Human and Economic Storm Damage Impact in the United States

Synopsis

This report indentifies the most impactful weather events affecting human populations within the United States in terms of fatalities, injuries and economic losses. The Storm Events Database complied by the National Oceanic and Atmospheric Administration (NOAA) from 1950 to 2011 will be used to answer the following questions:

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

Data Processing

Required Packages

To download and read the data, we need the following packages:

```
install.packages("downloader", repos="http://cran.us.r-project.org")
install.packages("R.utils", , repos="http://cran.us.r-project.org")

library(downloader)
library(tools)    # md5sum
library(R.utils)  # bunzip2d
```

Source Data

```
if(!file.exists("data")){dir.create("data")}

if(!file.exists("data/repdata-data-StormData.csv")) {
  fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(fileUrl,destfile="./data/repdata-data-StormData.csv.bz2",method="curl")
  bunzip2("data/repdata-data-StormData.csv.bz2")
}
stormDataRaw <- read.csv("data/repdata-data-StormData.csv")
```

Dimensions and Structure of the Storm Events Data

```
dim(stormDataRaw)
```

```
## [1] 902297     37
```

```
str(stormDataRaw)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 2 2 ...
```

```
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels ""," Christiansburg",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels ""," 0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels ""," CANTON"," TULIA",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels ""," CI","%SD",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ 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   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Data Cleaning

We will be using `dplyr` package for data summarystormData$eventType[grepl("thun.*orm", stormData$EVTYPE, ignore.case = TRUE)] and manipulation.

```r
# Use dplyr
install.packages("dplyr", repos="http://cran.us.r-project.org")
library(dplyr)
```

Subset data

First take a subset of relevant column variables:

- BGN_DATE
- EVTYPE
- FATALITIES
- INJURIES
- PROPDMG
- PROPDMGEXP
- CROPDMG
- CROPDMGEXP

```r
stormData <- dplyr::select(stormDataRaw, BGN_DATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CR
```

Date Conversion

Convert the text date to a date type.

```r
# Convert date
stormData$beginDate <- as.Date(strptime(stormData$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"))

minDate <- stormData %>% filter(beginDate == min(beginDate))
maxDate <- stormData %>% filter(beginDate == max(beginDate))

minDate$beginDate[1]
```

```
## [1] "1950-01-03"
```

```r
maxDate$beginDate[1]
```

```
## [1] "2011-11-30"
```

Standardize Events

In many cases, there are multiple event types for the same meteorological event. For example, there are 88 event types that describe thunderstorms. The following code will consolidate related event types into single event types.

```r
length(unique(stormData$EVTYPE[grepl("thun.*orm", stormData$EVTYPE, ignore.case = TRUE)]))
```

```
## [1] 88
```

```r
# Create another column to consolidate event types
stormData <- mutate(stormData, eventType = as.character(stormData$EVTYPE))

stormData$eventType[grepl("tornad", stormData$eventType, ignore.case = TRUE)]         <- "TORNADO"
stormData$eventType[grepl("thun.*orm", stormData$eventType, ignore.case = TRUE)]      <- "THUNDERSTORM
stormData$eventType[grepl("tstm", stormData$eventType, ignore.case = TRUE)]           <- "THUNDERSTORM
stormData$eventType[grepl("snow", stormData$eventType, ignore.case = TRUE)]           <- "SNOW"
stormData$eventType[grepl("blizzard", stormData$eventType, ignore.case = TRUE)]       <- "BLIZZARD"
stormData$eventType[grepl("hail", stormData$eventType, ignore.case = TRUE)]           <- "HAIL"
stormData$eventType[grepl("rain", stormData$eventType, ignore.case = TRUE)]           <- "RAIN"
stormData$eventType[grepl("precip", stormData$eventType, ignore.case = TRUE)]         <- "RAIN"
stormData$eventType[grepl("hurricane", stormData$eventType, ignore.case = TRUE)]      <- "HURRICANE"
stormData$eventType[grepl("tropical.*storm", stormData$eventType, ignore.case = TRUE)] <- "TROPICALSTORM
stormData$eventType[grepl("flood", stormData$eventType, ignore.case = TRUE)]          <- "FLOOD"
stormData$eventType[grepl("fire", stormData$eventType, ignore.case = TRUE)]           <- "FIRE"
stormData$eventType[grepl("lightn", stormData$eventType, ignore.case = TRUE)]         <- "LIGHTNING"
stormData$eventType[grepl("wind", stormData$eventType, ignore.case = TRUE)]           <- "WIND"
stormData$eventType[grepl("cold", stormData$eventType, ignore.case = TRUE)]           <- "COLD"
stormData$eventType[grepl("heat", stormData$eventType, ignore.case = TRUE)]           <- "HEAT"
stormData$eventType[grepl("storm surge", stormData$eventType, ignore.case = TRUE)]    <- "STORM SURGE"
stormData$eventType[grepl("freeze", stormData$eventType, ignore.case = TRUE)]         <- "FREEZE"
stormData$eventType[grepl("frost", stormData$eventType, ignore.case = TRUE)]         <- "FREEZE"
stormData$eventType <- as.factor(stormData$eventType)
```

Results

Load knitr package for table display.

