Evaluating The Costs of Natural Disasters

Brett Wiens

October 2, 2017

Evaluating the Costs of Natural Disasters

Synopsis

The National Oceanic and Atmospheric Administration, an American agency within the department of commerce collects information about atmospheric events, commonly known as natural disasters. The damages caused by severe weather events can include loss of property and crops, in addition to injury and loss of life.

Development of this analysis began by downloading the database and cleaning up spelling and other errors typical of such large data development efforts. Once cleaned and tidied, three figures were created, a map showing the distribution and damage of natural disasters, and a bar graph of the damages by each type.

The results show that there are far more events on the Eastern part of the country (due in part to larger population) but there have been several several events in the western USA. In terms of fatalities, the most damaging storm type are tornados, in terms of economic cost, the big four are Hurricanes, Floods, Storm Surges, and again, Tornados.

This information could be valuable for emergency planners preparing for response provisionment.

Data Processing

the data was first downloaded from the source [https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2] and read in using read.csv and bzfile functions.

```
sourceURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(sourceURL, "StormData.csv.bz2")
stormData <- read.csv(bzfile("StormData.csv.bz2"), stringsAsFactors = FALSE)
```

Data Cleanup

There are a few glaring and important weaknesses in the data.

One of the most pressing issues in the dataset must clearly be that there are spelling mistakes and inconsistencies in the EVTYPE field. Examples such as “Avalance” should obviously be converted into “Avalanche” so that they are appropriately evaluated.

```
##First, remove all events with no cost or injury.
CleanStorm <- stormData[,c(2,6,7,8,23:28)]
CleanStorm <- CleanStorm[(CleanStorm$PROPDMG + CleanStorm$CROPDMG + CleanStorm$INJURIES + CleanStorm$FATALITIES > 0),]
```

```
## Then make them all capitals
CleanStorm$Event <- toupper(CleanStorm$EVTYPE)

## Consolidate and Spelling Corrections
CleanStorm$Event <-  ifelse(grepl("(.*)CHILL(.*)", CleanStorm$Event),"WND CHILL",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("AVALANCE", CleanStorm$Event),"AVALANCHE",CleanS
torm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)BLIZZARD(.*)", CleanStorm$Event),"BLIZZARD"
,CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)BLOW-OUT(.*)", CleanStorm$Event),"BLOW-OUT T
IDE",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)BLOWING SNOW(.*)", CleanStorm$Event),"BLOWI
NG SNOW",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)BRUSH(.*)", CleanStorm$Event),"BRUSH FIRE",
CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)COASTAL(.*)", CleanStorm$Event),"COASTAL FLO
OD",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)DAM(.*)", CleanStorm$Event),"DAM BREACH",Cle
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FREEZ(.*)", CleanStorm$Event),"COLD WEATHER"
,CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)COLD(.*)", CleanStorm$Event),"COLD WEATHER",
CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)DOWNBURST(.*)", CleanStorm$Event),"DOWNBURST
",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)DRY(.*)", CleanStorm$Event),"DRY WEATHER",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)DROUGHT(.*)", CleanStorm$Event),"DROUGHT",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)DUST(.*)", CleanStorm$Event),"DUST STORM",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)EARLY SNOW(.*)", CleanStorm$Event),"EARLY SN
OW",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HEAT(.*)", CleanStorm$Event),"HOT WEATHER",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FLOOD(.*)", CleanStorm$Event),"FLOOD",Clean
