damages** result from **Ice Storms**, **River floods** and **Droughts**.