# Reproducible Research: Peer Assessment 2

Sanjeev Shrestha May 30, 2016

## Analysis of Impacts on Public Health and Economy of USA due to harsh weather

### ABSTRACT:

Severe weather causes impacts on both economy of country and health of people living there. The U.S. National Oceanic and Atmospheric Administration (NOAA) Storm Database has tracked economic losses, fatalities, and injuries associated with major storm events from 1950 onwards to 2011.

In this report, we will use the NOAA database to analyze the total fatality, total injury, and total economic loss over this time frame due to different storms.

The raw data can be easily accessed from [National Weather Service Data][1]. [1]: https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 "National Weather Service Data"

### FUNDAMENTAL SETTINGS BEFORE DATA PROCESSING:

```
library(R.utils)

## Warning: package 'R.utils' was built under R version 3.2.5

## Loading required package: R.oo

## Loading required package: R.methodsS3

## R.methodsS3 v1.7.1 (2016-02-15) successfully loaded. See ?R.methodsS3 for
help.

## R.oo v1.20.0 (2016-02-17) successfully loaded. See ?R.oo for help.

##
## Attaching package: 'R.oo'

## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods

## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save

## R.utils v2.3.0 (2016-04-13) successfully loaded. See ?R.utils for help.
```

```
##
## Attaching package: 'R.utils'

## The following object is masked from 'package:utils':
##
##     timestamp

## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.5

library(plyr)

## Warning: package 'plyr' was built under R version 3.2.5

require(gridExtra)

## Loading required package: gridExtra

## Warning: package 'gridExtra' was built under R version 3.2.5
```

## DATA PROCESSING:

Initial step is to download the data file and unzip it then subset for variables of interest.

```
if (!"stormData.csv.bz2" %in% dir("./repdata-data-StormData.csv/")) {
    print("hhhh")

download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormDat
a.csv.bz2", destfile = "repdata-data-StormData.csv/stormData.csv.bz2")
    bunzip2("repdata-data-StormData.csv/stormData.csv.bz2", overwrite=T,
remove=F)
}
```

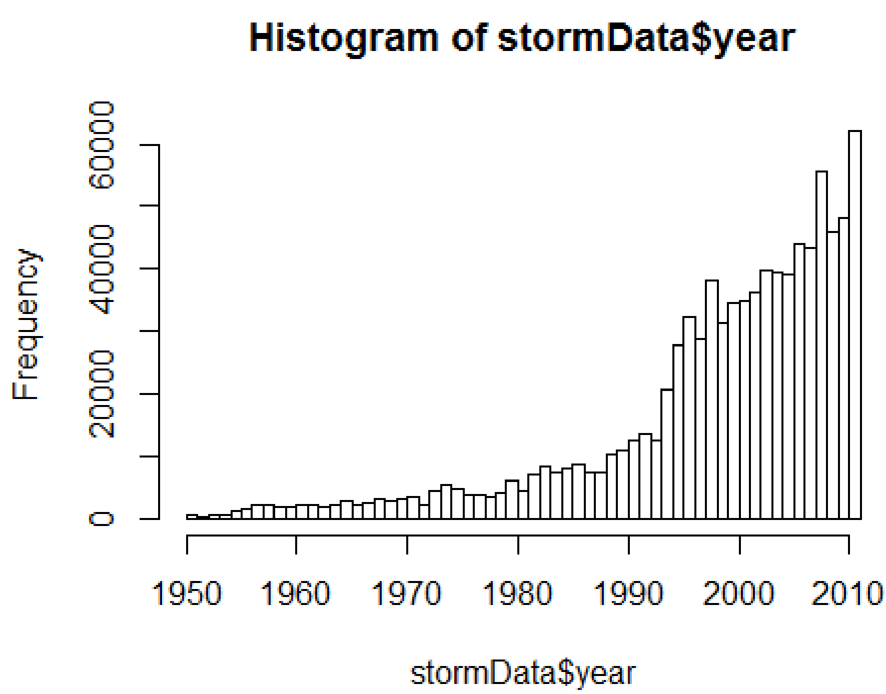
After this, we check the generated csv file. We do not need to load data if there is already an existence of datasets in the working environment.

