# Storm Damage Analysis

*Jeffrey Uslan*

*September 22, 2015*

## Synopsis

This report will analyse the impact of types of weather events on the public. The report focuses on both economic damage and threats to human health. The data is supplied courtesy of the U.S. National Oceanic and Atmospheric Administration.

## Settings and Libraries

```
echo = TRUE
library(ggplot2)
library(tidyr)
library(zoo)
library(dplyr)
library(stringr)
library(gridExtra)
```

## Data Processing

Download and import data.

```
if (!exists("storm")) storm <- read.csv(bzfile("repdata-data-StormData.csv.bz2"))
```

Glimpse the Shape of the data.

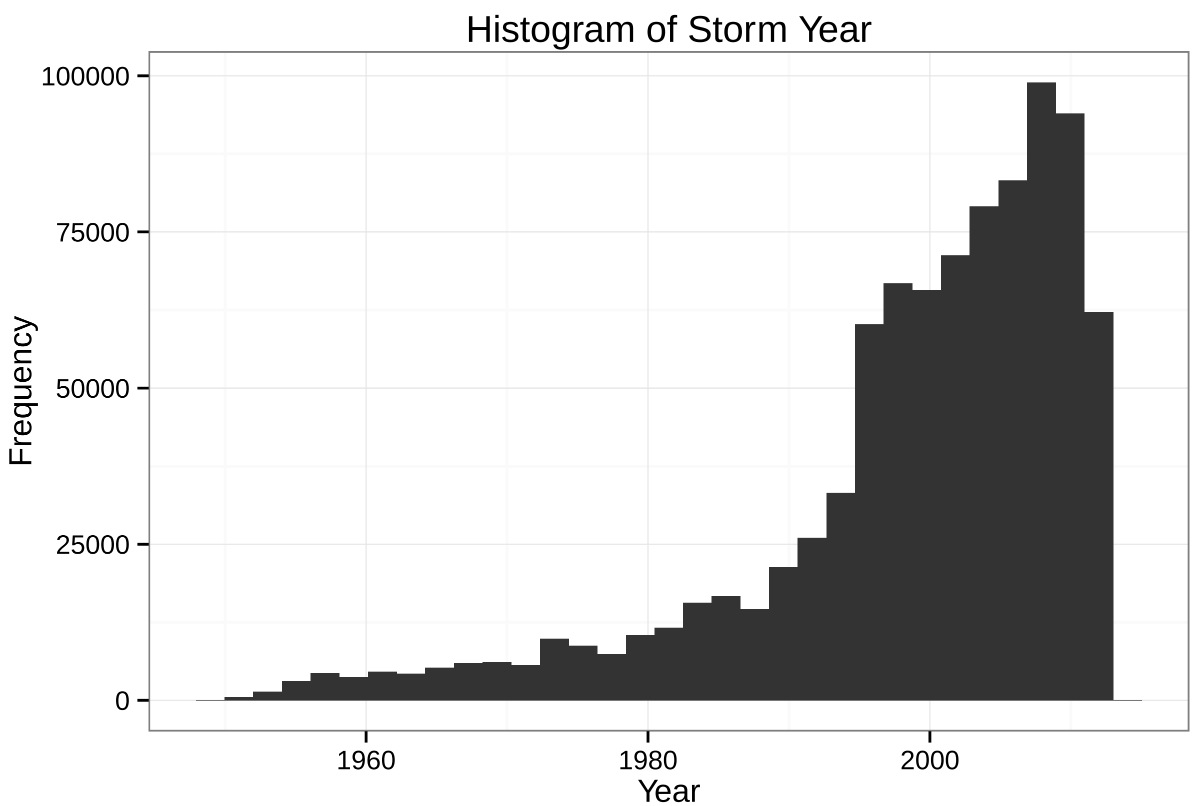
```
storm <- tbl_df(storm)
storm
```

```
## Source: local data frame [902,297 x 37]
##
##    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
## 3        1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4        1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5        1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6        1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
## 7        1 11/16/1951 0:00:00     0100       CST      9     BLOUNT    AL
## 8        1  1/22/1952 0:00:00     0900       CST    123 TALLAPOOSA    AL
## 9        1  2/13/1952 0:00:00     2000       CST    125 TUSCALOOSA    AL
## 10       1  2/13/1952 0:00:00     2000       CST     57    FAYETTE    AL
## ..     ...                ...      ...       ...    ...        ...   ...
```

```
## Variables not shown: EVTYPE (fctr), BGN_RANGE (dbl), BGN_AZI (fctr),
##   BGN_LOCATI (fctr), END_DATE (fctr), END_TIME (fctr), COUNTY_END (dbl),
##   COUNTYENDN (lgl), END_RANGE (dbl), END_AZI (fctr), END_LOCATI (fctr),
##   LENGTH (dbl), WIDTH (dbl), F (int), MAG (dbl), FATALITIES (dbl),
##   INJURIES (dbl), PROPDMG (dbl), PROPDMGEXP (fctr), CROPDMG (dbl),
##   CROPDMGEXP (fctr), WFO (fctr), STATEOFFIC (fctr), ZONENAMES (fctr),
##   LATITUDE (dbl), LONGITUDE (dbl), LATITUDE_E (dbl), LONGITUDE_ (dbl),
##   REMARKS (fctr), REFNUM (dbl)
```