Storm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FOG(.*)", CleanStorm$Event),"FOG",CleanStor
m$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FROST(.*)", CleanStorm$Event),"FROST",Clean
Storm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FUNNEL(.*)", CleanStorm$Event),"TORNADO",Cle
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)GLAZE(.*)", CleanStorm$Event),"ICE STORM",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)GRADIENT(.*)", CleanStorm$Event),"HIGH WIND
",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)GUST(.*)", CleanStorm$Event),"HIGH WIND",Cle
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HAIL(.*)", CleanStorm$Event),"HAIL",CleanSto
rm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)WIND(.*)", CleanStorm$Event),"HIGH WIND",Cle
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HEAVY(.*)", CleanStorm$Event),"HEAVY PRECIPI
TATION",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)SURF(.*)", CleanStorm$Event),"HIGH SURF",Cle
```

```
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HOT(.*)", CleanStorm$Event),"HOT WEATHER",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HURRICANE(.*)", CleanStorm$Event),"HURRICANE
",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HVY(.*)", CleanStorm$Event),"HEAVY PRECIPIT
ATION",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)ICE(.*)", CleanStorm$Event),"COLD WEATHER",
CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)SLIDE(.*)", CleanStorm$Event),"LANDSLIDE",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)SLUMP(.*)", CleanStorm$Event),"LANDSLUMP",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)LIGHT(.*)SNOW(.*)", CleanStorm$Event),"LIGH
T SNOW",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)LIGHTING(.*)", CleanStorm$Event),"THUNDERST
ORM",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)LIGHTNING(.*)", CleanStorm$Event),"THUNDERST
ORM",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)LOW(.*)TEMP(.*)", CleanStorm$Event),"COLD WE
ATHER",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)MICROBURST(.*)", CleanStorm$Event),"MICROBUR
ST",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)PRECIP(.*)", CleanStorm$Event),"HEAVY PRECI
PITATION",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)MUD(.*)", CleanStorm$Event),"MUD SLIDE",Cle
anStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)RIP CURRENT(.*)", CleanStorm$Event),"RIP CUR
RENT",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)THUNDERSTORM(.*)", CleanStorm$Event),"THUND
ERSTORM",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)HAIL(.*)", CleanStorm$Event),"HAIL",CleanSto
rm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)STORM SURGE(.*)", CleanStorm$Event),"STORM
SURGE",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)TORNADO(.*)", CleanStorm$Event),"TORNADO",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)TROPICAL(.*)", CleanStorm$Event),"TROPICAL
STORM",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)TSTM(.*)", CleanStorm$Event),"TROPICAL STORM
",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)VOLCAN(.*)", CleanStorm$Event),"VOLCANIC",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FLD(.*)", CleanStorm$Event),"FLOOD",CleanSt
orm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)FIRE(.*)", CleanStorm$Event),"FIRE",CleanSto
rm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)WINTER(.*)", CleanStorm$Event),"WINTER STOR
M",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)WINTRY(.*)", CleanStorm$Event),"WINTER STOR
M",CleanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)SNOW(.*)", CleanStorm$Event),"SNOW",CleanSto
rm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)TORNDAO(.*)", CleanStorm$Event),"TORNADO",C
leanStorm$Event)
CleanStorm$Event <-  ifelse(grepl("(.*)WATERSPOUT(.*)", CleanStorm$Event),"WATERSPO
UT",CleanStorm$Event)
```

```
CleanStorm$Event <-  ifelse(grepl("(.*)WND(.*)", CleanStorm$Event),"WIND CHILL",Cl
eanStorm$Event)
```

