Reproducible Research Project 2

Dhruba

September 13, 2017

Reproducible Research Project 2

Health and Economic Impact of Weather Events in the US

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 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, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

1. Synopsis

The dataset used in this analysis is the U.S. NOAA storm database which tracks characteristics of major storms and severe weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries and property damage. The questions we will be answering are which events are the most harmful with respect to the population health and economy. The analysis will start with data transformation and end with quantitative analysis of the imppact via plot illustration. Results reveal "tornadoes" followed by "excessive heat" are the most dangerous event wrt to population health. On the economic front "flash floods" and "thunderstorm" wreck havoc. Crops are majorly damaged by "drought", "flood", "hails".

Mention worthy, the data set has skewed data in favour of tornadoes. It is only since 1996 that all events has been recorded. That is why the

results are skewed in favour(or against) tornadoes.

2. Data Processing

```
### Loading the Dataset into R environment
library(data.table)
storm <- fread('StormData.csv', header = T, sep = ',')
```

```
##
Read 0.0% of 967216 rows
Read 13.4% of 967216 rows
Read 24.8% of 967216 rows
Read 36.2% of 967216 rows
Read 50.7% of 967216 rows
Read 60.0% of 967216 rows
Read 73.4% of 967216 rows
Read 79.6% of 967216 rows
Read 87.9% of 967216 rows
Read 902297 rows and 37 (of 37) columns from 0.523 GB file in 00:00:11
```

Events are aggregated in the next section as required for further analysis.

```
# number of unique event types
length(unique(storm$EVTYPE))
```

```
## [1] 985
```

```
# translate all letters to lowercase
event_types <- tolower(storm$EVTYPE)

# replace all punct. characters with a space
event_types <- gsub("[[:blank:][:punct:]+]", " ", event_types)
length(unique(event_types))
```

```
## [1] 874
```

```
# update the data frame
storm$EVTYPE <- event_types
```

3. Results

Dangerous Events wrt Population Health

The number of casualties, both injuries and fatalities are aggregated event wise.

Plyr package is installed from CRAN for a good visual experience.

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.4.1
```

```
casualties <- ddply(storm, .(EVTYPE), summarize,
                    fatalities = sum(FATALITIES),
                    injuries = sum(INJURIES))

# Find events that caused most death and injury
fatal_events <- head(casualties[order(casualties$fatalities, decreasing = T), ], 10)
injury_events <- head(casualties[order(casualties$injuries, decreasing = T), ], 10)
```

Table Representation of Fatalities

```
fatal_events[, c("EVTYPE", "fatalities")]
```

```
##             EVTYPE fatalities
## 741        tornado       5633
## 116 excessive heat       1903
## 138    flash flood        978
## 240           heat        937
## 410      lightning        816
## 762      tstm wind        504
## 154          flood        470
## 515    rip current        368
## 314      high wind        248
## 19       avalanche        224
```

Table Representation of Injuries

```
injury_events[, c("EVTYPE", "injuries")]
```

```
##                EVTYPE injuries
## 741           tornado    91346
## 762         tstm wind     6957
## 154             flood     6789
## 116    excessive heat     6525
## 410         lightning     5230
## 240              heat     2100
## 382         ice storm     1975
## 138       flash flood     1777
## 671 thunderstorm wind     1488
## 209              hail     1361
```

Economic damages

The available property damage and crop damage reportings/estimates were used to estimate economic damages.

```
exp_transform <- function(e) {
  # h -> hundred, k -> thousand, m -> million, b -> billion
  if (e %in% c('h', 'H'))
    return(2)
  else if (e %in% c('k', 'K'))
    return(3)
  else if (e %in% c('m', 'M'))
    return(6)
  else if (e %in% c('b', 'B'))
    return(9)
  else if (!is.na(as.numeric(e))) # if a digit
    return(as.numeric(e))
  else if (e %in% c('', '-', '?', '+'))
    return(0)
  else {
    stop("Invalid exponent value.")
  }
}
```

Compiling the results to form the datasets which will be used for the visualization of the results.

