

Impact of weather events using storm database

KHAWLA

2023-04-20

Reproducible Research Project 2

Synopsis

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 study explores the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. Tornadoes are wreaking havoc among the catastrophes with the most fatalities and injuries while property damage is due to floods.

Data

The events in the database start in the year 1950 and end in November 2011.

- **DataSet:** Storm Data (<https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>)
- **Documentation:** National Weather Service (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)

Data Processing

Loading storm data

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

Data Preview

```
str(data.storm)
```

```
## '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 ...
```

```
summary(data.storm)
```

```

##      STATE__      BGN_DATE      BGN_TIME      TIME_ZONE
## Min.   : 1.0    Length:902297    Length:902297    Length:902297
## 1st Qu.:19.0    Class :character    Class :character    Class :character
## Median :30.0    Mode  :character    Mode  :character    Mode  :character
## Mean   :31.2
## 3rd Qu.:45.0
## Max.   :95.0
##
##      COUNTY      COUNTYNAME      STATE      EVTYPE
## Min.   : 0.0    Length:902297    Length:902297    Length:902297
## 1st Qu.: 31.0    Class :character    Class :character    Class :character
## Median : 75.0    Mode  :character    Mode  :character    Mode  :character
## Mean   :100.6
## 3rd Qu.:131.0
## Max.   :873.0
##
##      BGN_RANGE      BGN_AZI      BGN_LOCATI      END_DATE
## Min.   : 0.000    Length:902297    Length:902297    Length:902297
## 1st Qu.: 0.000    Class :character    Class :character    Class :character
## Median : 0.000    Mode  :character    Mode  :character    Mode  :character
## Mean   : 1.484
## 3rd Qu.: 1.000
## Max.   :3749.000
##
##      END_TIME      COUNTY_END COUNTYENDN      END_RANGE
## Length:902297      Min.   :0      Mode:logical    Min.   : 0.0000
## Class :character    1st Qu.:0      NA's:902297      1st Qu.: 0.0000
## Mode  :character    Median :0                      Median : 0.0000
##                      Mean   :0                      Mean   : 0.9862
##                      3rd Qu.:0                      3rd Qu.: 0.0000
##                      Max.   :0                      Max.   :925.0000
##
##      END_AZI      END_LOCATI      LENGTH      WIDTH
## Length:902297      Length:902297      Min.   : 0.0000    Min.   : 0.000
## Class :character    Class :character    1st Qu.: 0.0000    1st Qu.: 0.000
## Mode  :character    Mode  :character    Median : 0.0000    Median : 0.000
##                      Mean   : 0.2301    Mean   : 7.503
##                      3rd Qu.: 0.0000    3rd Qu.: 0.000
##                      Max.   :2315.0000    Max.   :4400.000
##
##      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    Length:902297      Min.   : 0.000    Length:902297
## 1st Qu.: 0.00    Class :character    1st Qu.: 0.000    Class :character
## Median : 0.00    Mode  :character    Median : 0.000    Mode  :character
## Mean   : 12.06                      Mean   : 1.527
## 3rd Qu.: 0.50                      3rd Qu.: 0.000
## Max.   :5000.00                      Max.   :990.000

```

```
##
##      WFO              STATEOFFIC      ZONENAMES      LATITUDE
## Length:902297      Length:902297      Length:902297      Min.   : 0
## Class :character    Class :character    Class :character    1st Qu.:2802
## Mode  :character    Mode  :character    Mode  :character    Median :3540
##                                     Mean  :2875
##                                     3rd Qu.:4019
##                                     Max.   :9706
##                                     NA's   :47
##      LONGITUDE      LATITUDE_E      LONGITUDE_      REMARKS
## Min.   :-14451      Min.   : 0      Min.   :-14455      Length:902297
## 1st Qu.: 7247      1st Qu.: 0      1st Qu.: 0      Class :character
## Median : 8707      Median : 0      Median : 0      Mode  :character
## Mean   : 6940      Mean   :1452      Mean   : 3509
## 3rd Qu.: 9605      3rd Qu.:3549      3rd Qu.: 8735
## Max.    :17124      Max.    :9706      Max.    :106220
##                                     NA's    :40
##      REFNUM
## Min.   : 1
## 1st Qu.:225575
## Median :451149
## Mean   :451149
## 3rd Qu.:676723
## Max.    :902297
##
```

```
head(data.storm)
```