Property and Crop Damage are represented in two columns, a number and a multiplier. These need to be consolidated for proper analysis. Once they are corrected, the data is aggregated by both sums and means to demonstrate the total and average damage for each event type. (Fatalities, Injuries, Property Damage, Crop Damage)

```
## Property and Crop Damage is undervalued (represented by H = 100, K = 1,000, M =
1,000,000, B = 1,000,000,000)
## Need to clean that up a little bit.

CleanStorm$FullPropertyCost <- as.numeric(ifelse(CleanStorm$PROPDMGEXP == "H", Clea
nStorm$PROPDMG*100,
                                ifelse(CleanStorm$PROPDMGEXP == "K", CleanStorm$PROP
DMG*1000,
                                ifelse(CleanStorm$PROPDMGEXP == "M", CleanStorm$PROP
DMG*1000000,
                                ifelse(CleanStorm$PROPDMGEXP == "B", CleanStorm$PROP
DMG*1000000000,
                                ifelse(CleanStorm$PROPDMGEXP == "", CleanStorm$PROPD
MG,""))))))
CleanStorm$FullCropCost <- as.numeric(ifelse(CleanStorm$CROPDMGEXP == "H", CleanSto
rm$CROPDMG*100,
                                ifelse(CleanStorm$CROPDMGEXP == "K", CleanStorm$CROP
DMG*1000,
                                ifelse(CleanStorm$CROPDMGEXP == "M", CleanStorm$CROP
DMG*1000000,
                                ifelse(CleanStorm$CROPDMGEXP == "B", CleanStorm$CROP
DMG*1000000000,
                                ifelse(CleanStorm$CROPDMGEXP == "", CleanStorm$CROPD
MG,""))))))

## There is a major typo in the NAPA valley flood, 2006.  The actual damage was $11
6 Million, but here it was reported as $116 Billion.
CleanStorm$FullPropertyCost <- ifelse((CleanStorm$FullPropertyCost == "1.15e+11"),
CleanStorm$FullPropertyCost/1000, CleanStorm$FullPropertyCost)

SummarizedDamage <- aggregate(FullCropCost ~ Event, CleanStorm, FUN = sum)
SummarizedDamage <- merge(SummarizedDamage, aggregate(FullPropertyCost ~ Event, Cle
anStorm, FUN = sum), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(FATALITIES ~ Event, CleanStor
m, FUN = sum), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(INJURIES ~ Event, CleanStorm,
FUN = sum), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(FullPropertyCost ~ Event, Cle
anStorm, FUN = mean), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(FullCropCost ~ Event, CleanS
torm, FUN = mean), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(FATALITIES ~ Event, CleanStor
m, FUN = mean), by = "Event")
SummarizedDamage <- merge(SummarizedDamage, aggregate(INJURIES ~ Event, CleanStorm,
FUN = mean), by = "Event")
names(SummarizedDamage) <- c("Event Type", "Total Crop Damage",  "Total Property Da
mage",  "Total Fatalities", "Total Injuries", "Mean Property Damage", "Mean Crop Da
```

```
mage", "Mean Fatalities", "Mean Injuries")


SummarizedDamage$'Total Cost' <- SummarizedDamage$`Total Crop Damage` + SummarizedD
amage$`Total Property Damage`
SummarizedDamage$'Human Impact' <- SummarizedDamage$`Total Injuries` + SummarizedDa
mage$`Total Fatalities`



TotalCost <- SummarizedDamage[order(SummarizedDamage$'Total Cost', decreasing = TRU
E),c(1,2,3,10)]
TotalCost <- TotalCost[TotalCost$'Total Cost' > 10000,]
TotalHuman <- SummarizedDamage[order(SummarizedDamage$'Human Impact', decreasing =
TRUE),c(1,4,5,11)]
TotalHuman <- TotalHuman[TotalHuman$'Human Impact' > 10,]

## Prettying up the numbers makes them more readable.
TotalCost$'$ Total Cost' <- prettyNum(TotalCost$`Total Cost`, big.mark=",", scienti
fic = FALSE)
TotalCost$'$ Total Crop Damage' <- prettyNum(TotalCost$`Total Crop Damage`, big.mar
k=",", scientific = FALSE)
TotalCost$'$ Total Property Damage' <- prettyNum(TotalCost$`Total Property Damage`,
big.mark=",", scientific = FALSE)

TotalHuman$'| Human Impact |' <- prettyNum(TotalHuman$`Human Impact`, big.mark=",",
scientific = FALSE)
TotalHuman$'| Total Fatalities |' <- prettyNum(TotalHuman$`Total Fatalities`, big.m
ark=",", scientific = FALSE)
TotalHuman$'| Total Injuries |' <- prettyNum(TotalHuman$`Total Injuries`, big.mark=
",", scientific = FALSE)

DollarPost <- head(TotalCost[,c(1,5,6,7)],10)
HumanPost <- head(TotalHuman[, c(1,5,6,7)],10)
```

