Course Project 2

Peiyu Xiao

8/27/2020

An Analysis on the U.S. National Oceanic and Atmospheric Administration’s Storm Dataset

Summary

This analysis employs a storm dataset collected by the U.S. National Oceanic and Atmospheric Administration (NOAA) to answer two fundamental questions:

1. Across the United States, which types of events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

The results show that tornadoes cause the most damage in terms of population health while flood and drought have the most negative economic consequences in terms of property and crop damage respectively. The following sections discuss the whole analysis in details, which includes 1) data importing and processing, 2) data transformation and visualization, and 3) results.

Data Importing and Processing

```
                                                                                              ,
# download data download.file(url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz
"E:/Data Science Specialization/Reproducible Research/Course Project 2/storm_da

# read data storm_data <- read.csv(file = "E:/Data Science Specialization/Reproducible Research/Course Project 2/st
sep = ",")

# glimpse data str(storm_data)
```

```
## ’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 ...
```

```
head(storm_data)
```

```
##    STATE__            BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE EVTYPE
## 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

##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5           0                                                   0          NA
## 6           0                                                   0          NA
##        END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1           0                                14.0     100 3    0           0       15      25.0

## 2           0                                 2.0     150 2    0           0        0       2.5

## 3           0                                 0.1     123 2    0           0        2      25.0

## 4           0                                 0.0     100 2    0           0        2       2.5

## 5           0                                 0.0     150 2    0           0        2       2.5

## 6           0                                 1.5     177 2    0           0        6       2.5
##     PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1            K       0                                             3040      8812

## 2            K       0                                             3042      8755

## 3            K       0                                             3340      8742

## 4            K       0                                             3458      8626

## 5            K       0                                             3412      8642

## 6            K       0                                             3450      8748

##      LATITUDE_E LONGITUDE_ REMARKS REFNUM

## 1          3051        8806             1

## 2             0           0             2

## 3             0           0             3

## 4             0           0             4

## 5             0           0             5

## 6             0           0             6
```

```
tail(storm_data)
```

```
##            STATE__           BGN_DATE      BGN_TIME TIME_ZONE COUNTY

## 902292          47 11/28/2011 0:00:00 03:00:00 PM         CST     21

## 902293          56 11/30/2011 0:00:00 10:30:00 PM         MST      7

## 902294          30 11/10/2011 0:00:00 02:48:00 PM         MST      9

## 902295           2 11/8/2011 0:00:00 02:58:00 PM          AKS    213

## 902296           2 11/9/2011 0:00:00 10:21:00 AM          AKS    202

## 902297           1 11/28/2011 0:00:00 08:00:00 PM         CST      6

##                                    COUNTYNAME STATE        EVTYPE BGN_RANGE
## 902292 TNZ001>004 - 019>021 - 048>055 - 088      TN WINTER WEATHER       0
## 902293                          WYZ007 - 017      WY       HIGH WIND       0
## 902294                          MTZ009 - 010      MT       HIGH WIND       0
## 902295                                AKZ213      AK       HIGH WIND       0
## 902296                                AKZ202      AK        BLIZZARD       0
## 902297                                ALZ006      AL      HEAVY SNOW       0
##        BGN_AZI BGN_LOCATI           END_DATE    END_TIME COUNTY_END COUNTYENDN
## 902292                     11/29/2011 0:00:00 12:00:00 PM          0         NA
## 902293                     11/30/2011 0:00:00 10:30:00 PM          0         NA
## 902294                     11/10/2011 0:00:00 02:48:00 PM          0         NA
## 902295                      11/9/2011 0:00:00 01:15:00 PM          0         NA
## 902296                      11/9/2011 0:00:00 05:00:00 PM          0         NA
## 902297                     11/29/2011 0:00:00 04:00:00 AM          0         NA
##          END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES
## 902292           0                         0     0 NA   0          0        0
## 902293           0                         0     0 NA 66          0        0
## 902294           0                         0     0 NA 52          0        0
## 902295           0                         0     0 NA 81          0        0
## 902296           0                         0     0 NA   0          0        0
## 902297           0                         0     0 NA   0          0        0
##        PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO               STATEOFFIC
## 902292       0          K       0          K MEG          TENNESSEE, West
## 902293       0          K       0          K RIW WYOMING, Central and West
## 902294       0          K       0          K TFX         MONTANA, Central
## 902295       0          K       0          K AFG         ALASKA, Northern
## 902296       0          K       0          K AFG         ALASKA, Northern
## 902297       0          K       0          K HUN           ALABAMA, North
##                                                                                                                                                                        
## 902292 LAKE - LAKE - OBION - WEAKLEY - HENRY - DYER - GIBSON - CARROLL - LAUDERDALE - TIPTON - HAYWOO
## 902293                                                                                                                                                          OWL CREEK & BRIDG
## 902294                                                                                                                                                                 NORTH ROCK
## 902295                                                                                                                                                                        
## 902296                                                                                                                                                                        
## 902297                                                                                                                                                                        
##        LATITUDE LONGITUDE LATITUDE_E LONGITUDE_
## 902292        0         0          0          0
## 902293        0         0          0          0
## 902294        0         0          0          0
## 902295        0         0          0          0
## 902296        0         0          0          0
## 902297        0         0          0          0
```

```
##
## 902292
## 902293
## 902294
## 902295 EPISODE NARRATIVE: A 960 mb low over the southern Aleutians at 0300AKST on the 8th intensified
## 902296 EPISODE NARRATIVE: A 960 mb low over the southern Aleutians at 0300AKST on the 8th intensified
## 902297                                                          EPISODE NARRATIVE: An intense upper level low developed on the 28th
##             REFNUM
## 902292 902292
## 902293 902293
## 902294 902294
## 902295 902295
## 902296 902296 ##
902297 902297
```

```
summary(storm_data)
```

```
##          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
##
```