```
install.packages("knitr", repos="http://cran.us.r-project.org")
library(knitr)
```

Aggregate event damage

To compute human and property damage, aggregate the sums of damage from relevant columns. For instance, total economic damage per event will combine property and crop damage aggregate sums. Total human damage by event will be combine fatalities and injuries.

Compute Economic Damage

Use formattable for currency formatting

```
install.packages("formattable", repos="http://cran.us.r-project.org")
library(formattable)
```

```
# Calculate property damage amounts
PROPDMGEXP  <- levels(stormData$PROPDMGEXP)
pMultiplier <- c(1,1,1,1,1,10,100,1000,10000,100000,1000000,10000000,100000000,1000000000,100,100,1000,
propLookup  <- data.frame(cbind(PROPDMGEXP,pMultiplier))
propLookup$pMultiplier <- as.numeric(as.character(propLookup$pMultiplier))
stormData <- merge(stormData,propLookup)
stormData$totalPropDamage <- stormData$PROPDMG*stormData$pMultiplier
# Compute crop damage
CROPDMGEXP <- levels(stormData$CROPDMGEXP)
cMultiplier <- c(1,1,1,100,1000000000,1000,1000,1000000,1000000)
cropLookup  <- data.frame(cbind(CROPDMGEXP, cMultiplier))
cropLookup$cMultiplier <- as.numeric(as.character(cropLookup$cMultiplier))
stormData <- merge(stormData,cropLookup)
stormData$totalCropDamage <- stormData$CROPDMG*stormData$cMultiplier
stormData$totalDamage <- stormData$totalPropDamage + stormData$totalCropDamage
# generate top economic damage
topEconDamage <- stormData %>% group_by(eventType) %>% summarise(sumTotalDamage=sum(totalDamage)) %>% a
```

```
## Selecting by sumTotalDamage
```

```
topEconDamage$sumTotalDamage <- currency(topEconDamage$sumTotalDamage, symbol = "$")
```

Top Economic Storm Damage

```
kable(topEconDamage, col.names=c("Event","Total Property and Crop Damage"), caption = "Top 12 Economic
```

Table 1: Top 12 Economic Impact from Storm Events

Event	Total Property and Crop Damage
FLOOD	$180,456,574,932.91
HURRICANE	$90,271,472,810.00
TORNADO	$59,020,779,446.50
STORM SURGE	$47,965,579,000.00

Event	Total Property and Crop Damage
HAIL	$19,024,452,135.70
DROUGHT	$15,018,672,000.00
THUNDERSTORM	$12,457,075,688.10
ICE STORM	$8,967,041,360.00
FIRE	$8,904,910,130.00
TROPICALSTORM	$8,409,286,550.00
WIND	$7,038,656,523.00
WINTER STORM	$6,715,441,251.00

The table above shows the 12 top most impactful storm events in terms of economic damage. Across the United States, floods, hurricanes, tornados are the top 3 most harmful with respect to economic consequences.

Compute Human Damage

```
# Compute human damage
topHumanFatalities <- stormData %>% group_by(eventType) %>% summarise(sumHumanFatalities=sum(FATALITIES)
topHumanInjuries <- stormData %>% group_by(eventType) %>% summarise(sumHumanInjuries=sum(INJURIES)) %>%

# Combine all top damage data sets
allDamage <- merge(topEconDamage, topHumanFatalities)
allDamage <- merge(allDamage, topHumanInjuries)
allDamage$sumHumanImpact <- allDamage$sumHumanFatalities + allDamage$sumHumanInjuries

# Generate top human damage
topHumanDamage <- arrange(allDamage,desc(sumHumanImpact))  %>% select(eventType,sumHumanImpact, sumHuman
```