A histogram shows that data records became consistent around 1995, so we will omit earlier events from our analysis.

```r
storm$year <- as.numeric(format(as.Date(storm$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"))
ggplot(data=storm)+theme_bw()+
  geom_histogram(aes(x=year))+xlab("Year")+ylab("Frequency")+ggtitle("Histogram of Storm Year")
```




```r
# Limit to dense data after 1995
storm = storm[storm$year>=1995,]
```

## Public Health

The recorded weather events are not consistent:

```r
storm$EVTYPE= as.factor(tolower(storm$EVTYPE))
arrange(storm %>% group_by(EVTYPE) %>% summarise(Count=n()),desc(Count))
```

```
## Source: local data frame [716 x 2]
##
##              EVTYPE  Count
## 1              hail 215932
## 2         tstm wind 128925
## 3  thunderstorm wind  81746
## 4       flash flood  52673
## 5             flood  24642
## 6           tornado  24335
## 7         high wind  19958
## 8        heavy snow  14710
## 9         lightning  14280
## 10       heavy rain  11640
## ..              ...    ...
```

Rename events based on manual exploration.

```r
storm$simplfied_EVTYPE=as.character(storm$EVTYPE)

word="fld"
rep_ind=grep(word,storm$simplfied_EVTYPE)
storm$simplfied_EVTYPE[rep_ind]="flood"

replacement_words=c("hail","wind","flood","snow","lightning","rain","winter","funnel cloud",
                    "waterspout","fire","blizzard","drought","ice","heat","fog","freeze",
                    "surf","tropical","cold","dust","avalanche","hurricane","tide",
                    "landslide","warm","current","dry","tornado","storm")
for (word in replacement_words) {

 rep_ind=grep(word,storm$simplfied_EVTYPE)
 storm$simplfied_EVTYPE[rep_ind]=word

}
```

This simplifies our categories greatly, taking us from 716 to 216

```r
Storm_Health=storm %>% group_by(simplfied_EVTYPE) %>%
          summarise(Fatalities=sum(FATALITIES),Injuries=sum(INJURIES),
                    Count=n())
Storm_Health %>% arrange(desc(Count))
```

```
## Source: local data frame [202 x 4]
##
##    simplfied_EVTYPE Fatalities Injuries  Count
## 1              wind       1082     7234 259950
## 2              hail         15     1021 217619
## 3             flood       1414     8598  82607
## 4           tornado       1545    21783  24365
## 5            winter        256     1836  19477
```

```
## 6            snow        149       829  16481
## 7       lightning        730      4633  14287
## 8            rain        101       268  11977
## 9    funnel cloud          0         1   6408
## 10           fire         87      1458   4215
## ..            ...        ...       ...    ...
```