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

## [1] 902297     37

head(stormData, 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 database consists of storm events from the year 1950 to November 2011. There are 902297 rows and 37 columns in total in the given database.

```r
if (dim(stormData)[2] == 37) {
    stormData$year <- as.numeric(format(as.Date(stormData$BGN_DATE, format = 
"%m/%d/%Y %H:%M:%S"), "%Y"))
}
hist(stormData$year, breaks = 60)
```




Produced histogram shows that the number of events tracked starts to significantly increase starting from 1995. Now we will use the subset of the data from 1990 to 2011 to try to get near to the precise records.

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

## [1] 681500     38
```

This gives us 681500 rows and 38 columns in total.

### LETS ANALYSE HOW IT IMPACTS ON PUBLIC HEALTH:

In this part, we will analyse the number of **fatalities** and **injuries** that are caused by the severe weather events. We would try to get the first 15 most severe types of weather events.

```
sortHelper <- function(fieldName, top = 15, dataset = stormData) {
    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)
```

### LETS ANALYSE HOW IT IMPACTS ON ECONOMY OF THE COUNTRY:

In this part, we will convert the **property damage** and **crop damage** data into comparable numerical forms according to the meaning of units described in the code book (Storm Events). 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)
}
```

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

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

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

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

```r
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"
```

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

## RESULTS:

As for the impact on public health, we have got two sorted lists of severe weather events below by the number of people badly affected.

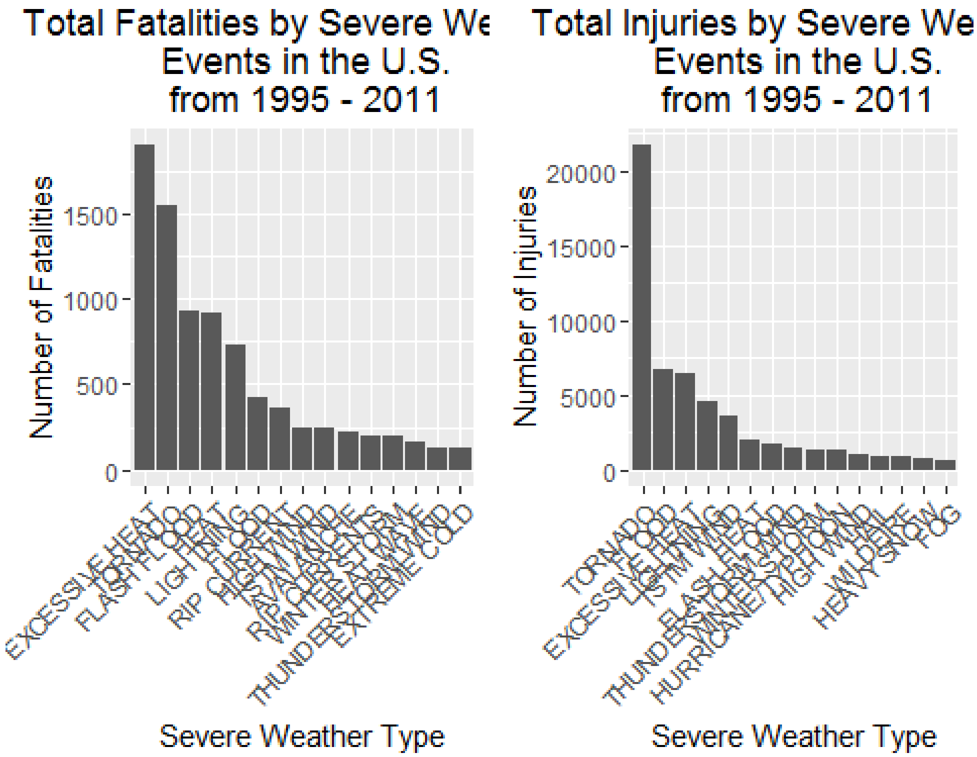
```r
fatalities
```

```
##                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
## 11       RIP CURRENTS        204
## 12       WINTER STORM        195
## 13          HEAT WAVE        161
## 14  THUNDERSTORM WIND        131
## 15       EXTREME COLD        126
```

```
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
## 11          HIGH WIND     1093
## 12               HAIL      916
## 13           WILDFIRE      911
## 14         HEAVY SNOW      751
## 15                FOG      718
```

And the following shows a pair of graphs of total fatalities and total injuries affected by these severe weather events.

```
fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, stat
= "count", width = 1) +
    scale_y_continuous("Number 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")

## Warning: `stat` is deprecated

injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, stat =
"count", width = 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")

## Warning: `stat` is deprecated

grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)
```




Based on the above statistics achieved, we can summaries that **excessive heat** and **tornado** caused most fatalities where as **tornado** caused most injuries in the United States from 1995 to 2011.