Data Transformation and Visualization

Disaster Events and Population Health

I reorganized the storm data by disaster events types and summarized the total number of fatalities and injuries by each type of events. Then, I plotted two barcharts to visualize the top 5 most harmful events by total fatalities and injuries respectively. Last, I combined the two charts into Figure 1.

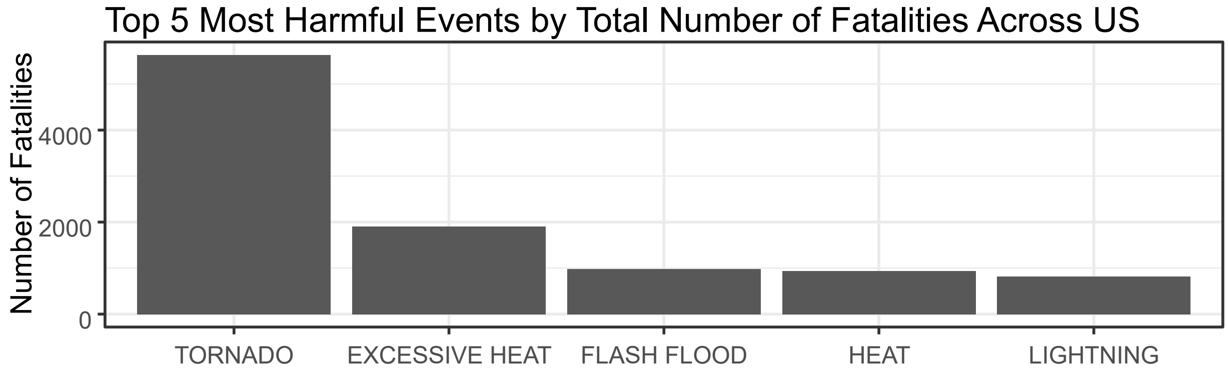
```
# grouping data by EVTYPE and summarizing population health related variables library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2                v purrr          0.3.4
## v tibble 3.0.3                 v dplyr          1.0.1
## v tidyr            1.1.1         v stringr 1.4.0
## v readr            1.3.1         v forcats 0.5.0
## -- Conflicts ------------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()                       masks stats::lag()
```

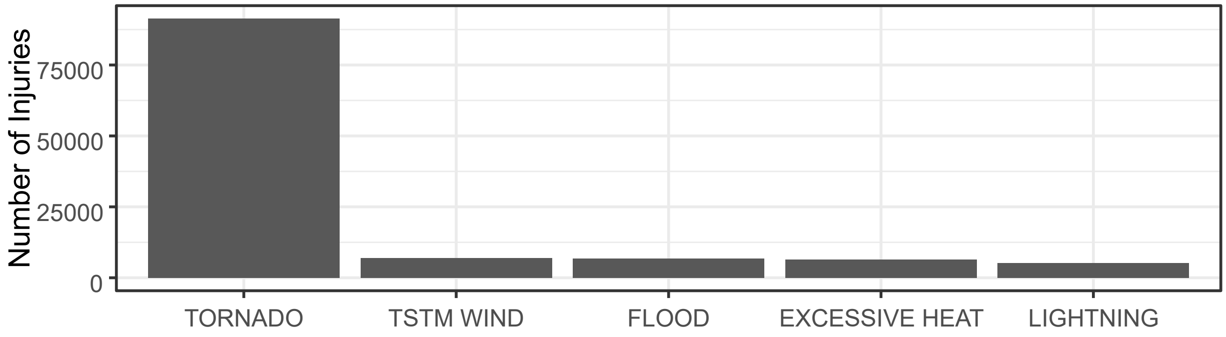
```
storm_data_ph <- storm_data %>% mutate(event_type =
        as.factor(EVTYPE)) %>% group_by(event_type) %>%
        summarize(total_fatalities = sum(FATALITIES, na.rm = TRUE),
                          total_injuries = sum(INJURIES, na.rm = TRUE))
```