| STAT... | BGN_DATE | BGN_T... | TIME_Z... | CO... | COUNTY... | ST... | EVTY... | BGN |
|---------|----------------------|----------|-----------|-------|------------|-------|---------|-----|
| <dbl> | <chr> | <chr> | <chr> | <dbl> | <chr> | <chr> | <chr> | |
| 1 | 1 4/18/1950 0:00:00 | 0130 | CST | 97 | MOBILE | AL | TORNADO | |
| 2 | 1 4/18/1950 0:00:00 | 0145 | CST | 3 | BALDWIN | AL | TORNADO | |
| 3 | 1 2/20/1951 0:00:00 | 1600 | CST | 57 | FAYETTE | AL | TORNADO | |
| 4 | 1 6/8/1951 0:00:00 | 0900 | CST | 89 | MADISON | AL | TORNADO | |
| 5 | 1 11/15/1951 0:00:00 | 1500 | CST | 43 | CULLMAN | AL | TORNADO | |
| 6 | 1 11/15/1951 0:00:00 | 2000 | CST | 77 | LAUDERDALE | AL | TORNADO | |

6 rows | 1-10 of 38 columns

```
names(data.storm) <- tolower(names(data.storm))
```

The data set contains 37 variables and 902297 observations.

Data Preparation

```
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##      filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
storm <- data.storm %>%
  select(c("evtype", "fatalities", "injuries", "propdmg", "propdmgexp", "cropdmg", "cropdmgexp"))
head(storm)
```

| evtype <chr> | fatalities <dbl> | injuries <dbl> | propdmg <dbl> | propdmgexp <chr> | cropdmg <dbl> | cropdmgexp <chr> |
|-----------------|---------------------|-------------------|------------------|---------------------|------------------|---------------------|
| 1 TORNADO | 0 | 15 | 25.0 | K | 0 | |
| 2 TORNADO | 0 | 0 | 2.5 | K | 0 | |
| 3 TORNADO | 0 | 2 | 25.0 | K | 0 | |
| 4 TORNADO | 0 | 2 | 2.5 | K | 0 | |
| 5 TORNADO | 0 | 2 | 2.5 | K | 0 | |
| 6 TORNADO | 0 | 6 | 2.5 | K | 0 | |

6 rows

```
summary(storm)
```

```
##      evtype      fatalities      injuries      propdmg
## Length:902297  Min.   : 0.0000  Min.   : 0.0000  Min.   : 0.00
## Class :character 1st Qu.: 0.0000  1st Qu.: 0.0000  1st Qu.: 0.00
## Mode  :character Median : 0.0000  Median : 0.0000  Median : 0.00
##                Mean   : 0.0168  Mean   : 0.1557  Mean   : 12.06
##                3rd Qu.: 0.0000  3rd Qu.: 0.0000  3rd Qu.: 0.50
##                Max.   :583.0000  Max.   :1700.0000 Max.   :5000.00
##      propdmgexp      cropdmg      cropdmgexp
## Length:902297  Min.   : 0.000  Length:902297
## Class :character 1st Qu.: 0.000  Class :character
## Mode  :character Median : 0.000  Mode  :character
##                Mean   : 1.527
##                3rd Qu.: 0.000
##                Max.   :990.000
```

```
str(storm)
```

```
## 'data.frame':    902297 obs. of  7 variables:
## $ evtype      : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ 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   "" "" "" "" ...
```

```
View(storm)
sum(is.na(storm))
```

```
## [1] 0
```

We'll use only 7 variables:

evtype: type of event

fatalities: number of fatalities for each event

injuries: number of injuries for each event

propdmg: property damage

propdmgexp: crop damage

cropdmg: property damage exponents (in B, M, K and H)

cropdmgexp: crop damage exponents (in B, M, K and H)

```
rm(data.storm)
```

we remove the raw data set and keep on working on the *storm* data set

Data Cleaning

```
sum(unique(storm$fatalities))
```

```
## [1] 2342
```

```
sum(unique(storm$injuries))
```

```
## [1] 35453
```

```
sum(unique(storm$propdmg))
```

```
## [1] 280155
```

```
unique(storm$propdmgexp)
```

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

```
unique(storm$cropdmgexp)
```