## Economic Damage

Adjusting damage estimates according to unit notation provided to be in millions of dollars

```
storm$PROPDMG[storm$PROPDMGEXP=="B"]=(storm$PROPDMG[storm$PROPDMGEXP=="B"]*(10^9))/(10^6)
storm$PROPDMG[storm$PROPDMGEXP=="M"]=(storm$PROPDMG[storm$PROPDMGEXP=="M"]*(10^6))/(10^6)
storm$PROPDMG[storm$PROPDMGEXP=="K"]=(storm$PROPDMG[storm$PROPDMGEXP=="K"]*(10^3))/(10^6)
storm$PROPDMG[storm$PROPDMGEXP=="H"]=(storm$PROPDMG[storm$PROPDMGEXP=="H"]*(10^2))/(10^6)

storm$CROPDMG[storm$CROPDMGEXP=="B"]=(storm$CROPDMG[storm$CROPDMGEXP=="B"]*(10^9))/(10^6)
storm$CROPDMG[storm$CROPDMGEXP=="M"]=(storm$CROPDMG[storm$CROPDMGEXP=="M"]*(10^6))/(10^6)
storm$CROPDMG[storm$CROPDMGEXP=="K"]=(storm$CROPDMG[storm$CROPDMGEXP=="K"]*(10^3))/(10^6)
storm$CROPDMG[storm$CROPDMGEXP=="H"]=(storm$CROPDMG[storm$CROPDMGEXP=="H"]*(10^2))/(10^6)
```

Producing summary table grouped by weather event type

```
Storm_Economy=storm %>% group_by(simplfied_EVTYPE) %>%
  summarise(Property=sum(PROPDMG),Crop=sum(CROPDMG),
          Count=n())
```

# Results

## Public Health

Public health damage in descending order of fatalities.

```
Storm_Health %>% arrange(desc(Fatalities))
```

```
## Source: local data frame [202 x 4]
##
##    simplfied_EVTYPE Fatalities Injuries  Count
## 1              heat       3081     9088   2587
## 2           tornado       1545    21783  24365
## 3             flood       1414     8598  82607
## 4              wind       1082     7234 259950
## 5         lightning        730     4633  14287
## 6           current        564      524    763
## 7            winter        256     1836  19477
## 8         avalanche        223      159    380
## 9              cold        169      125    823
## 10             surf        160      245   1057
## ..              ...        ...      ...    ...
```

Public health damage in descending order of injuries.

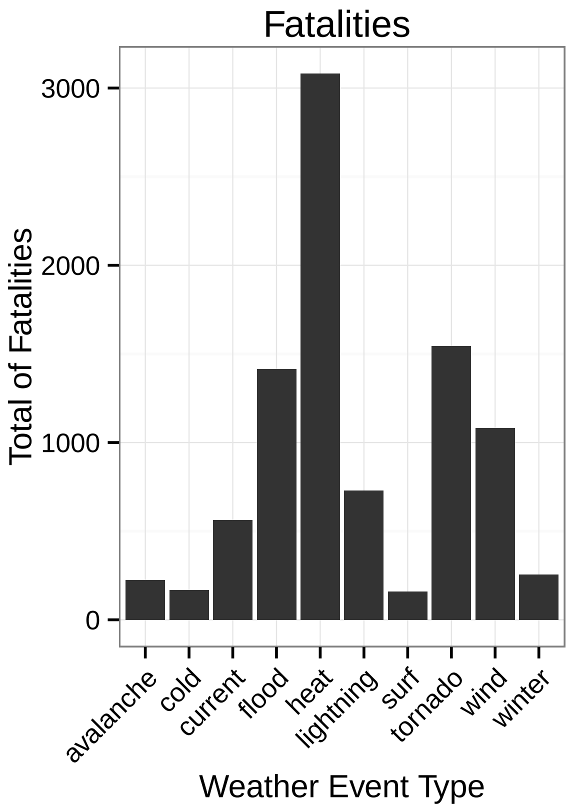
```
Storm_Health %>% arrange(desc(Injuries))
```

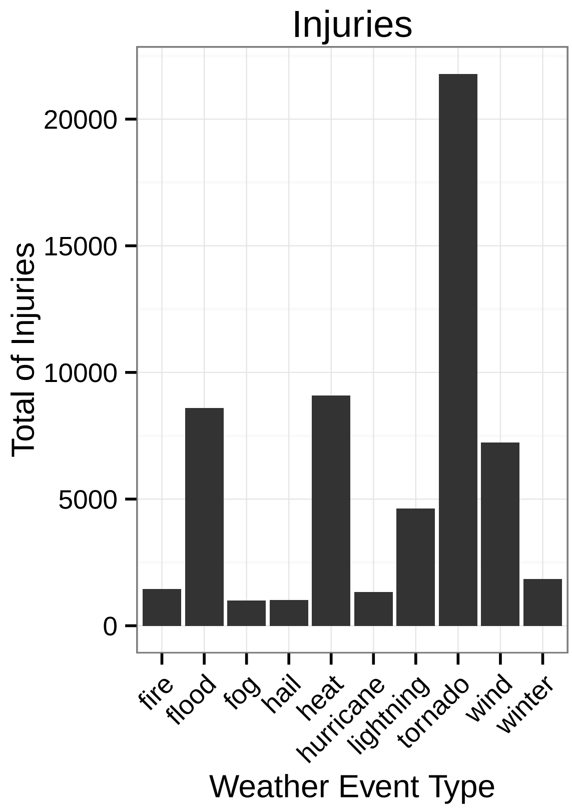
```
## Source: local data frame [202 x 4]
##
##    simplfied_EVTYPE Fatalities Injuries  Count
## 1           tornado       1545    21783  24365
## 2              heat       3081     9088   2587
## 3             flood       1414     8598  82607
## 4              wind       1082     7234 259950
## 5         lightning        730     4633  14287
## 6            winter        256     1836  19477
## 7              fire         87     1458   4215
## 8         hurricane        133     1327    281
## 9              hail         15     1021 217619
## 10              fog         71      994   1851
## ..              ...        ...      ...    ...
```