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

```
# Top 5 most harmful events according to total number of fatalities across US top_fatalities <-
storm_data_ph %>% arrange(desc(total_fatalities)) %>% filter(row_number()<=5) %>% select(1:2)

plot1 <- top_fatalities %>% ggplot(aes(reorder(event_type, - total_fatalities), total_fatalities)) +
        geom_bar(stat = "identity") + labs(x = "", y = "Number of Fatalities",
                  title = "Top 5 Most Harmful Events by Total Number of Fatalities Across US") +
        theme_bw()

# Top 5 most harmful events according to total number of injuries across US top_injuries <-
storm_data_ph %>% arrange(desc(total_injuries)) %>% filter(row_number()<=5) %>%







        select(1,3)

plot2 <- top_injuries %>% ggplot(aes(reorder(event_type, - total_injuries), total_injuries)) + geom_bar(stat =
        "identity") + labs(x = "", y = "Number of Injuries", title = "Top 5 Most Harmful Events by Total Number of
        Injuries Across US") +
        theme_bw()

# Figure 1 (Top 5 most harmful events with respect to population health) library("ggpubr")
ggarrange(plot1, plot2, nrow = 2)
```



Top 5 Most Harmful Events by Total Number of Injuries Across US



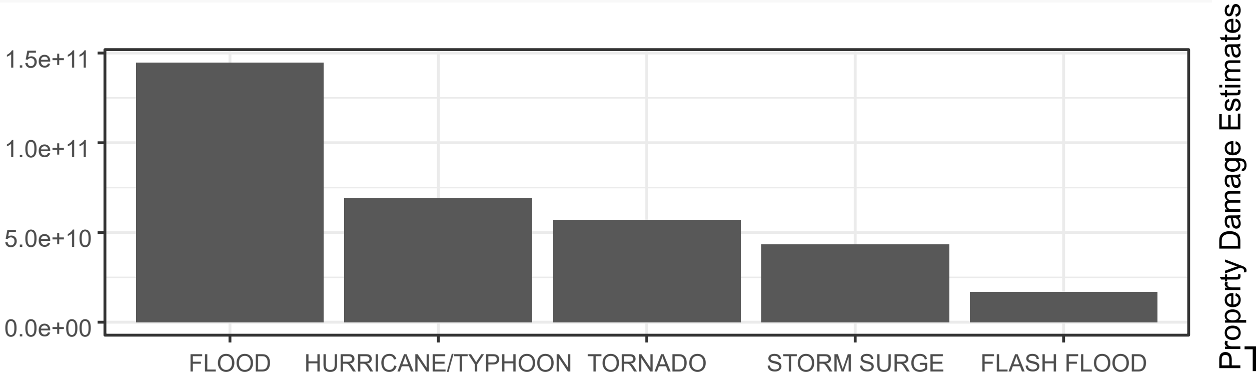
Disaster Events and Economic Consequences