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

we look for any unusual data

```
storm$propdmgexp[storm$propdmgexp %in% c("", "+", "-", "?")] <- "0"  
storm$propdmgexp[storm$propdmgexp %in% c("B")] <- "9"  
storm$propdmgexp[storm$propdmgexp %in% c("M", "m")] <- "6"  
storm$propdmgexp[storm$propdmgexp %in% c("K")] <- "3"  
storm$propdmgexp[storm$propdmgexp %in% c("H", "h")] <- "2"
```

```
storm$cropdmgexp[storm$cropdmgexp %in% c("", "?")] <- "0"  
storm$cropdmgexp[storm$cropdmgexp %in% c("B")] <- "9"  
storm$cropdmgexp[storm$cropdmgexp %in% c("M", "m")] <- "6"  
storm$cropdmgexp[storm$cropdmgexp %in% c("K", "k")] <- "3"
```

```
storm$propdmgttotal <- storm$propdmg*(10^as.numeric(storm$propdmgexp))  
storm$cropdmgttotal <- storm$cropdmg*(10^as.numeric(storm$cropdmgexp))  
storm$dmgttotal <- storm$propdmgttotal+ storm$cropdmgttotal
```

we found that the last two variables do not have clean data so we clean it.

```
storm_by_event <- storm %>%  
  group_by(evtype) %>%  
  summarize(sum_fatalities = sum(fatalities),  
            sum_injuries = sum(injuries),  
            sum_propdmg = sum(propdmgttotal),  
            sum_cropdmg = sum(cropdmgttotal),  
            sum_totaldmg = sum(dmgttotal))  
head(storm_by_event)
```

| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_prop... <dbl> | sum_crop... <dbl> | sum_totaldmg <dbl> |
|-----------------------|-------------------------|-----------------------|----------------------|----------------------|-----------------------|
| HIGH SURF ADVISORY | 0 | 0 | 200000 | 0 | 200000 |
| COASTAL FLOOD | 0 | 0 | 0 | 0 | 0 |
| FLASH FLOOD | 0 | 0 | 50000 | 0 | 50000 |
| LIGHTNING | 0 | 0 | 0 | 0 | 0 |
| TSTM WIND | 0 | 0 | 8100000 | 0 | 8100000 |
| TSTM WIND (G45) | 0 | 0 | 8000 | 0 | 8000 |
| 6 rows | | | | | |

we summarized the data by events showing each event with its summarized fatalities, injuries and damages.

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

```
# Events that caused most fatalities
storm_fatalities <- arrange(storm_by_event, desc(sum_fatalities))
fatalitydata <- head(storm_fatalities)
fatalitydata
```

| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_prop... <dbl> | sum_crop... <dbl> | sum_totaldmg <dbl> |
|-------------------|-------------------------|-----------------------|----------------------|----------------------|-----------------------|
| TORNADO | 5633 | 91346 | 56947380677 | 414953270 | 57362333947 |
| EXCESSIVE HEAT | 1903 | 6525 | 7753700 | 492402000 | 500155700 |
| FLASH FLOOD | 978 | 1777 | 16822673979 | 1421317100 | 18243991079 |
| HEAT | 937 | 2100 | 1797000 | 401461500 | 403258500 |
| LIGHTNING | 816 | 5230 | 930379430 | 12092090 | 942471520 |
| TSTM WIND | 504 | 6957 | 4484928495 | 554007350 | 5038935845 |

6 rows

```
# Events that caused most injuries
storm_injuries <- arrange(storm_by_event, desc(sum_injuries))
injurydata <- head(storm_injuries)
injurydata
```

| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_propdmg <dbl> | sum_crop... <dbl> | sum_totaldmg <dbl> |
|-------------------|-------------------------|-----------------------|----------------------|----------------------|-----------------------|
| TORNADO | 5633 | 91346 | 56947380677 | 414953270 | 57362333947 |
| TSTM WIND | 504 | 6957 | 4484928495 | 554007350 | 5038935845 |
| FLOOD | 470 | 6789 | 144657709807 | 5661968450 | 150319678257 |
| EXCESSIVE HEAT | 1903 | 6525 | 7753700 | 492402000 | 500155700 |
| LIGHTNING | 816 | 5230 | 930379430 | 12092090 | 942471520 |
| HEAT | 937 | 2100 | 1797000 | 401461500 | 403258500 |

6 rows

we retain only the first 6 events that caused the most fatalities and injuries.