```
## Selecting by sumHumanInjuries
```

Top Human Damage

```
kable(topHumanDamage, col.names=c("Event", "Human Impact", "Fatalites", "Injuries"), caption = "Top 12 
```

Table 2: Top 12 Human Impact from Storm Events

Event	Human Impact	Fatalites	Injuries
TORNADO	97068	5661	91407
THUNDERSTORM	10274	730	9544
FLOOD	10128	1524	8604
WIND	2640	689	1951
ICE STORM	2064	89	1975
FIRE	1698	90	1608
WINTER STORM	1527	206	1321
HURRICANE	1463	135	1328
HAIL	1386	15	1371
TROPICALSTORM	449	66	383
STORM SURGE	67	24	43
DROUGHT	4	0	4

The table above shows the 12 top most impactful storm events on human populations in terms of fatalities and injuries. Across the United States, tornado, thunderstorm and flood events are most harmful with respect to population health.

Data Visualization

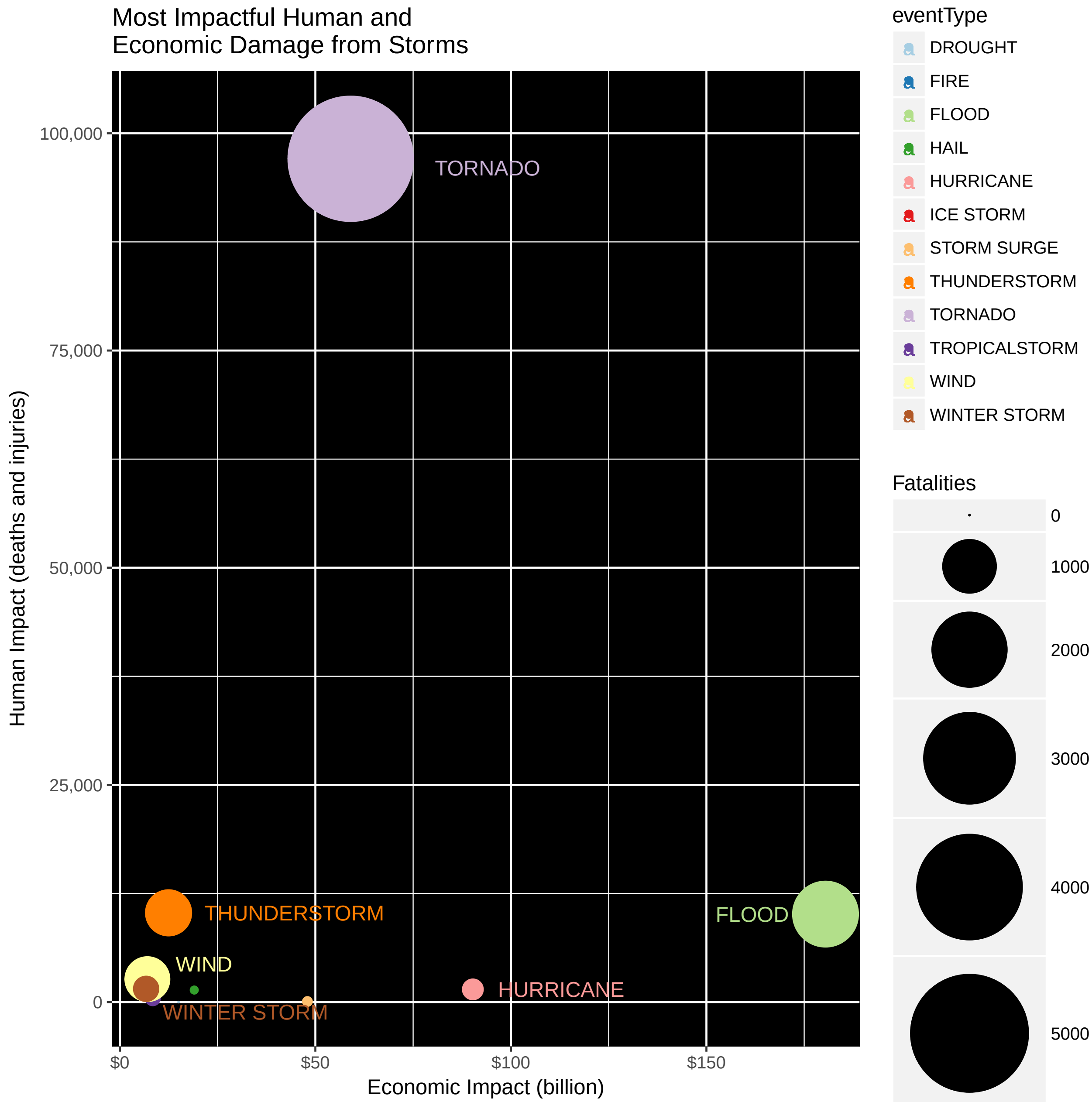
Required Packages

```
install.packages("ggplot2", repos="http://cran.us.r-project.org")
library(ggplot2)
```

Show total damage as 2d scatter plot with total economic damage as x-axis, total human impact as y-axis and fatalities as point size.