To learn about the economic consequences of disaster events, I rearranged the storm data by disaster events types and summarized the total property and crop damage estimates by each type of events. Then, I plotted two barcharts to visualize the top 5 most harmful events by total property and crop damage estimates. The results are shown in Figure 2.

```
# grouping data by EVTYPE and summarizing economic consequences related variables storm_data_ec <-
storm_data %>% select(EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP) %>%
    mutate(property_damage_parameters = case_when(PROPDMGEXP == "" ~ 1,
                                                  PROPDMGEXP == "-" ~ 1,
                                                  PROPDMGEXP == "?" ~ 1,
                                                  PROPDMGEXP == "+" ~ 1,
                                                  PROPDMGEXP == "0" ~ 1,
                                                  PROPDMGEXP == "1" ~ 10,
                                                  PROPDMGEXP == "2" ~ 100,
                                                  PROPDMGEXP == "3" ~ 1000,
                                                  PROPDMGEXP == "4" ~ 10000,
                                                  PROPDMGEXP == "5" ~ 100000,
                                                  PROPDMGEXP == "6" ~ 1000000,
                                                  PROPDMGEXP == "7" ~ 10000000,
                                                  PROPDMGEXP == "8" ~ 100000000,
                                                  PROPDMGEXP == "B" ~ 1000000000,
                                                  PROPDMGEXP == "h" ~ 1,
                                                  PROPDMGEXP == "H" ~ 1,
                                                  PROPDMGEXP == "K" ~ 1000,
                                                  PROPDMGEXP == "m" ~ 1000000, PROPDMGEXP
                                                  == "M" ~ 1000000),
           crop_damage_parameters = case_when(CROPDMGEXP == "" ~ 1,
                                              CROPDMGEXP == "?" ~ 1,
                                              CROPDMGEXP == "0" ~ 1,
                                              CROPDMGEXP == "2" ~ 100,
                                             CROPDMGEXP == "B" ~ 1000000000,
                                              CROPDMGEXP == "k" ~ 1000,
                                              CROPDMGEXP == "K" ~ 1000,
                                             CROPDMGEXP == "m" ~ 1000000,
                                             CROPDMGEXP == "M" ~ 1000000),
           property_damage = PROPDMG*property_damage_parameters, crop_damage =
    CROPDMG*crop_damage_parameters, event_type = as.factor(EVTYPE)) %>%
    group_by(event_type) %>% summarize(total_property_damage =
    sum(property_damage, na.rm = TRUE),
              total_crop_damage = sum(crop_damage, na.rm = TRUE))
```

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

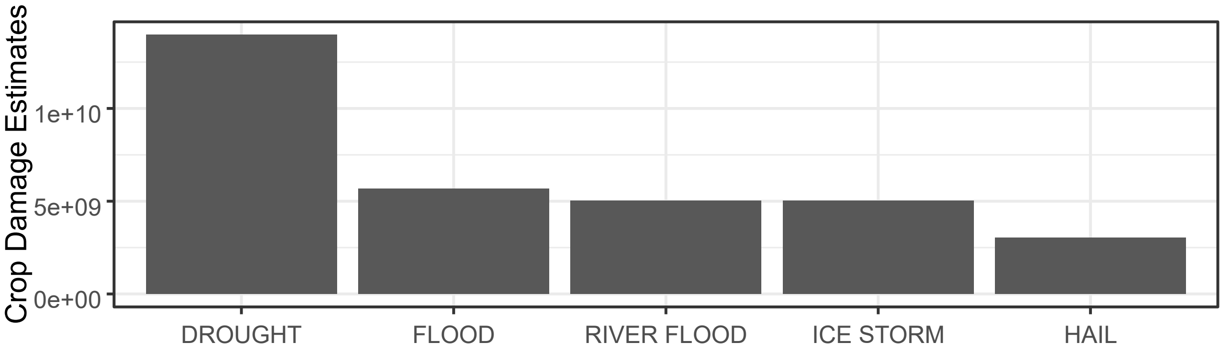
```
# Top 5 most harmful events according to total property damage across US top_pd <-
storm_data_ec %>% arrange(desc(total_property_damage)) %>%
filter(row_number()<=5) %>% select(1:2)

plot3 <- top_pd %>% ggplot(aes(reorder(event_type, - total_property_damage), total_property_damage)) +
        geom_bar(stat = "identity") + labs(x = "", y = "Property Damage Estimates", title = "Top 5 Most Harmful
        Events by Total Property Damage Across US") +
        theme_bw()

# Top 5 most harmful events according to total crop damage across US top_cd <-
storm_data_ec %>%




        arrange(desc(total_crop_damage)) %>%
        filter(row_number()<=5) %>% select(1,3)

plot4 <- top_cd %>% ggplot(aes(reorder(event_type, - total_crop_damage), total_crop_damage)) +
        geom_bar(stat = "identity") + labs(x = "", y = "Crop Damage Estimates", title = "Top 5 Most
        Harmful Events by Total Crop Damage Across US") +
        theme_bw()

# Figure 2 (Top 5 most harmful events with respect to economic consequences) ggarrange(plot3, plot4, nrow = 2)
```



Top 5

Most Harmful Events by Total Property Damage Across US

Top 5 Most Harmful Events by Total Crop Damage Across US



Results

Based on Figure 1, it is clear that the most harmful events with respect to population health is the tornado, which caused about 5,600 deaths and 91,000 injuries during the period between 1950 to November 2011. Figure 2 further indicates that flood and drought are most detrimental events to the economy given that the flood has brought property damages for about 0.15 trillion dollars since 1950, whereas the crop damage triggered by the drought is about 14 billion dollars until 2011.