2. Across the United States, which types of events have the greatest economic consequences?

```
# Events with greatest economic consequences
storm_damage <- arrange(storm_by_event, desc(sum_totaldmg))
damagedata <- head(storm_damage)
damagedata
```


| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_propdmg <dbl> | sum_crop... <dbl> | sum_totald <dbl> |
|-------------------|-------------------------|-----------------------|----------------------|----------------------|---------------------|
| FLOOD | 470 | 6789 | 144657709807 | 5661968450 | 1503196782 |
| HURRICANE/TYPHOON | 64 | 1275 | 69305840000 | 2607872800 | 719137128 |
| TORNADO | 5633 | 91346 | 56947380677 | 414953270 | 573623339 |
| STORM SURGE | 13 | 38 | 43323536000 | 5000 | 433235410 |
| HAIL | 15 | 1361 | 15735267513 | 3025954473 | 187612219 |
| FLASH FLOOD | 978 | 1777 | 16822673979 | 1421317100 | 182439910 |

6 rows

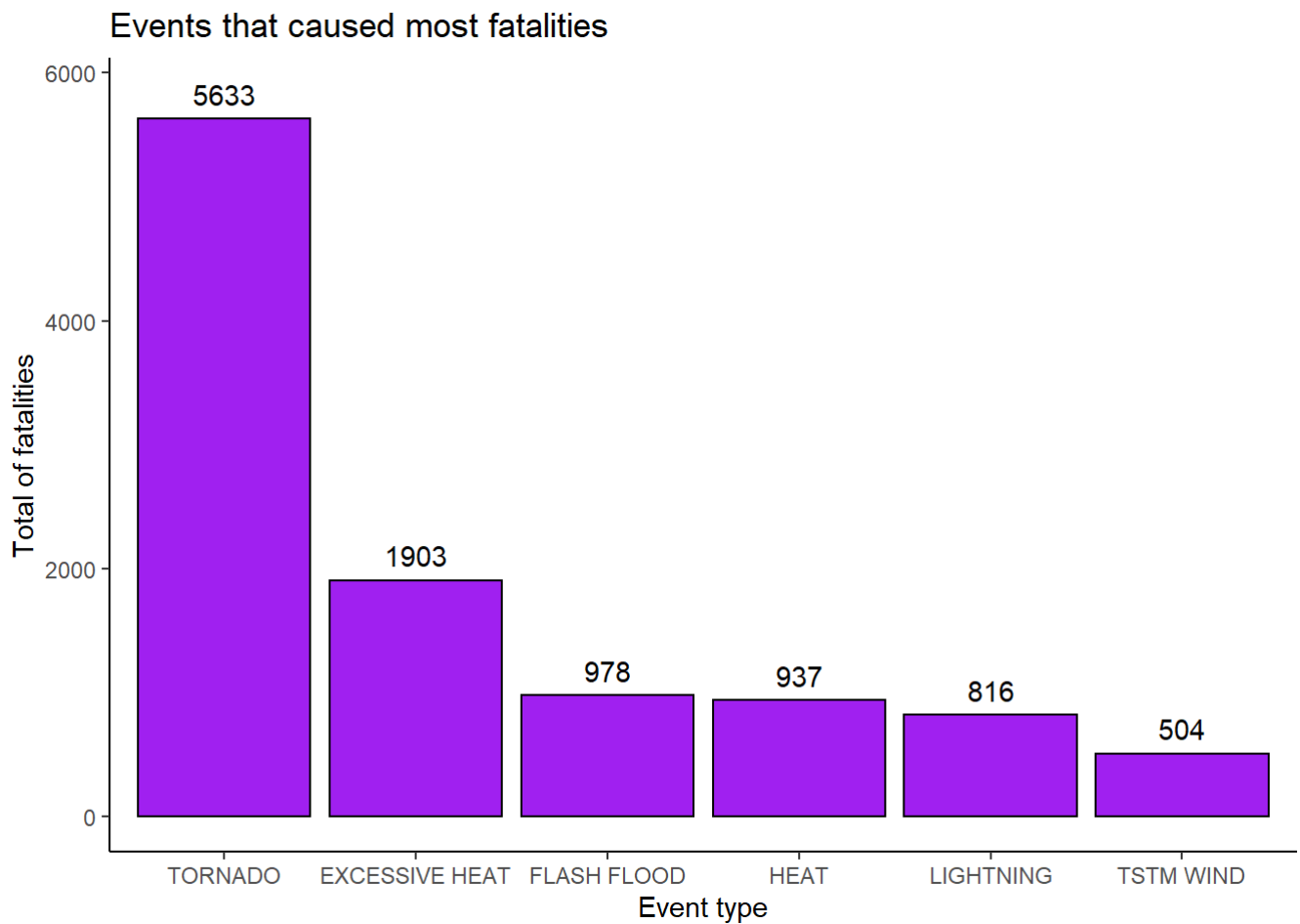
we retain only the first 6 events that caused the most economic consequences.