```
topDamage <- allDamage[1:12,]

div = 1000000000
ggplot(topDamage, aes(x=sumTotalDamage/div, y=sumHumanImpact, color=eventType)) +
    geom_point(aes(size = sumHumanFatalities)) +
    labs(size="Fatalities") +
    geom_text(data = topDamage[topDamage$eventType=="TORNADO" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.8, vjust = 1.1) +
    geom_text(data = topDamage[topDamage$eventType=="HEAT" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = 1.5) +
    geom_text(data = topDamage[topDamage$eventType=="FLOOD" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = 1.5) +
    geom_text(data = topDamage[topDamage$eventType=="LIGHTNING" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.3, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="THUNDERSTORM" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.2, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="HURRICANE" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.2, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="WIND" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.5, vjust = -.5) +
    geom_text(data = topDamage[topDamage$eventType=="RIP CURRENT" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = 1, vjust = -1) +
    geom_text(data = topDamage[topDamage$eventType=="AVALANCHE" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), vjust = -.3) +
    geom_text(data = topDamage[topDamage$eventType=="COLD" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = .5, vjust = -1) +
    geom_text(data = topDamage[topDamage$eventType=="WINTER STORM" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.1, vjust = 2) +
    scale_colour_brewer(palette="Paired")   +
    scale_size_area(max_size=30) +
    scale_x_continuous(labels = scales::dollar, limits = c(min(topDamage$sumTotalDamage)/div,
                                                           max(topDamage$sumTotalDamage)/div))  +
    scale_y_continuous(labels = scales::comma, limits = c(min(topDamage$sumHumanImpact),
                                                          max(topDamage$sumHumanImpact)+5000)) +
    xlab("Economic Impact (billion)") +
    ylab("Human Impact (deaths and injuries)") +
    theme(panel.background = element_rect(fill = "black")) +
    ggtitle("Most Impactful Human and\nEconomic Damage from Storms")
```



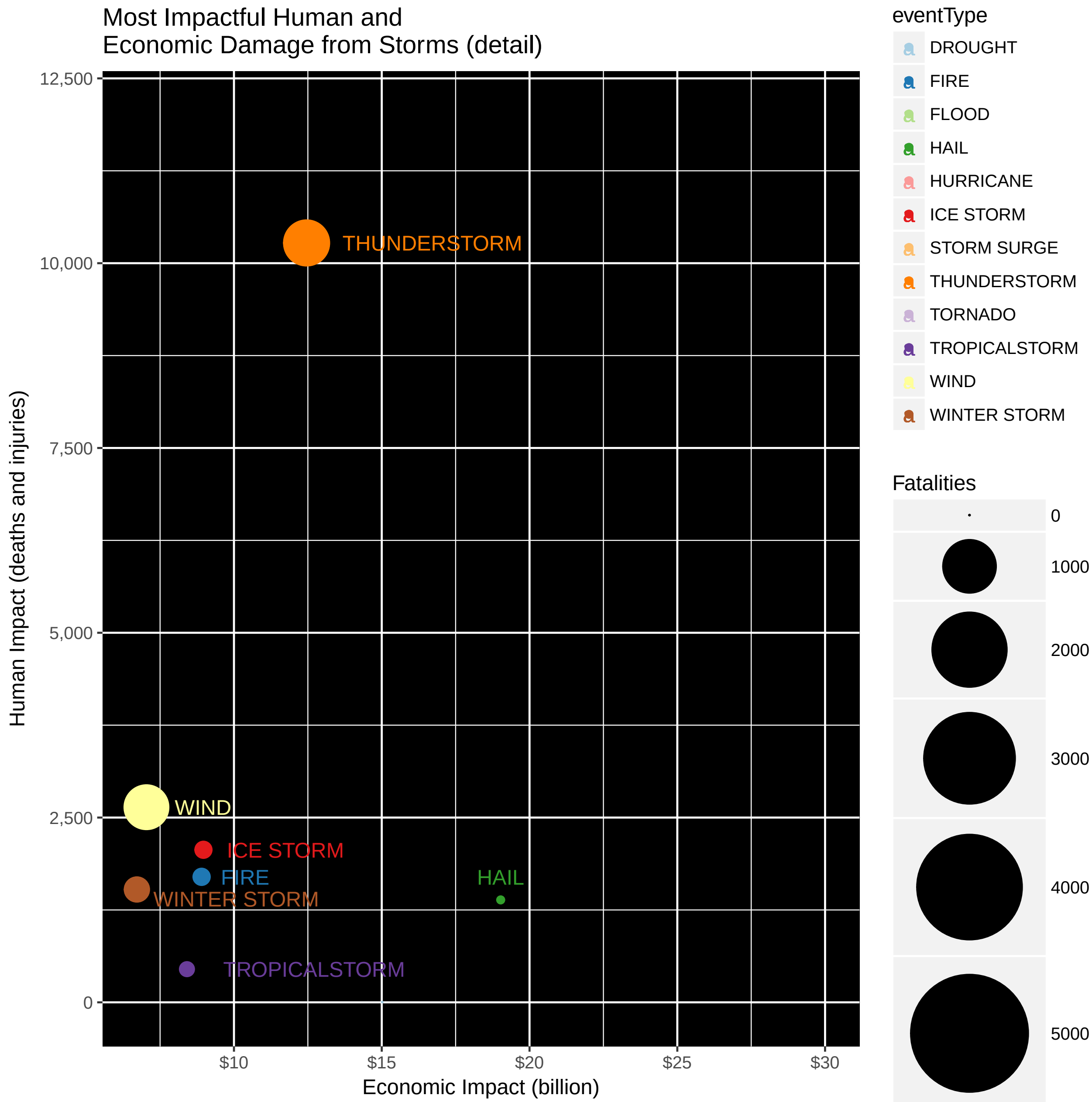
The graph above shows the magnitude of harm to human life that comes from tornados with moderate economic damage. Tornados are far and away the most dangerous for human safety. Where as floods produce the most economic damage and relatively moderate human impact. Hurricanes produce the second most impactful economic damage and less human harm than thunderstorms and wind.

To better understand economic damage less than $100 billion and 12,500 human lives affected, zoom into storm data to focus on these events.