Results / Mapping

The top ten natural disasters in terms of economic damage are:

```
DollarPost
```

```
##        Event Type   $ Total Cost $ Total Crop Damage
## 33      HURRICANE 90,131,472,810       5,495,292,800
## 20          FLOOD 65,092,953,394      12,388,587,200
## 59        TORNADO 57,396,759,093         414,961,310
## 57    STORM SURGE 47,965,579,000             855,000
## 23           HAIL 20,728,717,154       3,113,795,853
## 31      HIGH WIND 17,918,407,334       1,953,921,608
## 13        DROUGHT 15,018,927,780      13,972,621,780
## 8    COLD WEATHER 12,261,981,160       8,031,060,850
## 19           FIRE  8,899,910,130         403,281,630
## 61 TROPICAL STORM  8,411,073,550         694,896,000
##    $ Total Property Damage
## 33          84,636,180,010
## 20          52,704,366,194
## 59          56,981,797,783
## 57          47,964,724,000
## 23          17,614,921,301
## 31          15,964,485,726
## 13           1,046,306,000
## 8            4,230,920,310
## 19           8,496,628,500
## 61           7,716,177,550
```

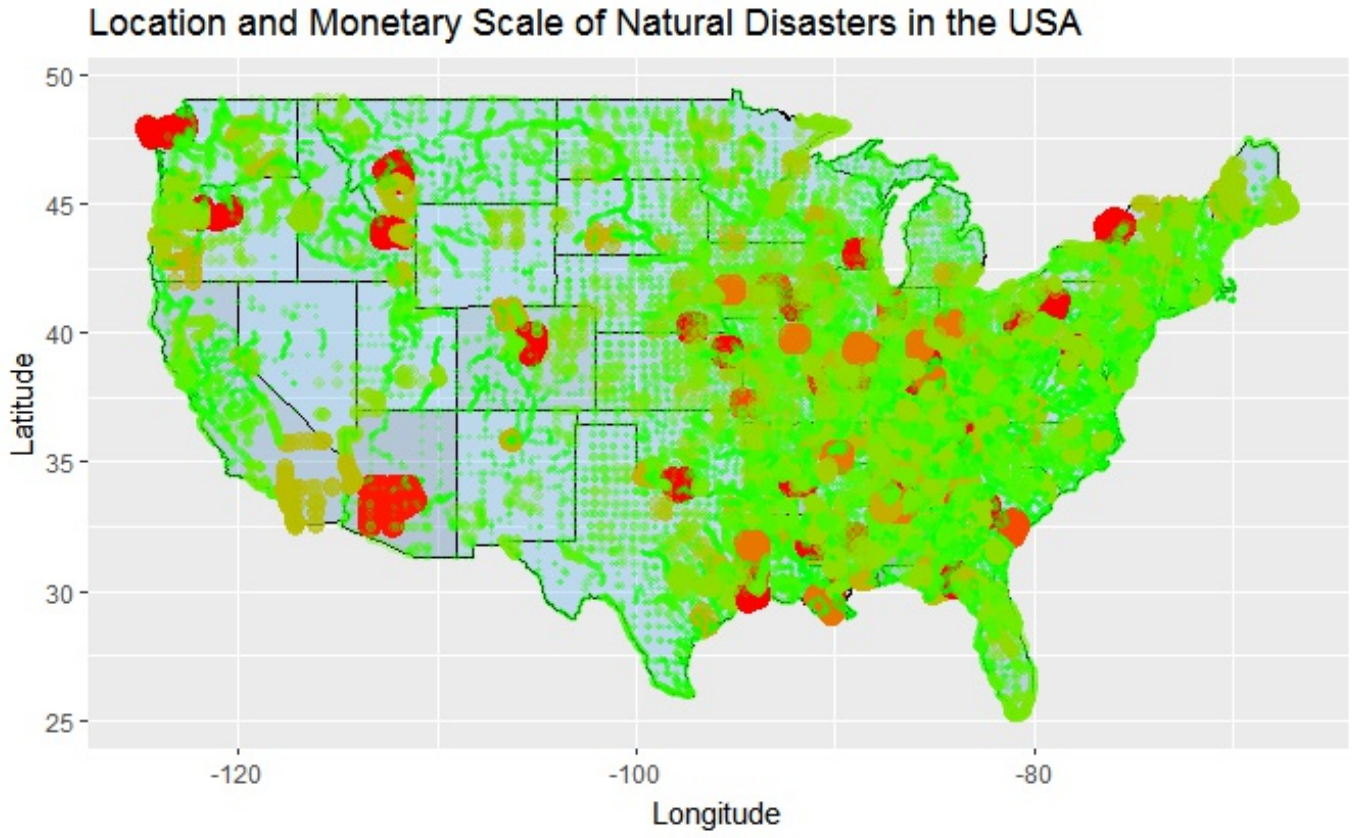
The top ten natural disasters in terms of human casualties are:

```
HumanPost
```

```
##             Event Type | Human Impact | | Total Fatalities |
## 59             TORNADO           97,046                5,636
## 31           HIGH WIND           12,532                1,183
## 32         HOT WEATHER           12,341                3,132
## 20               FLOOD           10,242                1,557
## 58        THUNDERSTORM            6,089                  818
## 8         COLD WEATHER            3,051                  359
## 73        WINTER STORM            2,231                  278
## 19                FIRE            1,698                   90
## 24 HEAVY PRECIPITATION            1,687                  285
## 23                HAIL            1,512                   45
##    | Total Injuries |
## 59             91,410
## 31             11,349
## 32              9,209
## 20              8,685
## 58              5,271
## 8               2,692
## 73              1,953
## 19              1,608
## 24              1,402
## 23              1,467
```