The graphs below illustrate the events posing most risk to public health.

```
worst_fatalities=(Storm_Health %>% arrange(desc(Fatalities)))$simplfied_EVTYPE[1:10]
temp_plot_data=Storm_Health %>% arrange(desc(Fatalities))
temp_plot_data=temp_plot_data[temp_plot_data$simplfied_EVTYPE %in% worst_fatalities,]
fatality_plot <- ggplot(temp_plot_data) + theme_bw() +
  geom_bar(aes(x =simplfied_EVTYPE,weight=Fatalities ))+
  scale_y_continuous("Total of Fatalities") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  xlab("Weather Event Type") +
  ggtitle("Fatalities ")

worst_Injuries=(Storm_Health %>% arrange(desc(Injuries)))$simplfied_EVTYPE[1:10]
temp_plot_data=Storm_Health %>% arrange(desc(Injuries))
temp_plot_data=temp_plot_data[temp_plot_data$simplfied_EVTYPE %in% worst_Injuries,]
injury_plot <-  ggplot(temp_plot_data) + theme_bw() +
  geom_bar(aes(x =simplfied_EVTYPE,weight=Injuries ))+
  scale_y_continuous("Total of Injuries") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  xlab("Weather Event Type") +
  ggtitle("Injuries")
grid.arrange(fatality_plot, injury_plot, ncol = 2)
```




### Economic Damage

Economic damage in descending order of property damage.

```
Storm_Economy %>% arrange(desc(Property))
```

```
## Source: local data frame [202 x 4]
##
##    simplfied_EVTYPE   Property      Crop  Count
## 1             flood 160599.941 6937.8045  82607
## 2         hurricane  84550.180 5504.7928    281
## 3             storm  44396.884   18.2050    356
## 4           tornado  25117.691  456.6028  24365
## 5              wind  20697.734 1941.7631 259950
## 6              hail  15669.297 3069.4353 217619
## 7              fire   7761.007  402.2676   4215
## 8          tropical   7660.073  693.8860    749
## 9              tide   4650.933    0.8500    427
## 10              ice   3644.372   15.6608   1981
## ..              ...        ...       ...    ...
```

Economic damage in descending order of crop damage.