```
div = 1000000000
ggplot(topDamage, aes(x=sumTotalDamage/div, y=sumHumanImpact, color=eventType)) +
    geom_point(aes(size = sumHumanFatalities)) +
    labs(size="Fatalities") +
    geom_text(data = topDamage[topDamage$eventType=="THUNDERSTORM" ,],
              aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.2, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="WIND" ,],
```

```r
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.5, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="ICE STORM" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.2, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="COLD" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = .5, vjust = -1) +
    geom_text(data = topDamage[topDamage$eventType=="FIRE" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.4, vjust = .5) +
    geom_text(data = topDamage[topDamage$eventType=="WINTER STORM" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.1, vjust = 1.1) +
    geom_text(data = topDamage[topDamage$eventType=="HAIL" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = .5, vjust = -1) +
    geom_text(data = topDamage[topDamage$eventType=="TROPICALSTORM" ,],
             aes(sumTotalDamage/div,sumHumanImpact, label = eventType), hjust = -.2, vjust = .5) +
    scale_colour_brewer(palette="Paired")   +
    scale_size_area(max_size=30) +
    scale_x_continuous(labels = scales::dollar, limits = c(min(topDamage$sumTotalDamage)/div,
                                                         max(30)))  +
    scale_y_continuous(labels = scales::comma, limits = c(min(topDamage$sumHumanImpact),
                                                          max(12000))) +
    xlab("Economic Impact (billion)") +
    ylab("Human Impact (deaths and injuries)") +
    theme(panel.background = element_rect(fill = "black")) +
    ggtitle("Most Impactful Human and\nEconomic Damage from Storms (detail)")
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



The graph above is a zoomed in view to focus on storm events clustered under 12,500 human impact and less than $30 billion dollars.

Being able to see this data more clearly after zooming in, an insight that I personally gained from this graph was that indeed, as the Talking Heads sang, "air can hurt you too". In this case wind is pretty effective at killing you.

References

coursera: Reproducible Research by John Hopkins University Bloomberg School of Public Health Roger D. Peng, PhD, Associate Professor, Biostatistics, Jeff Leek, PhD, Associate Professor, Biostatistics, and Brian Caffo, PhD, Professor, Biostatistics.

Storm Events Database National Oceanic and Atmospheric Administration. United States Department of

Commerce.

Read Headed Step Data blog by Keith Helfrich Data visualization ideas.

knitr Yihui Xie

Fear of Music by Talking Heads Sire Records 1979