```
prop_dmg_exp <- sapply(storm$PROPDMGEXP, FUN=exp_transform)
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
storm$prop_dmg <- storm$PROPDMG * (10 ** prop_dmg_exp)
crop_dmg_exp <- sapply(storm$CROPDMGEXP, FUN=exp_transform)
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
storm$crop_dmg <- storm$CROPDMG * (10 ** crop_dmg_exp)
```

```
library(plyr)
econ_loss <- ddply(storm, .(EVTYPE), summarize,
                   prop_dmg = sum(prop_dmg),
                   crop_dmg = sum(crop_dmg))

# filter out events that caused no economic loss
econ_loss <- econ_loss[(econ_loss$prop_dmg > 0 | econ_loss$crop_dmg > 0), ]
prop_dmg_events <- head(econ_loss[order(econ_loss$prop_dmg, decreasing = T), ], 10)
crop_dmg_events <- head(econ_loss[order(econ_loss$crop_dmg, decreasing = T), ], 10)
```

```
### Table Representation of Property Damage
prop_dmg_events[, c("EVTYPE", "prop_dmg")]
```

```
##                EVTYPE     prop_dmg
## 154             flood 144657709807
## 366 hurricane typhoon  69305840000
## 741           tornado  56947380677
## 585       storm surge  43323536000
## 138       flash flood  16822673979
## 209              hail  15735267513
## 357         hurricane  11868319010
## 755    tropical storm   7703890550
```

```
## 866      winter storm   6688497251
## 314         high wind   5270046295
```

```
### Table Representation of Crop Damage
crop_dmg_events[, c("EVTYPE", "crop_dmg")]
```

```
##                EVTYPE    crop_dmg
## 84            drought 13972566000
## 154             flood  5661968450
## 519       river flood  5029459000
## 382         ice storm  5022113500
## 209              hail  3025954473
## 357         hurricane  2741910000
## 366 hurricane typhoon  2607872800
## 138       flash flood  1421317100
## 125      extreme cold  1312973000
## 185      frost freeze  1094186000
```

4. Visulazing the results

cowplot is used to align the graphs side by side to show the effect of diasters wrt to injuries and fatalities side by side.

Impact on Population Health

Top dangerous weather event types.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
# Set the levels in order
p1 <- ggplot(data=fatal_events,
             aes(x=reorder(EVTYPE, fatalities), y=fatalities, fill=fatalities)) +
  geom_bar(stat="identity") +
  coord_flip() +
  ylab("Total number of fatalities") +
  xlab("Event type") +
  theme(legend.position="none")

p2 <- ggplot(data=injury_events,
             aes(x=reorder(EVTYPE, injuries), y=injuries, fill=injuries)) +
  geom_bar(stat="identity") +
  coord_flip() +
```