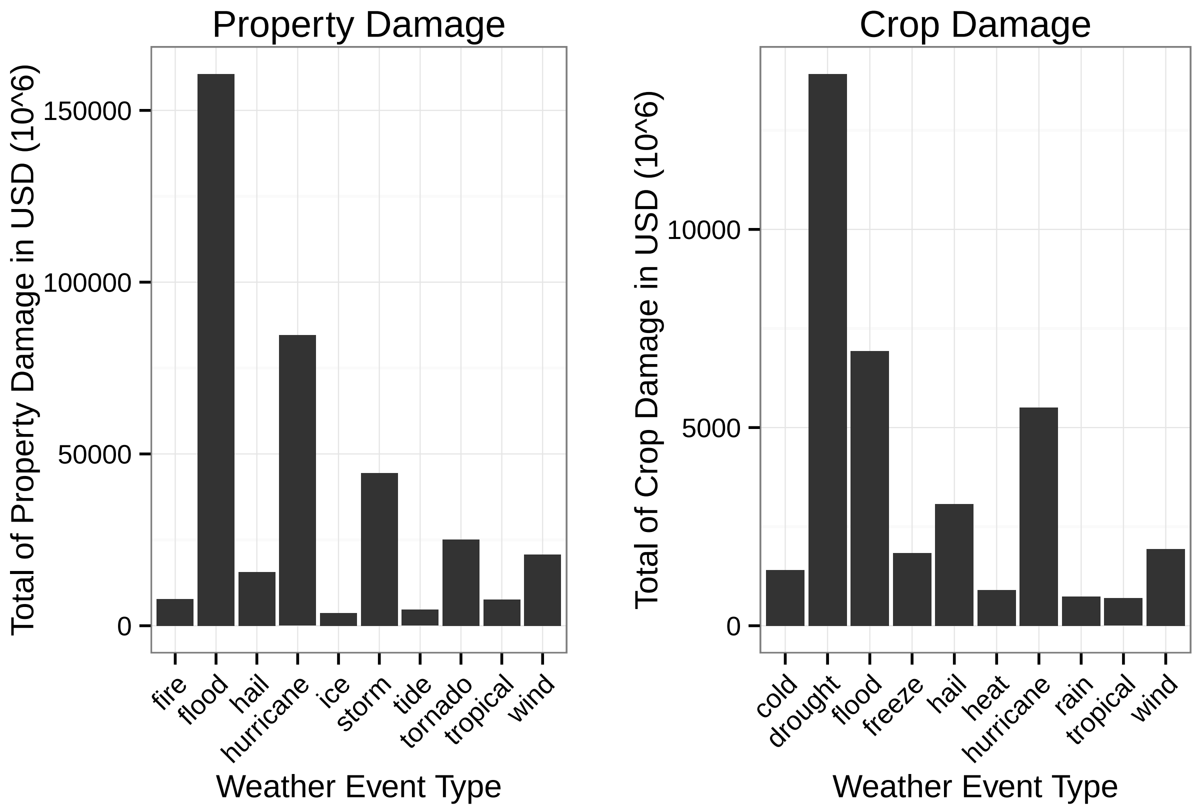
```r
Storm_Economy %>% arrange(desc(Crop))
```

```
## Source: local data frame [202 x 4]
##
##    simplfied_EVTYPE     Property       Crop  Count
## 1           drought   1046.30600 13922.1218   2486
## 2             flood 160599.94142  6937.8045  82607
## 3         hurricane  84550.18001  5504.7928    281
## 4              hail  15669.29732  3069.4353 217619
## 5              wind  20697.73367  1941.7631 259950
## 6            freeze     18.68000  1839.0610   1439
## 7              cold     46.33640  1408.6155    823
## 8              heat     20.12075   899.3135   2587
## 9              rain   3158.28544   740.1528  11977
## 10         tropical   7660.07255   693.8860    749
## ..              ...          ...        ...    ...
```

The graphs below illustrate the events posing most risk to public health.

```r
ind_1=(Storm_Economy %>% arrange(desc(Property)))$simplfied_EVTYPE[1:10]
temp_plot_data=Storm_Economy %>% arrange(desc(Property))
temp_plot_data=temp_plot_data[temp_plot_data$simplfied_EVTYPE %in% ind_1,]
property_plot <- ggplot(temp_plot_data) + theme_bw() +
  geom_bar(aes(x =simplfied_EVTYPE,weight=Property ))+
  scale_y_continuous("Total of Property Damage in USD (10^6)") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  xlab("Weather Event Type") +
  ggtitle("Property Damage")

ind_2=(Storm_Economy %>% arrange(desc(Crop)))$simplfied_EVTYPE[1:10]
temp_plot_data=Storm_Economy %>% arrange(desc(Crop))
temp_plot_data=temp_plot_data[temp_plot_data$simplfied_EVTYPE %in% ind_2,]
crop_plot <-  ggplot(temp_plot_data) + theme_bw() +
  geom_bar(aes(x =simplfied_EVTYPE,weight=Crop ))+
  scale_y_continuous("Total of Crop Damage in USD (10^6)") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  xlab("Weather Event Type") +
  ggtitle("Crop Damage")
grid.arrange(property_plot, crop_plot, ncol = 2)
```




## Conclusion

This Analysis shows that:

- Heat, Tornadoes, and Flooding, respectively, pose the greatest risk to human life.
- Tornadoes, Heat, and Flooding, respectively, pose the greatest risk to human injury.
- Flooding, Hurricanes, and Storms, respectively, inflict the greatest property damage.
- Drought, Hurricanes, and Heat, respectively, inflict the greatest crop damage.