Results

```
library(ggplot2)
```

FATALITIES

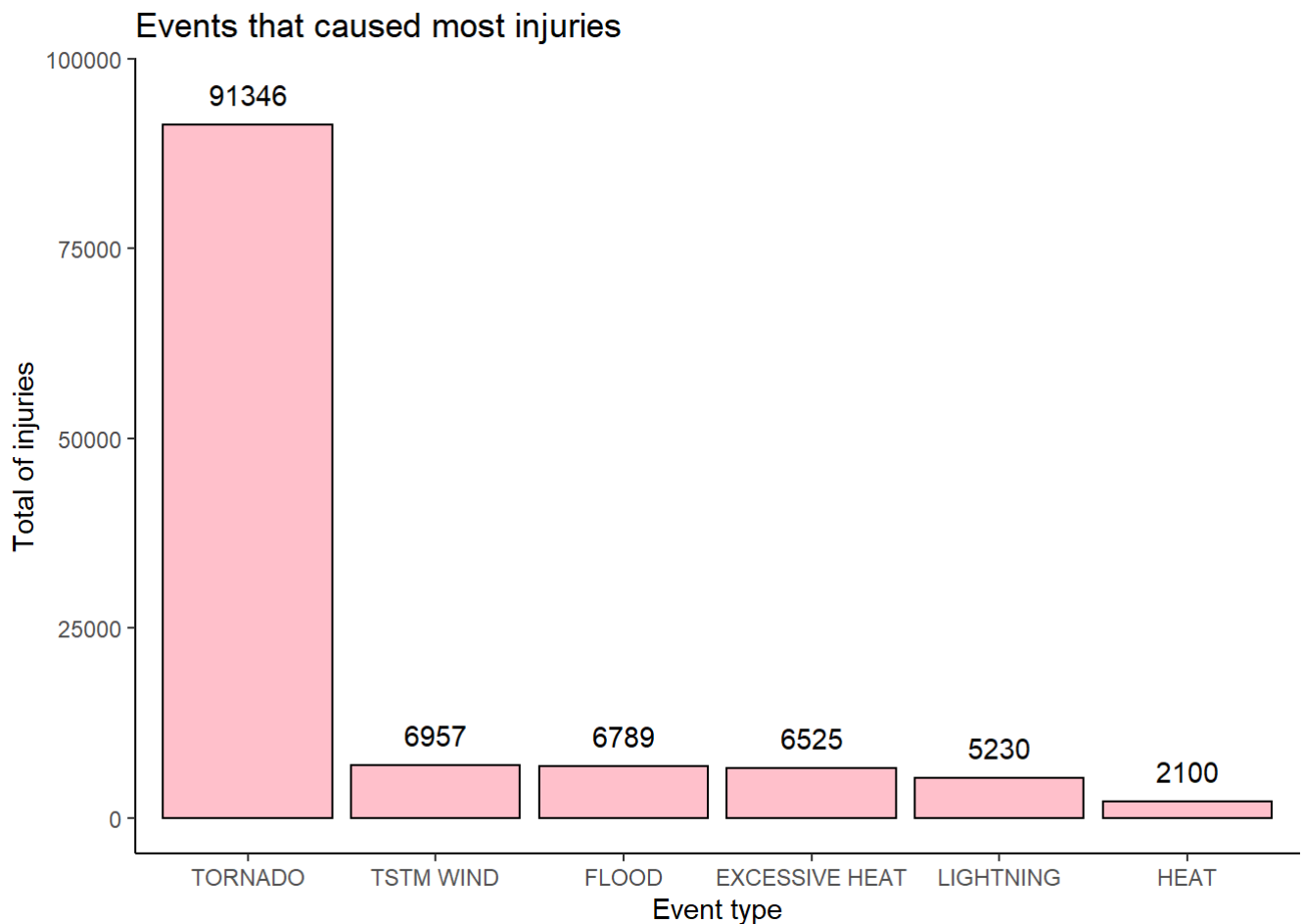
```
ggplot(fatalitydata, aes(reorder(evtype, -sum_fatalities),sum_fatalities))+
  geom_bar(stat = "identity",color= "black", fill="purple")+
  geom_text(aes(label = sum_fatalities), nudge_y = 200)+
  ggtitle("Events that caused most fatalities")+
  xlab("Event type") + ylab("Total of fatalities") +
  theme_classic()
```



it looks like **Tornadoes** causes the most fatalities with 6533 and exceeds by far **excessive heat** which registered 1903 fatalities.