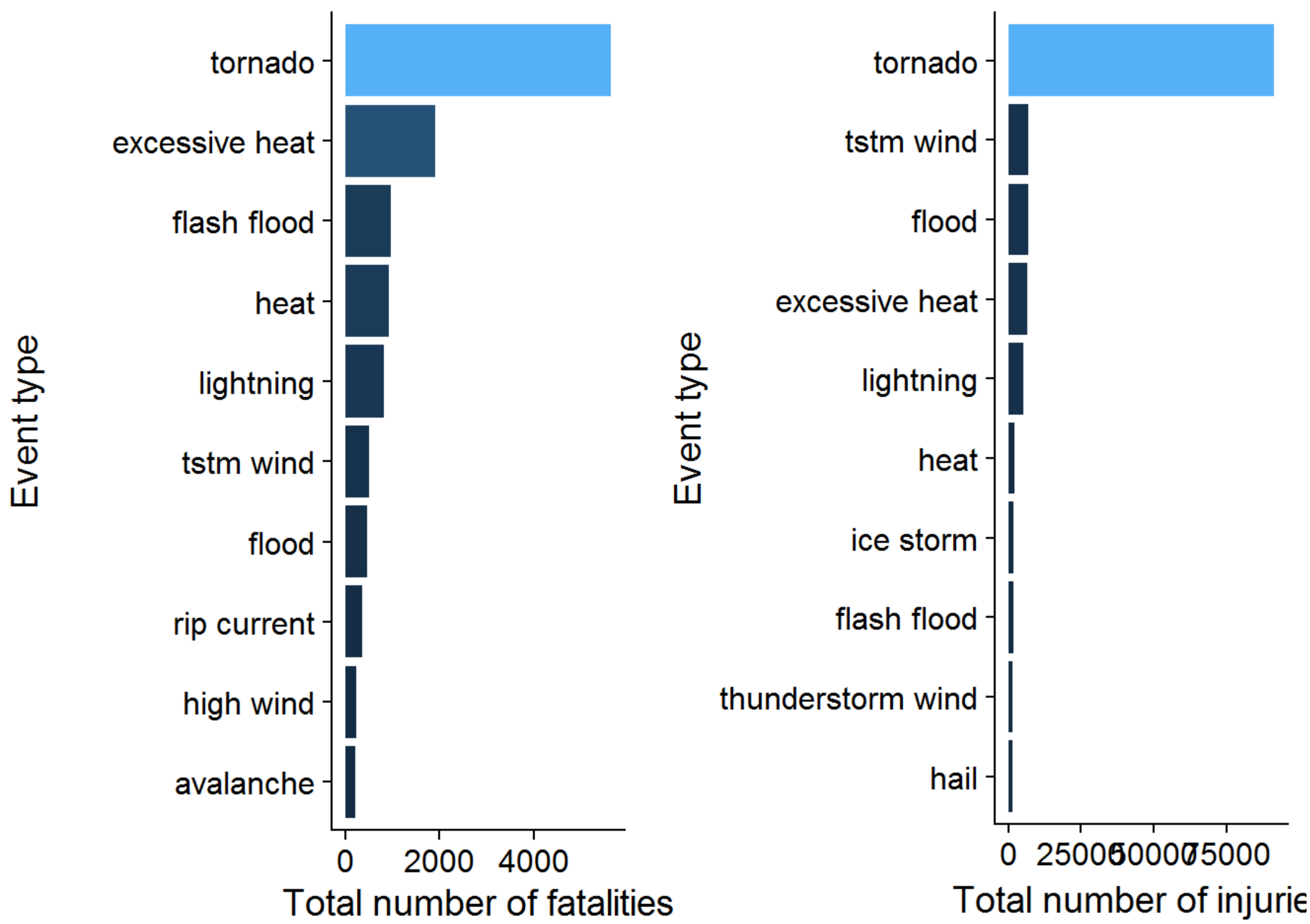
```
  ylab("Total number of injuries") +
  xlab("Event type") +
  theme(legend.position="none")


library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 3.4.1
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