The assignment suggested we consider this as though we were a government manager responsible for emergency preparedness. If I was the manager of FEMA, I wouldn't be satsified preparing simply for the largest disasters generally, but by geographic extent. The first plot I am including in this assignment is a map showing the extent and location of all reported natural disasters.

```
CountyAggregateCost <- aggregate((CleanStorm$FullPropertyCost + CleanStorm$FullCrop
Cost) ~ tolower(CleanStorm$COUNTYNAME), CleanStorm, sum)
states = map_data("state")
counties = map_data("county")
names(CountyAggregateCost) <- c("subregion", "cost")
CountyCost <- merge(counties, CountyAggregateCost, by = "subregion", all = TRUE)
CountyCost <- CountyCost[complete.cases(CountyCost),]

mapBase <- ggplot(data = CountyCost, mapping = aes(x = long, y = lat, group = grou
p)) + coord_fixed(1.3) + geom_polygon(data=states, aes(fill = states$order), alpha
= 0.3, color = "black") + geom_point(data = CountyCost, aes(color = cost, alpha =
cost,cex = cost), pch = 20) + scale_fill_gradient(trans = "log") + labs(title = "L
ocation and Monetary Scale of Natural Disasters in the USA", x = "Longitude", y = "
Latitude") + scale_color_gradient2(low = "darkblue", mid = "green", high = "red",
space = "Lab") + theme(legend.position = "none")
mapBase
```