INJURIES

```
ggplot(injurydata, aes(reorder(evtype, -sum_injuries),sum_injuries))+  
  geom_bar(stat = "identity",color= "black", fill="pink")+  
  geom_text(aes(label = sum_injuries), nudge_y = 4000)+  
  ggtitle("Events that caused most injuries")+  
  xlab("Event type") + ylab("Total of injuries") +  
  theme_classic()
```



it looks like **Tornadoes** causes the most injuries with 91346 and exceeds by far **TSTM Wind** which registered 6957 injuries.

ECONOMIC CONSEQUENCES/DAMAGES

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble 3.1.6      v purrr 0.3.4
## v tidyr 1.2.0      v stringr 1.4.0
## v readr 2.1.2      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

```
head(damagedata)
```

| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_propdmg <dbl> | sum_crop... <dbl> | sum_totald <dbl> |
|-------------------|-------------------------|-----------------------|----------------------|----------------------|---------------------|
| FLOOD | 470 | 6789 | 144657709807 | 5661968450 | 1503196782 |
| HURRICANE/TYPHOON | 64 | 1275 | 69305840000 | 2607872800 | 719137128 |
| TORNADO | 5633 | 91346 | 56947380677 | 414953270 | 573623339 |
| STORM SURGE | 13 | 38 | 43323536000 | 5000 | 433235410 |
| HAIL | 15 | 1361 | 15735267513 | 3025954473 | 187612219 |

| evtype <chr> | sum_fatalities <dbl> | sum_injuries <dbl> | sum_propdmg <dbl> | sum_crop... <dbl> | sum_totald... <dbl> |
|-----------------|-------------------------|-----------------------|----------------------|----------------------|------------------------|
| FLASH FLOOD | 978 | 1777 | 16822673979 | 1421317100 | 182439910 |

```
damage <- damagedata %>%
  gather(key = "Type", value = "TotalDamage", c("sum_propdmg", "sum_cropdmg")) %>%
  select(evtype, Type, TotalDamage)

damage
```

| evtype <chr> | Type <chr> | TotalDamage <dbl> |
|-------------------|---------------|----------------------|
| FLOOD | sum_propdmg | 144657709807 |
| HURRICANE/TYPHOON | sum_propdmg | 69305840000 |
| TORNADO | sum_propdmg | 56947380677 |
| STORM SURGE | sum_propdmg | 43323536000 |
| HAIL | sum_propdmg | 15735267513 |
| FLASH FLOOD | sum_propdmg | 16822673979 |
| FLOOD | sum_cropdmg | 5661968450 |
| HURRICANE/TYPHOON | sum_cropdmg | 2607872800 |
| TORNADO | sum_cropdmg | 414953270 |
| STORM SURGE | sum_cropdmg | 5000 |
| 1-10 of 12 rows | | Previous 1 2 Next |

```
damage$Type[damage$Type %in% c("sum_propdmg")] <- "property damage"
damage$Type[damage$Type %in% c("sum_cropdmg")] <- "crop damage"

damage
```

| evtype <chr> | Type <chr> | TotalDamage <dbl> |
|-------------------|-----------------|----------------------|
| FLOOD | property damage | 144657709807 |
| HURRICANE/TYPHOON | property damage | 69305840000 |
| TORNADO | property damage | 56947380677 |
| STORM SURGE | property damage | 43323536000 |
| HAIL | property damage | 15735267513 |
| FLASH FLOOD | property damage | 16822673979 |
| FLOOD | crop damage | 5661968450 |
| HURRICANE/TYPHOON | crop damage | 2607872800 |