```
plot_grid(p1, p2, align='v')
```

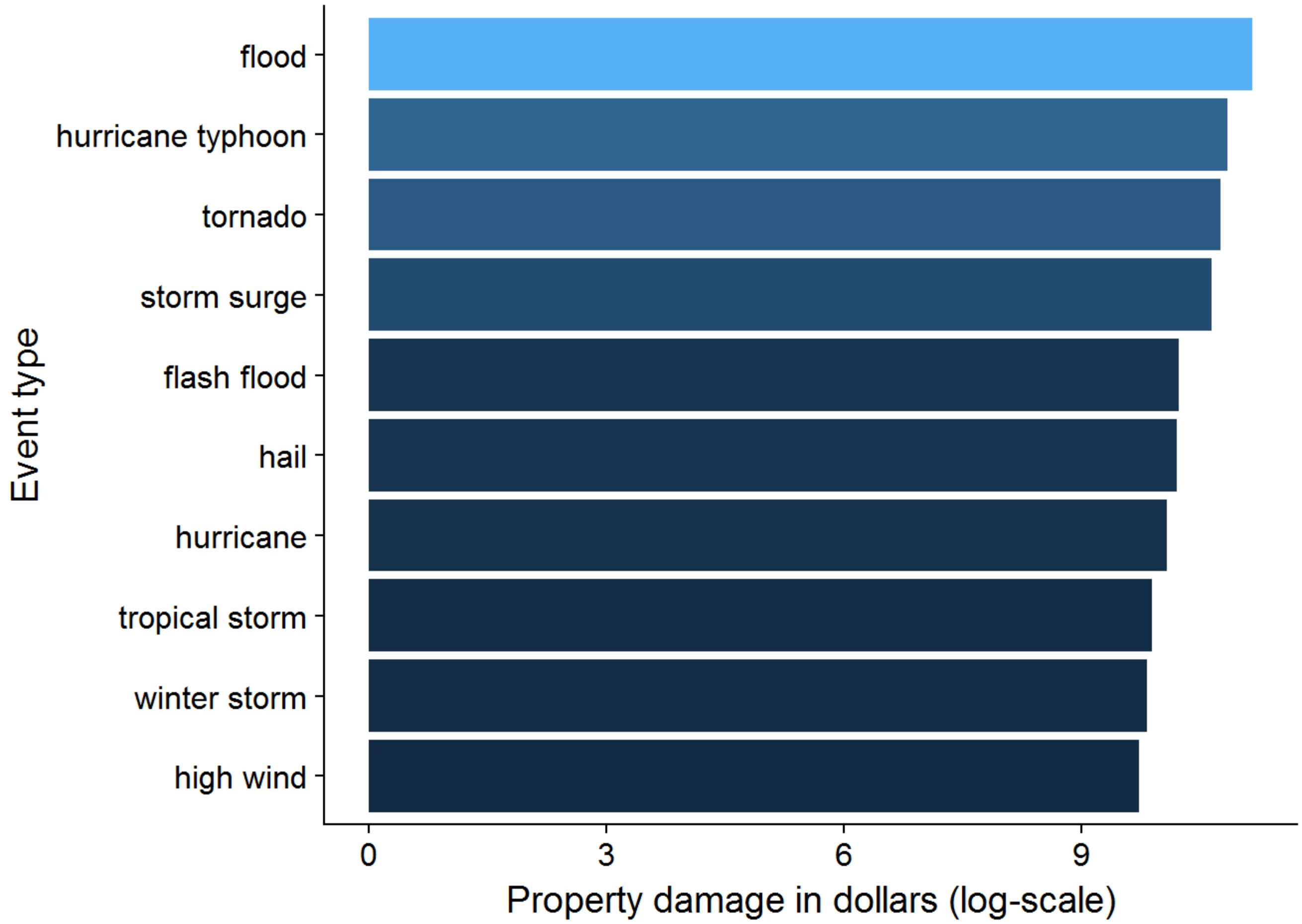


Tornadoes, Excessive Heat and Flash Floods are the most dangerous events leading to the maximum number of fatalities. On the injury front tornado still is the most dominant force, followed by tstm wind and flood. Overall tornadoes are a major force of havoc.

Economic Impact

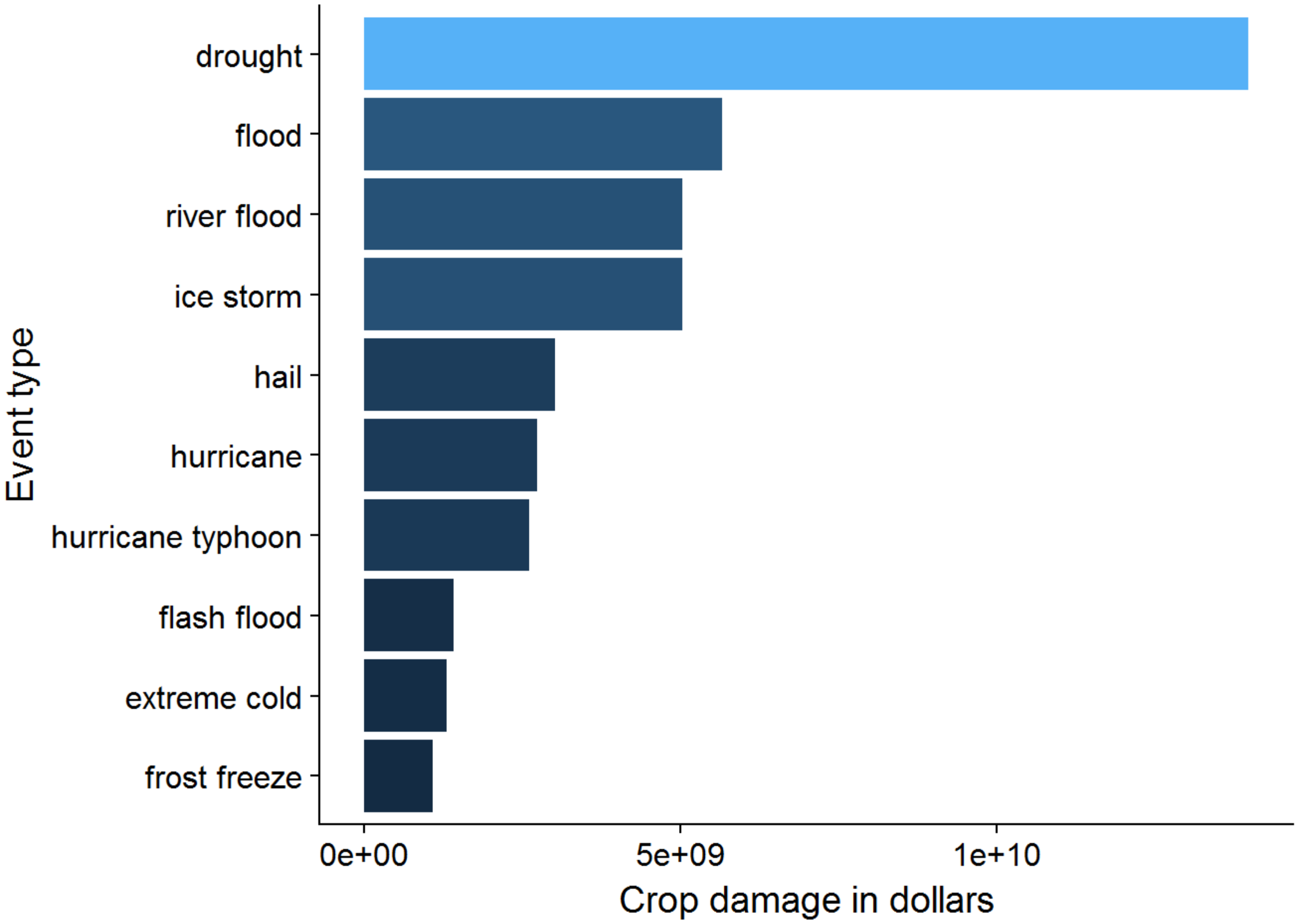
Plots shows the most damage-prone disasters from 1950s

```
library(ggplot2)

# Set the levels in order
p1 <- ggplot(data=prop_dmg_events,
             aes(x=reorder(EVTYPE, prop_dmg), y=log10(prop_dmg), fill=prop_dmg )) +
  geom_bar(stat="identity") +
  coord_flip() +
  xlab("Event type") +
  ylab("Property damage in dollars (log-scale)") +
  theme(legend.position="none")

p2 <- ggplot(data=crop_dmg_events,
             aes(x=reorder(EVTYPE, crop_dmg), y=crop_dmg, fill=crop_dmg)) +
  geom_bar(stat="identity") +
  coord_flip() +
  xlab("Event type") +
  ylab("Crop damage in dollars") +
  theme(legend.position="none")

p1
```



Flood , hurricane and tornado are the major sources of devastation on the front of property damage.

```
p2
```



The graph shows the damage on crops on the major disasters. As can be guessed by intuition, drought appears to be the major source of destruction on crops followed by flood, where as river flood and ice storm are in major contestants of the 3rd place.