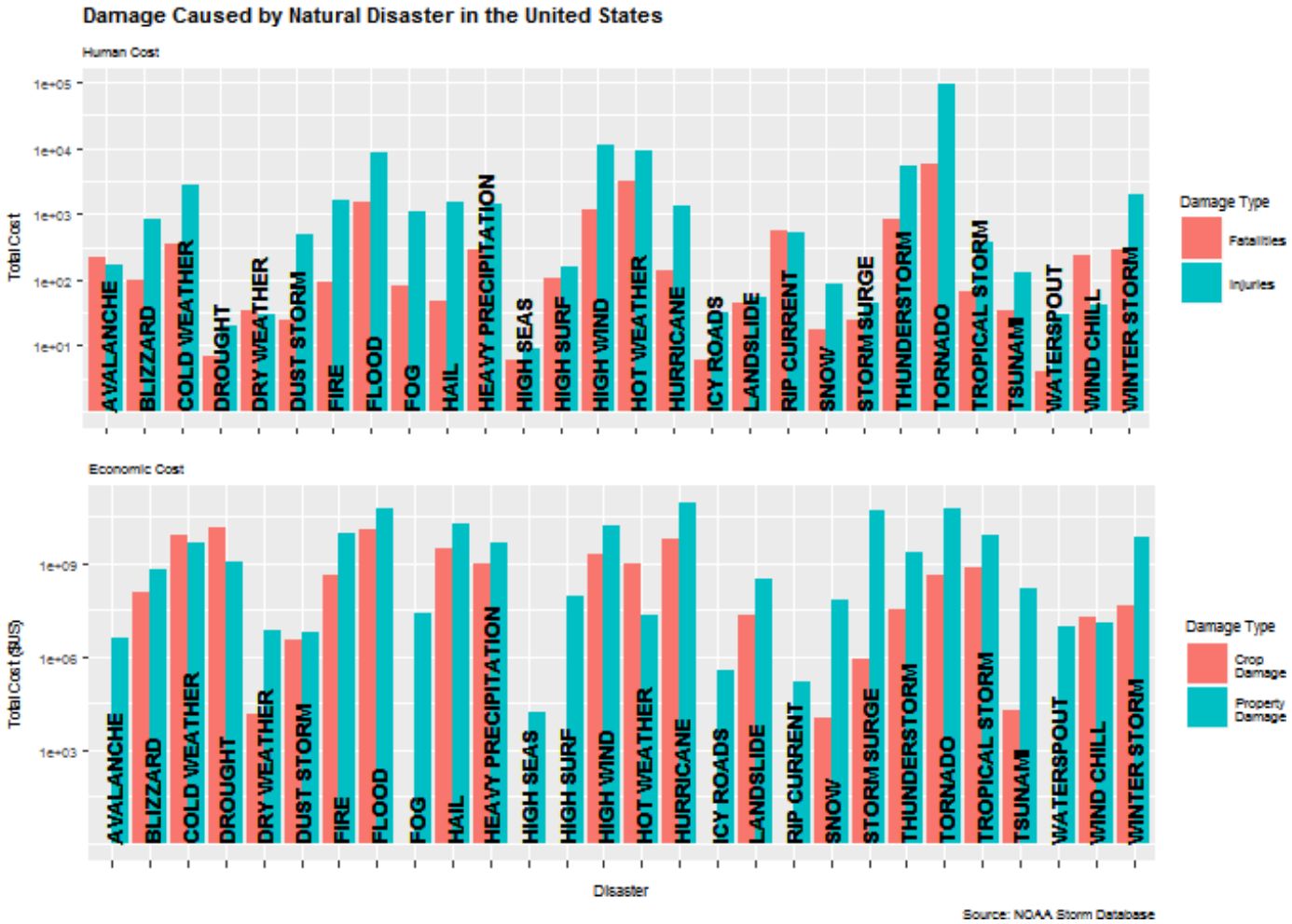
We can see that the density of events that have caused economic or humanitarian misery is greatest on the Eastern half of the country, however, this correlates highly with the larger population base in those areas. The largest damages are spread out throughout the country, with some of the largest in Arizona and Washington, and a surprising lack of enormous events on the Eastern Seabord.

It should be noted that the first time I created this map, it looked nonsensical due to an apparent 115B flood in the NAPA valley, 2006. I looked into this in greater detail and found that it was actually a 115M flood. Taking this into account shows hot-spots throughout the country where additional emergency materials might be best situated.