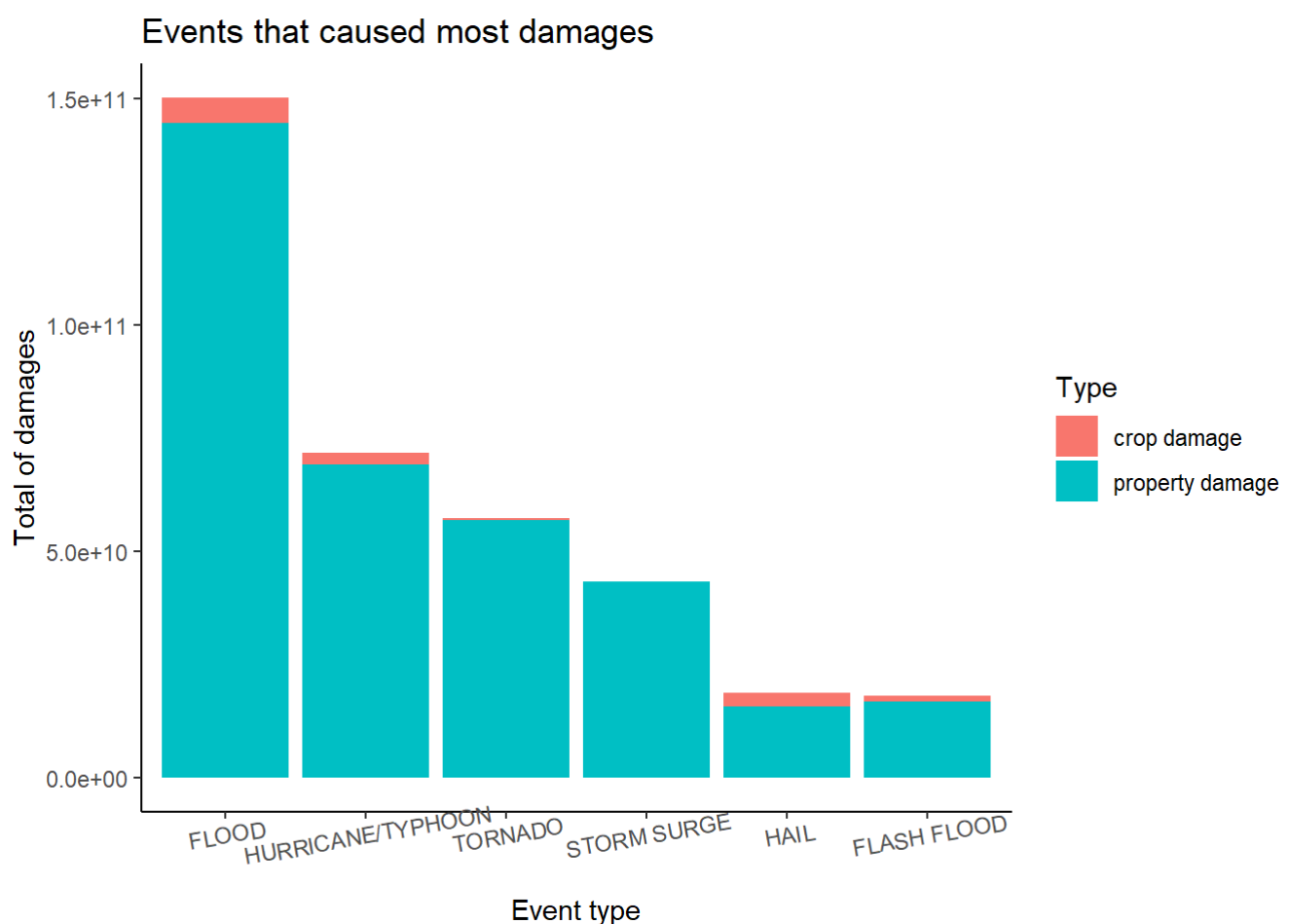
| evtype <chr> | Type <chr> | TotalDamage <dbl> |
|-----------------|---------------|----------------------|
| TORNADO | crop damage | 414953270 |
| STORM SURGE | crop damage | 5000 |

1-10 of 12 rows

Previous 1 2 Next

```
damage$evtype <- with(data = damage, reorder(evtype, -TotalDamage))

ggplot(damage, aes(x = evtype, y = TotalDamage, fill = Type))+
  geom_bar(position = "stack" , stat = "identity")+
  ggtitle("Events that caused most damages")+
  xlab("Event type") + ylab("Total of damages") +
  theme_classic()+ theme(axis.text.x = element_text(angle = 10))
```



it looks like **Floods** cause the most damage in terms of crop and property damages followed by **Hurricanes** and **Tornadoes**