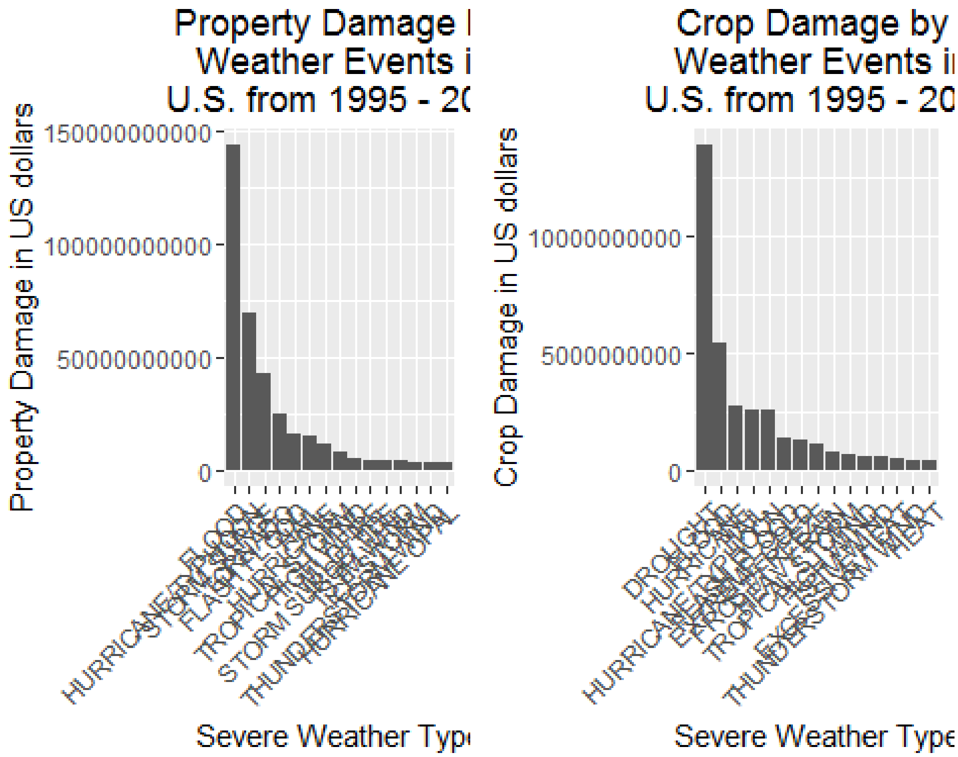
Now for the impact on economy, we have got two sorted lists below by the amount of money cost by damages.

```
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
## 11  STORM SURGE/TIDE      4641188000
## 12          TSTM WIND     4482361440
## 13          ICE STORM     3643555810
## 14 THUNDERSTORM WIND      3399282992
## 15     HURRICANE OPAL     3172846000

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
## 11          HIGH WIND   633561300
## 12          TSTM WIND   553947350
## 13     EXCESSIVE HEAT   492402000
## 14 THUNDERSTORM WIND   414354000
## 15               HEAT   401411500
```

Now follows a pair of graphs of total property damage and total crop damage affected by these severe weather events.

```
propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, stat
= "count", width = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
scale_y_continuous("Property Damage in US dollars")+
    xlab("Severe Weather Type") + ggtitle("Property Damage by\n Weather
Events in\n U.S. from 1995 - 2011")

## Warning: `stat` is deprecated

cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, stat = "count",
width = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
scale_y_continuous("Crop Damage in US dollars") +
    xlab("Severe Weather Type") + ggtitle("Crop Damage by \n Weather Events
in\n U.S. from 1995 - 2011")

## Warning: `stat` is deprecated

grid.arrange(propertyPlot, cropPlot, ncol = 2)
```




Again,based on the above statistics, we can analyse that **flood** and **hurricane/typhoon** caused most property damage where as **drought** and **flood** caused most crop damage in the United States from 1995 to 2011.

## ANALYSIS SUMMARY:

From this analysis, we came to know that **excessive heat** and **tornado** have the most impacts on population health, while **flood**, **drought**, and **hurricane/typhoon** have the greatest impact on the field of economy.