The second figure I'm including (is actually two combined ggplots) shows the various damages by natural

disaster.

```
ResultsData <- merge(TotalCost, TotalHuman, by = "Event Type")

## Introduced some characters before, correcting to numeric
ResultsData[,c(2:4,8:10)] <- sapply(ResultsData[,c(2:4, 8:10)], as.numeric)

## Data isn't tidy.  Melting the data down so the Damages can be displayed side-by-side in chart
TidyResultsData <- ResultsData %>% gather('Damage Type', 'Damage', c(2:3,8:9))
TidyResultsData <- TidyResultsData[,c(1,10,11)]
TidyResultsData <- TidyResultsData[order(TidyResultsData$`Event Type`),]
TidyMoney <- TidyResultsData[(TidyResultsData$`Damage Type` == "Total Crop Damage"|
TidyResultsData$`Damage Type` == "Total Property Damage"),]
TidyHumans <- TidyResultsData[!(TidyResultsData$`Damage Type` == "Total Crop Damag
e"|TidyResultsData$`Damage Type` == "Total Property Damage"),]
TidyHumans$Damage <- as.numeric(TidyHumans$Damage)

## Create a plot showing injuries and fatalities as they vary by natural disaster.
HumanCostPlot <-
  ggplot(TidyHumans, aes(x = TidyHumans$'Event Type', y = TidyHumans$'Damage'+1))
+ geom_bar(aes(fill = TidyHumans$`Damage Type`), stat = 'identity', position = "dod
ge") + theme(axis.text.x = element_blank()) + labs(title = "Damage Caused by Natur
al Disaster in the United States", x = "", y = "Total Cost", subtitle = "Human Cos
t") + scale_y_log10(breaks = c(10,100,1000,10000,100000)) + scale_fill_discrete(na
me="Damage Type", labels=c("Fatalities","Injuries")) + theme(legend.position = "ri
ght", axis.text = element_text(size = 5), axis.title = element_text(size = 6), plo
t.title = element_text(size = 8), plot.subtitle = element_text(size = 5), legend.t
ext = element_text(size = 5), legend.title = element_text(size = 6)) + geom_text(a
es(label = TidyHumans$`Event Type`, x = TidyHumans$`Event Type`, y = 1, fontface =
"bold"), angle = 90, size = 3,hjust = 0) + theme(plot.margin = unit(c(0,0,0,0),"cm
")) + theme(legend.box.margin = margin(0.01)) + theme(legend.margin = margin(0,0,0
,0, unit = "cm")) + theme(plot.title = element_text(face = "bold"))

## Create a plot showing crop and property damage by disaster type.
EconomicCostPlot <- ggplot(TidyMoney, aes(x = TidyMoney$'Event Type', y = TidyMone
y$'Damage'+1)) + geom_bar(aes(fill = TidyMoney$`Damage Type`), stat = 'identity',
position = "dodge") + theme(axis.text.x = element_blank()) + labs(x = "Disaster",
y = "Total Cost ($US)", subtitle = "Economic Cost", caption = "Source: NOAA Storm
Database") + scale_y_log10(breaks = c(1000,1000000,1000000000)) + scale_fill_discr
ete(name="Damage Type", labels=c("Crop\nDamage","Property\nDamage")) + theme(legend
.position = "right", axis.text = element_text(size = 5), axis.title = element_text
(size = 6), plot.title = element_text(size = 8), plot.subtitle = element_text(size
= 5), legend.text = element_text(size = 5), legend.title = element_text(size = 6),
plot.caption = element_text(size=5))  + geom_text(aes(label = TidyMoney$`Event Typ
e`, x = TidyMoney$`Event Type`, y = 1, fontface = "bold"), angle = 90, size = 3,hj
ust = 0) + theme(plot.margin = unit(c(0,0,0,0),"cm")) + theme(legend.margin = marg
in(0,0,0,0, unit = "cm"))

## Using grid.arrange, two ggplot2 elements (or more) can be placed in the same plo
t space.  Obviously I played with the options and tools a bunch.
grid.arrange(HumanCostPlot,EconomicCostPlot, nrow = 2)
```



It is apparent that Tornados cause the most deaths in the United States, and the second biggest killer is “High Wind.” After cleaning the data, it is clear that a number of weather events cause significant property and crop damage. Hurricanes, Floods, Tornados, and Storm Surges have all created tens of billions of dollars of damage in the US.